Macroeconomic Dynamics, **6**, 2002, 523–547. Printed in the United States of America.
DOI: 10.1017.S1365100501010069

COVARIANCE EFFECT

BARTHOLOMEW MOORE
Fordham University

HUNTLEY SCHALLER
Carleton University

If one drops the strong assumption that firms and households know all of the relevant parameters, and instead models agents as learning these parameters, the estimated parameters become random variables. Taking expectations several periods into the future may then involve taking the expectation of a product of random variables. Because the resulting problem is difficult, previous research has avoided it. This paper makes some progress using both analytical and numerical techniques. Focusing especially on consumption, we find that the resulting covariance terms could account for the well-documented empirical result that consumption displays excess sensitivity to lagged income. We show that a similar covariance effect could play a role in many widely used economic models.

Keywords: Learning, Consumption, Excess Sensitivity

1. INTRODUCTION

The assumption of rational expectations appeals to economists at least in part because we prefer to model agents as rational. However, agents with rational expectations are also assumed to know the true parameters of all relevant stochastic processes. In view of the difficulty that economists have determining basic economic parameters, this is a strong assumption. An alternative is to assume that agents learn economic parameters using data based on their own experience. Of course, if we model agents as learning, we can no longer treat all parameters as known constants. Instead, for the agents in a model, these parameters become random variables (more specifically, statistics that the agents are estimating).

Fully taking into account that an agent's parameter estimates are random variables can significantly affect learning dynamics in a model that has not yet converged to rational expectations equilibrium. Many previous papers have looked at the implications of learning, but they assume a simplified form of learning to make the analysis of expectations more tractable.[1]

In this paper, we focus on models in which an exogenous variable follows a first-order autoregressive process. The AR(1) case is simple and very widely

Address correspondence to: Bartholomew Moore, Department of Economics, Dealy Hall, Fordham University, 441 East Fordham Rd., The Bronx, NY 10458, USA; e-mail: bmoore@fordham.edu.

 1365-1005/02 $9.50

employed in economics. Under rational expectations, the current expectation of what the exogenous variable will be two periods in the future is the square of the autoregressive parameter times the current value of the exogenous variable. In previous work on learning, researchers have treated the agent's estimate of an AR parameter as a known constant. This allowed them to write the expectation as the square of the parameter estimate times the current value of the exogenous variable.

In contrast, under a full modeling of learning, the expectation of an exogenous variable two periods in the future can no longer be expressed in this simple way. Instead, the parameter is unknown, and so, the expectation involves the product of two random variables—the estimate of the autoregressive parameter and the future value of the exogenous variable. The expectation of a product of random variables leads to a covariance, specifically the covariance between the parameter estimate and the future value of the exogenous variable.[2] Since the agent uses realizations of the exogenous variable to estimate the autoregressive parameter, the covariance will not be zero.

Taking the covariance effect into account makes models more complex, but it may also make them more realistic and perhaps more successful in accounting for what empirical researchers have found in the data.[3] To illustrate this possibility, we focus on the well-known rational expectations–permanent income hypothesis model of consumption. This model has a famous implication, namely that the marginal utility of consumption should follow a random walk (or, more precisely, a martingale). A variety of empirical studies have shown that lagged income has predictive power for the first difference of consumption, at both the aggregate and the household levels, violating the martingale prediction.[4] This is known as the "excess sensitivity" puzzle because, under the rational expectations–permanent income hypothesis, households incorporate all information into their current consumption decision and lagged variables should therefore be irrelevant.[5]

The simplest example in which the covariance effect enters is in a consumption problem with a two-period horizon. In this example (which yields the martingale property under rational expectations), we show that the covariance effect will make the first difference of household consumption depend on lagged income. We use closed-form analytical results, simulations based on analytical results, and numerical solution techniques to show how the short-horizon results extend to a long-horizon problem.

The paper is organized as follows. Section 2 formally introduces the covariance effect using an asset pricing example. Section 3 presents the short-horizon consumption model and shows how the covariance effect leads to evidence of "excess sensitivity." Section 4 introduces a numerical method that allows us to extend the horizon further. Section 5 illustrates how the covariance effect enters three other well-known models (specifically, models of money demand, exchange-rate dynamics, and fixed investment). Section 6 concludes.

2. FORMAL INTRODUCTION OF THE COVARIANCE EFFECT

To illustrate some basic ideas, we begin with a simple asset pricing model. Consider the special case of the Lucas (1978) model with a single asset and risk-neutral agents. The first-order condition for the household's problem implies

$$P_{t+j} = \left(\frac{1}{1+r}\right)\hat{E}_{t+j}[P_{t+j+1} + D_{t+j+1}], \quad \mathbf{(1)}$$

where P is the price of the asset, r is the interest rate, D is the dividend on the asset, and $\hat{E}_{t+j}$ is the expectation operator conditional on information in period $t+j$. (We use the notation "$\hat{E}$" here to emphasize that we are considering the subjective expectation of an agent who is learning. Of course, under rational expectations, these subjective expectations are replaced by the mathematical expectation.) Let the timing be such that the agent receives the dividend and then trades the asset. Suppose that the agent has observed the dividend process for t periods, so that her information set at time t is $I_t = \{D_0, D_1, \ldots, D_t\}$. For simplicity, we focus on the last three periods in a finite-horizon problem, and so, the last period will be period $t+2$. Dividends are paid in each period, specifically including periods $t+1$ and $t+2$. Since $t+2$ is the terminal period, the price in that period (after the dividend has already been paid) is zero. By successive substitution in (1), it is straightforward to show that

$$P_{t+1} = \left(\frac{1}{1+r}\right)\hat{E}_{t+1}D_{t+2} \quad \mathbf{(2)}$$

and

$$P_t = \left(\frac{1}{1+r}\right)\hat{E}_t D_{t+1} + \left(\frac{1}{1+r}\right)^2 \hat{E}_t \hat{E}_{t+1} D_{t+2}. \quad \mathbf{(3)}$$

Suppose that the stochastic process for dividends is[6]

$$D_{t+1} = \rho D_t + \varepsilon_{t+1}, \quad \mathbf{(4)}$$

where ε is i.i.d. $N(0, \sigma^2)$. Under rational expectations the subjective expectation, $\hat{E}_t$, is replaced by E_t, the mathematical expectation conditional on information at time t. In equations (2) and (3), we then have $E_{t+j}D_{t+j+1} = \rho D_{t+j}$ and $E_t E_{t+1} D_{t+2} = E_t D_{t+2} = \rho^2 D_t$, so that

$$P_{t+1} = \frac{\rho}{1+r} D_{t+1} \quad \mathbf{(5)}$$

and

$$P_t = \left[\frac{\rho}{1+r} + \frac{\rho^2}{(1+r)^2}\right] D_t. \quad \mathbf{(6)}$$

Now consider what happens if learning plays a role. Suppose that ρ is unknown and that agents estimate ρ using ordinary least squares, which is the minimum

https://doi.org/10.1017/S1365100501010069 Published online by Cambridge University Press

mean squared error estimator. The expressions for P_{t+1} and P_t are again given by (2) and (3), except that the agent's expectation is no longer the mathematical expectation. This leads to a complication when we consider the expectations terms in the equation for P_t. Note first that, since D_t is known at time t,

$$\hat{E}_t D_{t+1} = \hat{E}_t(\rho D_t + \varepsilon_{t+1}) = D_t \hat{E}_t(\rho),$$

and thus

$$\hat{E}_t D_{t+1} = \hat{\rho}_t D_t, \tag{7}$$

where $\hat{E}_t(\rho) = \hat{\rho}_t$ and $\hat{\rho}_t$ denotes the current parameter estimate. Using (7), the equation for P_{t+1} gives

$$P_{t+1} = \left(\frac{\hat{\rho}_{t+1}}{1+r}\right) D_{t+1}. \tag{8}$$

In equation (3), equation (7) gives

$$\hat{E}_{t+1} D_{t+2} = \hat{\rho}_{t+1} D_{t+1}. \tag{9}$$

Using (9) in the iterated expectation term on the right-hand side of (3) then gives that $\hat{E}_t \hat{E}_{t+1} D_{t+2} = \hat{E}_t \hat{\rho}_{t+1} D_{t+1}$. However, conditional on the agent's information set at time t, $\hat{\rho}_{t+1} D_{t+1}$ is a product of random variables. The expectation of this product will be the product of the expectations of $\hat{\rho}_{t+1}$ and D_{t+1} plus a covariance term:

$$\hat{E}_t \hat{E}_{t+1} D_{t+2} = \hat{E}_t \hat{\rho}_{t+1} D_{t+1} = \hat{E}_t \hat{\rho}_{t+1} \hat{E}_t D_{t+1} + \widehat{\mathrm{Cov}}_t(\hat{\rho}_{t+1}, D_{t+1}),$$

and thus, using $\hat{E}_t \hat{\rho}_{t+1} = \hat{\rho}_t$ and (7),

$$\hat{E}_t \hat{E}_{t+1} D_{t+2} = \hat{\rho}_t^2 D_t + \widehat{\mathrm{Cov}}_t(\hat{\rho}_{t+1}, D_{t+1}), \tag{10}$$

where

$$\widehat{\mathrm{Cov}}_t(\hat{\rho}_{t+1}, D_{t+1}) \equiv \hat{E}_t[(\hat{\rho}_{t+1} - \hat{E}_t \hat{\rho}_{t+1})(D_{t+1} - \hat{E}_t D_{t+1})].$$

In general, the covariance term will not be equal to zero because a high realization of D will tend to increase the agent's estimate of ρ, creating a positive covariance.[7] Thus $\hat{E}_t \hat{E}_{t+1} D_{t+2} \neq \hat{\rho}_t^2 D_t$, and so, P_t cannot be derived from (6) by simply replacing the parameter ρ with the parameter estimate $\hat{\rho}_t$.

This is a simple example, but the issue is broad ranging. If agents are learning an autoregressive parameter in a driving process such as (4), then the expectations terms are not the same as they would be if the estimated parameter were nonstochastic: $\hat{E}_t D_{t+j+1} \neq \hat{\rho}_t^{j+1} D_t$ for $j > 0$. Nonetheless, in the learning literature, expectations typically are resolved by treating the agent's parameter estimate as nonstochastic. As noted earlier, Timmermann (1996) makes this assumption to render tractable an asset pricing model in which agents estimate the parameters of an autoregressive dividend process. Lewis (1989) studies a monetary model of exchange rates in which agents are learning the autoregressive parameters of the money supply and income processes. She resolves expectations by treating the estimated AR parameters as known constants. Also, in their influential papers on the convergence of learning to rational expectations, Marcet and Sargent (1989a,b) study an application in which firms estimate the autoregressive parameter in the process governing aggregate capital. In their application, the representative firm must evaluate expectations of the future aggregate capital stock. To resolve these expectations, Marcet and Sargent treat the estimate of the AR(1) parameter as a known constant.

A recent paper in which agents do treat their parameter estimates as random variables is Guidolin and Timmermann (2000).[8] In that paper, learning that explicitly accounts for future variation in parameter estimates is called "rational learning" and learning that treats parameter estimates as known constants is called "adaptive learning." In a version of the Lucas asset pricing model, Guidolin and Timmermann establish conditions under which rational learning will increase the expected level of asset prices and the variance of returns. They show that rational learning may generate serial correlation and volatility clustering in asset returns. Also, they argue that learning may cause a martingale test to reject the joint hypothesis of rational expectations and market efficiency.

The example in this section illustrates that a full analysis of learning introduces an effect that is missing in both rational expectations models and in models of learning that treat the estimated parameter as nonstochastic. We refer to this effect as the covariance effect. In subsequent sections, we show how the covariance effect can substantially change the predictions of an otherwise standard economic model.

3. SHORT-HORIZON CONSUMPTION MODEL

In this section, we show how the covariance effect can alter a well-known prediction of the rational expectations–permanent income hypothesis model of consumption, namely the prediction that consumption will be a martingale. We begin by setting up a simple model of consumption that yields the martingale prediction.

Let the household's problem be to choose a consumption sequence to maximize the expected present value of future utility subject to the dynamic budget constraint which governs the evolution of financial assets:

$$\max_{\{C_{t+j}\}_{j=0}^{\tau}} \hat{E}_t \sum_{j=0}^{\tau} \beta^j U(C_{t+j}) \tag{11}$$

so that

$$A_{t+1} = R(A_t + Y_t - C_t), \tag{12}$$

where β is the discount factor, U is the utility function, C is consumption, A is assets, R is the gross rate of return on the single, risk-free asset, and Y is labor income. For notational convenience, we assume $R = 1/\beta$. We assume quadratic utility because this yields closed-form solutions for consumption, which allow us

https://doi.org/10.1017/S1365100501010069 Published online by Cambridge University Press

to isolate the effects of learning. In addition, quadratic utility implies a coefficient of relative prudence[9] of zero and thus avoids introducing a precautionary savings motive, which is absent from the standard rational expectations–permanent income hypothesis formulation. Since the terminal period is $t+\tau$ and the current period is t, τ represents the agent's horizon.

The solution to this consumption problem will be a set of consumption rules for each period from t to $t+\tau$:

$$C_{t+\tau} = A_{t+\tau} + Y_{t+\tau}, \tag{13}$$

$$C_{t+\tau-1} = \frac{R}{1+R}\left[A_{t+\tau-1} + Y_{t+\tau-1} + \frac{1}{R}\hat{E}_{t+\tau-1}Y_{t+\tau}\right], \tag{14}$$

and, in general,[10]

$$C_t = \frac{R^\tau}{\sum_{i=0}^{\tau} R^i}\left[A_t + Y_t + \frac{1}{R}\hat{E}_t Y_{t+1} + \frac{1}{R^2}\hat{E}_t Y_{t+2} + \cdots + \frac{1}{R^\tau}\hat{E}_t Y_{t+\tau}\right]. \tag{15}$$

Under rational expectations, consumption in the model described above would be a martingale. The intuition is simple: Under rational expectations, the agent's expectation of a variable is the same as the mathematical expectation of that variable conditional on the information set at the time the expectation is formed. Consumption changes from one period to the next only if there is news. Based on current information, the best estimate of next period's consumption is today's consumption.

Empirically, consumption does not appear to be a martingale. In particular, many empirical studies regress the first difference of log consumption on lagged income[11]:

$$\ln C_{t+1} - \ln C_t = a_0 + a_1 \ln Y_t + u_{t+1}. \tag{16}$$

The typical finding is that the coefficient on lagged income is not zero, as predicted by the rational expectations–permanent income hypothesis, but instead negative and significantly different from zero. This is known as the excess sensitivity puzzle because consumption is more sensitive to lagged income than it should be under the rational expectations–permanent income hypothesis.

To explore whether this could be due to learning, consider the following stochastic process for labor income:

$$y_{t+1} = \rho y_t + \varepsilon_{t+1} \tag{17}$$

under the assumption that ε is i.i.d. $N(0, \sigma^2)$.[12] Here, y_t represents the deviation of income from its unconditional mean; that is, $y_t \equiv Y_t - \bar{y}$. We assume that the agent knows the form of the stochastic process for labor income, the unconditional mean of Y, and the variance of ε but does not know ρ.[13] More formally, the agent's information set is

$$I_t \equiv \left\{y_t, y_{t-1}, \ldots, y_0; \bar{y}, \sigma^2\right\}. \tag{18}$$

This means that, at the beginning of period t, the agent has a sample of size t with which to estimate ρ. The agent estimates ρ using ordinary least squares; thus,

$$\hat{\rho}_t = \left(\sum_{i=1}^{t} y_{i-1} y_i\right) \Big/ \left(\sum_{i=1}^{t} y_{i-1}^2\right). \qquad \textbf{(19)}$$

The assumption of OLS is attractive both because of its simplicity and because the linear projection of y on lagged y yields the MMSE forecast.[14] Moreover, if the agent begins with a diffuse prior, the OLS estimate of ρ is the Bayesian estimate. That is, $\hat{\rho}_t$ is the mean of the distribution of ρ, posterior to the information set I_t.[15]

3.1. Two-Period Horizon

The covariance effect arises when agents take the expectation of a product of random variables. Therefore, the first place the covariance effect will show up is in the forecast of y_{t+2}. We therefore set τ equal to 2 for the moment. Using (7), (10), (14), and (15), we can derive the first difference of consumption under learning[16]:

$$(C_{t+1} - C_t) = \frac{R}{1+R}\left(1 + \frac{\hat{\rho}_{t+1}}{R}\right)\varepsilon_{t+1} + \frac{R}{1+R}\left[\left(1 + \frac{\hat{\rho}_{t+1}}{R}\right)\rho y_t - \left(1 + \frac{\hat{\rho}_t}{R}\right)\hat{\rho}_t y_t\right] - \left(\frac{1}{1+R}\right)\hat{\sigma}_{\rho y}(t), \qquad \textbf{(20)}$$

where

$$\hat{\sigma}_{\rho y}(t) \equiv \widehat{\mathrm{Cov}}_t(\hat{\rho}_{t+1}, y_{t+1}). \qquad \textbf{(21)}$$

The first term on the right-hand side of (20) represents the annuitized present value of a shock to labor income. In the absence of learning, the estimated value of ρ would be replaced with the true value of ρ, $\hat{\sigma}_{\rho y}$ would equal zero, and this would be the only term. Then, consumption next period would be equal to consumption this period plus a disturbance orthogonal to this period's information set. Thus consumption would be a martingale as suggested by the rational expectations intuition presented earlier. In particular, in the absence of learning, a regression of the first difference of consumption on lagged income should yield a coefficient of zero since ε_{t+1} is orthogonal to y_t.

The second term is also an annuitized present value, specifically one that arises from changes in expected future income due to revisions in the agent's estimate of ρ. Under rational expectations, ρ is known and, therefore, new information does not alter the agent's beliefs about ρ. Consequently, under rational expectations the second term does not arise.

The third term reflects the covariance effect. It appears because high realizations of labor income induce the agent to revise the estimate of ρ upward, leading to higher expectations of future income and thus higher consumption in the current period (compared to the rational expectations–permanent income hypothesis benchmark). An increase in current consumption means a smaller first difference of consumption (smaller $C_{t+1} - C_t$) and thus a regression of the first difference of consumption on lagged income should yield a negative coefficient on lagged income.

More formally, it is possible to show that the covariance effect tends to make the first difference of consumption a negative function of lagged income using the following proposition, which is proved in the Appendix.

PROPOSITION 1. *Under learning, the expectation of the covariance between* $\hat{\rho}_{t+1}$ *and* y_{t+1} *conditional on current information is an increasing function of* y_t, *where the coefficient on* y_t *is* $\sigma^2 / \sum_0^{t-1} y_t^2$:

$$\hat{\sigma}_{\rho y}(t) = \frac{\sigma^2}{\sum_{i=0}^{t-1} y_i^2} y_t. \tag{22}$$

Since $\hat{\sigma}_{\rho y}$ enters the right-hand side of (20) with a negative coefficient, the covariance effect will tend to make $(C_{t+1} - C_t)$ negatively correlated with y_t.

3.2. Three-Period Horizon

The simple case of $\tau = 2$ shows that the covariance effect could account for empirical evidence of excess sensitivity of consumption to lagged income. Whether this generalizes as we lengthen the horizon turns out to be a surprisingly complicated question. We begin by considering the case of $\tau = 3$. Now, from equation (15), we see that C_t depends on $\hat{E}_t y_{t+3}$. Since the law of iterated expectations gives $\hat{E}_t y_{t+3} = \hat{E}_t \hat{E}_{t+1} \hat{E}_{t+2} y_{t+3}$, we can use (10) to write

$$\hat{E}_t y_{t+3} = \hat{E}_t \left(\hat{\rho}_{t+1}^2 y_{t+1}\right) + \hat{E}_t \widehat{\mathrm{Cov}}_{t+1}(\hat{\rho}_{t+2}, y_{t+2}). \tag{23}$$

This expression is harder to evaluate because it involves taking the expectation of a more complicated nonlinear function of random variables. In particular, compared to the case of $\tau = 2$, we now must take the expectation of the functions $\hat{\rho}_{t+1}^2 y_{t+1}$ and $\widehat{\mathrm{Cov}}_{t+1}(\hat{\rho}_{t+2}, y_{t+2})$ rather than taking the expectation of the simple product of $\hat{\rho}_{t+1}$ and y_{t+1}.

One approach is to use a second-order Taylor-series expansion to obtain the following proposition, which is proved in the Appendix.[17]

PROPOSITION 2. *Under learning, the expectation of* $\hat{\rho}_{t+1}^j y_{t+1}$ *conditional on current information is equal to* $\hat{\rho}_t^{j+1} y_t$ *plus a term that depends on* $\hat{\sigma}_{\rho y}(t)$ *and a term that depends on the estimated variance of* $\hat{\rho}_{t+1}$:

$$\hat{E}_t \hat{\rho}_{t+1}^j y_{t+1} \cong \hat{\rho}_t^{j+1} y_t + j \hat{\rho}_t^{j-1} \hat{\sigma}_{\rho y}(t) + \tfrac{1}{2} j(j-1) \hat{\rho}_t^{j-1} y_t \hat{\sigma}_\rho^2(t), \tag{24}$$

where

$$\hat{\sigma}_\rho^2(t) \equiv \hat{E}_t\left[(\hat{\rho}_{t+1} - \hat{E}_t \hat{\rho}_{t+1})^2\right] = \hat{E}_t\left[(\hat{\rho}_{t+1} - \hat{\rho}_t)^2\right].$$

The next proposition (proved in the Appendix) establishes that $\hat{\sigma}_\rho^2(t)$ is an increasing function of $\hat{\sigma}_{\rho y}(t)$.

PROPOSITION 3. *Under learning, the estimated variance of* $\hat{\rho}_{t+1}$ *is equal to the estimated covariance of* $\hat{\rho}_{t+1}$ *with* y_{t+1} *multiplied by the ratio of current income to* $\sum_0^t y_t^2$:

$$\hat{\sigma}_\rho^2(t) = \frac{y_t}{\sum_0^t y_i^2}\hat{\sigma}_{\rho y}(t).$$

Using (15), (23), and Propositions 2 and 3 to solve for $C_{t+1} - C_t$ yields[18]

$$(C_{t+1} - C_t) = \left(\frac{R^2}{1+R+R^2}\right)\left\{\left(1 + \frac{\hat{\rho}_{t+1}}{R} + \frac{\hat{\rho}_{t+1}^2}{R^2}\right)\varepsilon_{t+1} + \left[\left(1 + \frac{\hat{\rho}_{t+1}}{R} + \frac{\hat{\rho}_{t+1}^2}{R^2}\right)\rho - \left(1 + \frac{\hat{\rho}_t}{R} + \frac{\hat{\rho}_t^2}{R^2}\right)\hat{\rho}_t\right]y_t + \frac{1}{R^2}\left[\hat{\sigma}_{\rho y(t+1)} - \hat{E}_t\hat{\sigma}_{\rho y(t+1)}\right] - \left(\frac{1}{R} + \frac{2\hat{\rho}_t}{R^2} + \frac{\hat{\rho}_t y_t^2}{R^2\sum_{i=0}^t y_i^2}\right)\hat{\sigma}_{\rho y}(t)\right\}. \quad \textbf{(25)}$$

The terms in this expression are similar to those in the corresponding expression for $\tau = 2$. The annuitization factor $[R^2/(1 + R + R^2)]$ is smaller because the agent is spreading the effects of any news over more periods. (To see this, note that R is approximately 1, and so, the annuitization factor here is approximately one-third, since there are three periods left.) In the $\tau = 2$ case, there was a term arising from revisions in the agent's estimate of ρ. That term (the second term above) is still there, but there is now also a term (the third term above) arising from revisions in the estimated covariance between $\hat{\rho}$ and y.

The most interesting term is the covariance term (the last term above). Whether the covariance effect is stronger as the horizon lengthens from $\tau = 2$ to $\tau = 3$ depends on the coefficient on the covariance term, which is the product of the annuitization factor and an additional expression in parentheses. The magnitude of the expression in parentheses depends on R, the estimated value of ρ, and the ratio of squared income to the sum of squared income. R is close to 1 and reasonable variations will have only a modest effect on the expression. On average, the order of magnitude of the ratio of squared income to the sum of squared income will be approximately $1/t$. The estimate of ρ has the most effect on the expression. For example, for $R = 1.05$, $t = 40$, and estimated $\rho = 0.9$, the product of the annuitization factor and the expression in parentheses will be 0.91. For estimated $\rho = 0.7$ and $\rho = 0.5$, the product of the annuitization factor and the expression in parentheses will be 0.78 and 0.65, respectively. In comparison, the coefficient on the covariance for the $\tau = 2$ case in equation (20) is 0.49. Thus, for a wide range of parameter values, the coefficient on the covariance is higher in the $\tau = 3$ case than in the $\tau = 2$ case.

In analyzing the covariance effect, the simplest case is consumption close to the terminal period (e.g., $\tau = 2$), but the results above suggest that the covariance effect will not disappear in a longer-horizon problem. If anything, the foregoing analysis suggests that the covariance effect may become even stronger as the horizon lengthens. There is an intuitive explanation for a stronger effect at longer horizons. If the horizon is extended, changes in tomorrow's parameter estimates will matter for more future periods.

The analytical results show that learning and, in particular, the covariance effect move a standard permanent income hypothesis model in the right direction by making theoretical consumption behavior qualitatively more similar to actual consumption behavior. To examine the quantitative impact of learning, we turn to simulations based on the analytical results above. We set $\bar{y} = 100$, $A_t = 25$, $R = 1.05$, $\sigma = 1$, and we vary ρ. We generate a sequence of observations, y_i, $i = 0, 1, \ldots, t$, where y_0 is drawn from the unconditional distribution of y, and where y_1 through y_t are generated using equation (17). We use these observations to estimate ρ via OLS. Having obtained $\hat{\rho}_t$ and y_t, we obtain C_t from (15), (10), (23), and Propositions 1–3. We then use (17) and (12) to obtain y_{t+1}, and A_{t+1}. Equations (15) and (10) and Proposition 1 are then used to obtain C_{t+1}. We thus have the single observation $[C_{t+1}, C_t, y_t]$. We repeat this procedure 10,000 times, generating a sample of 10,000 observations of $[C_{t+1}, C_t, y_t]$. This sample represents a cross section of consumers, each of whom has observed her own realization of the income sequence.[19] Next, we use this sample to estimate (16). We report results for the models with $\tau = 2$ and $\tau = 3$ in Table 1. We consider four different values for ρ ($\rho = 0.7$, 0.9, 0.95, and 0.99) and several values of t.[20] Table 1 clearly shows that learning can generate evidence of excess sensitivity. All of the coefficients on lagged income are negative and most are significantly different from zero. To assess the economic significance, it is useful to note that the coefficients on lagged income are on the same order of magnitude as found, for example, by Zeldes (1989).

Although the covariance effect decreases as agents have more observations on which to base their estimate of the unknown parameter (i.e., as t increases), it still has an effect on consumption even when t is quite large. For example, in all cases the simulations in Table 1 show that the coefficient on lagged income is significantly different from zero when agents have 30 years of observations of their income process.

Another interesting point emerges from Table 1. The covariance effect seems to become stronger as the horizon lengthens. This is illustrated by the absolute magnitude of the coefficients on lagged income, which generally tends to increase as we move from $\tau = 2$ to $\tau = 3$. This reinforces the impression from the analytical results that the covariance effect is not simply an artifact of a short-horizon problem.

Proposition 1 shows that the covariance effect will tend to decrease with t (since the sum of y_i^2 from $i = 0$ to $i = t - 1$ will be increasing in t). The variable t represents the number of observations of the income process that the agent has,

Table 1. Regressions of simulated consumption growth under full learning on lagged income: Short-horizon cases

Household's years of previous observations (t)	Horizon $\tau = 2$: $\hat{\alpha}_1$ (standard error)	$\tau = 2$: t-statistic	Horizon $\tau = 3$: $\hat{\alpha}_1$ (standard error)	$\tau = 3$: t-statistic
(A) $\rho = 0.7$				
15	−0.050 (0.006)	−8.7	−0.047 (0.005)	−8.8
30	−0.021 (0.006)	−3.6	−0.021 (0.005)	−4.2
60	−0.003 (0.006)	−0.6	−0.010 (0.005)	−1.9
(B) $\rho = 0.9$				
15	−0.049 (0.004)	−12.3	−0.055 (0.004)	−14.0
30	−0.023 (0.004)	−6.1	−0.024 (0.004)	−6.4
60	−0.006 (0.004)	−1.7	−0.013 (0.004)	−3.4
(C) $\rho = 0.95$				
15	−0.030 (0.003)	−10.2	−0.041 (0.003)	−13.8
30	−0.016 (0.003)	−5.7	−0.015 (0.003)	−5.3
60	−0.008 (0.003)	−3.0	−0.005 (0.003)	−1.7
(D) $\rho = 0.99$				
15	−0.011 (0.001)	−8.1	−0.016 (0.001)	−11.1
30	−0.010 (0.001)	−7.3	−0.013 (0.001)	−9.5
60	−0.006 (0.001)	−4.8	−0.009 (0.001)	−7.1

and so, t is most naturally interpreted as the number of years since the agent began her working life. Thus Proposition 1 implies that the coefficient on lagged income should decline with age. To the best of our knowledge, no one has tested whether the coefficient on lagged income is more negative for people early in their

https://doi.org/10.1017/S1365100501010069 Published online by Cambridge University Press

working life than for people later in their working life. There is some indirect evidence on this point, however, in Zeldes (1989). Zeldes compares households with high and low levels of assets relative to income. The ratio of assets to income is likely to be strongly correlated with age. Zeldes finds that the coefficient on lagged income is typically more negative for low-asset households, a result that is consistent with what we would expect if the covariance effect was playing some role in determining consumption. Of course, an alternative possibility is that some households (particularly low-assets households) may be liquidity constrained. If the covariance effect plays a separate role, then the coefficient on lagged income should be more negative for younger agents (after controlling for liquidity constraints). It would be interesting to see if this implication is supported by the data.

3.3. Asymptotic Convergence and Bounded Memory

Intuitively, we would expect the covariance effect to disappear as the number of observations increases because the estimated value of ρ will converge to the true value as the sample size grows. This is confirmed by the simulation results and by Proposition 1, which shows that the denominator of the coefficient on y_t increases with the sample size t. More formally, since the law of motion for y_t, equation (17), is independent of the agent's beliefs, learning ρ is an example of the estimation of a time-invariant linear stochastic difference equation. As such we can apply the result of Marcet and Sargent (1989b, p. 354), which gives $\hat{\rho}_t \to \rho$ almost surely as $t \to \infty$. However, although it converges asymptotically, the least-squares estimator of an AR(1) parameter shows considerable variation even in moderately large samples. Nankervis and Savin (1988, Table 2) show that, if the true ρ is 0.9, the standard deviation of the least-squares estimator is 0.31 for a sample of size 30, 0.09 for a sample of size 50, and 0.06 for a sample of size 100. Thus, the covariance effect can be important for quite large samples.[21]

Although our results are not invariant to the interpretation of the time period, the most natural choice for the length of the period in a model of learning about income is 1 year, since this is the frequency at which many agents receive news (about raises, promotions, the renegotiation of contracts, etc.). Under this interpretation, substantial uncertainty about the true parameter is likely to exist throughout the typical agent's working life. In this sense the life-cycle model of consumption provides a natural example of learning with bounded memory. Specifically, the length of an agent's memory is limited to the past years of his or her working life. Since we treat all past income observations as being equally relevant, the life-cycle hypothesis provides an economic rationale for giving equal weight to a finite number of observations, one of the types of bounded memory analyzed by Honkapohja and Mitra (2000).

The life-cycle consumption case assumes agents have only a finite number of observations. However, learning may also be important asymptotically. For example, in an infinite-horizon model of consumption, an occasional break in the structural parameters of the income process may cause agents to discard or discount observations from the distant past. This economic justification for learning with bounded memory is by no means limited to models of consumption. For example, we have already shown, in Section 2, that the covariance effect will arise in a present-value model of stock prices with learning. Timmermann (1996, p. 533) provides the following argument to support learning with bounded memory in an infinite horizon version of the present-value model:

> Learning effects are unlikely to vanish even in the long run .... After a structural break in the data-generating process it is reasonable to assume that agents will attempt to re-estimate the dividend process using relatively recent data points rather than the entire historical sample. Similarly, slowly changing nonstationarities in the data-generating process could be handled by agents' use of a moving window of the data, in which case learning effects will persist even asymptotically.

In another example, we show in Section 5 that the covariance effect will arise in a monetary model of exchange rates. Lewis (1989, p. 92) points out a potential advantage to bounded memory following a structural break in the parameters of that model. Analyzing the change in U.S. money demand that occurred at the beginning of the 1980's, she states that a market using a diffuse prior, which represents no confidence in previous estimates of money demand parameters, will "on-average learn more quickly and make less systematic prediction errors than a market using a prior based on the past."

4. EXTENDING THE HORIZON FURTHER

As τ increases, the expressions for consumption become more cumbersome. To extend the horizon further, we therefore develop a numerical method that does not require that we resolve analytically the expectations of complicated nonlinear functions of random variables.

There are some unique features of our problem that make it particularly challenging. Because we are attempting to capture the agent's subjective expectations, in modeling those expectations we must use only information available to the agent. This means that we need to find functions that express these expectations in terms of the agent's current information set.

Let I_t be given by equation (18). Since each of the expectations that we consider is an expectation conditional on I_t, each could, in principle, be written as an explicit function of I_t. The object of the method is to find a polynomial function of I_t to approximate each of these conditional expectations. Since we wish to evaluate these functions for $t = 15$, 30, and 60, we cannot form polynomials in I_t directly; its dimension is too large. For example, with $t = 15$, the set of second-order polynomials in I_t has 136 elements. We must therefore find a compact way of summarizing I_t. We define the state vector, $\boldsymbol{s_t} = [y_t, \hat{\rho}_t, M_t^{-1}]'$, where

$$M_t \equiv \frac{1}{t+1}\sum_{i=0}^{t} y_i^2,$$

and form polynomials in its elements. Note that y_t is the exogenous state variable, $\hat{\rho}_t$ is the agent's point estimate, and M_t^{-1} is proportional to the variance of the agent's parameter estimate. Thus, $\boldsymbol{s_t}$ expands the state variable to include the mean and variance of the agent's beliefs [(see equation (32)].

We use this vector to summarize I_t for two reasons. First, having only three elements, its dimension is manageable—the set of second-order polynomials in $\boldsymbol{s_t}$ has 10 elements. Second, we can write next period's state vector, $\boldsymbol{s_{t+1}} = [y_{t+1}, \hat{\rho}_{t+1}, M_{t+1}^{-1}]'$, as a function of $\boldsymbol{s_t}$ and of random variables whose distribution is known to the agent. Specifically, using the recursive representation of OLS, we have

$$y_{t+1} = \hat{\rho}_t y_t + \nu_{t+1}, \quad \textbf{(26)}$$

$$\hat{\rho}_{t+1} = \hat{\rho}_t + \frac{1}{t+1} M_t^{-1} y_t (y_{t+1} - \hat{\rho}_t y_t), \quad \textbf{(27)}$$

and

$$M_{t+1} = M_t + \left(\frac{1}{t+2}\right)\left(y_{t+1}^2 - M_t\right), \quad \textbf{(28)}$$

where $\nu_{t+1} = (y_{t+1} - \hat{\rho}_t y_t) = (\rho - \hat{\rho}_t) y_t + \varepsilon_{t+1}$. Note that

$$(\rho - \hat{\rho}_t) = \left(-\sum_1^t y_{i-1}\varepsilon_i \bigg/ \sum_1^t y_{i-1}^2\right). \quad \textbf{(29)}$$

Since the ε_i are i.i.d. $N(0, \sigma^2)$, ν_{t+1} is $N(0, \sigma_\nu^2)$, where

$$\sigma_\nu^2 = \left[1 + \left(\frac{1}{\sum_1^t y_{i-1}^2}\right)\right]\sigma^2. \quad \textbf{(30)}$$

To keep the dimension of the polynomial set manageably small, we form the approximating functions from the complete set of polynomials of total degree k, which is described by Judd (1998, pp. 239–240). Briefly, this is the set of polynomial terms that would appear in the kth-order Taylor-series expansion. Denote the elements of this set by $\{\phi_i(\boldsymbol{s_t})\}$ for $i = 1, 2, \ldots, N$, where N is the number of elements in the polynomial set. For example, the set of second-order polynomials in $\boldsymbol{s_t}$ is $\{1, y_t, \hat{\rho}_t, M_t^{-1}, y_t^2, \hat{\rho}_t y_t, M_t^{-1} y_t, \hat{\rho}_t^2, \hat{\rho}_t M_t^{-1}, M_t^{-2}\}$. Thus, $\phi_1(\boldsymbol{s_t}) = 1$, $\phi_2(\boldsymbol{s_t}) = y_t$, $\phi_3(\boldsymbol{s_t}) = \hat{\rho}_t, \ldots, \phi_N(\boldsymbol{s_t}) = M_t^{-2}$. The results we report below are based on this second-order set. Define the polynomial function $\psi(\boldsymbol{a}_j, \boldsymbol{s_t})$ as follows:

$$\psi(\boldsymbol{a}_j, \boldsymbol{s_t}) \equiv \sum_{i=1}^N a_{ji}\, \phi_i(\boldsymbol{s_t}). \quad \textbf{(31)}$$

The approximation method finds a vector of coefficients, $\boldsymbol{a}_j = [a_{j1}, a_{j2}, \ldots, a_{jN}]'$ such that $\psi(\boldsymbol{a}_j, \boldsymbol{s_t})$ is a good approximation to $\hat{E}_t y_{t+j}$ for $j = 1, 2, \ldots, \tau$.

https://doi.org/10.1017/S1365100501010069 Published online by Cambridge University Press

We use a Monte Carlo method[22] to fix the parameters $\boldsymbol{a}_j$. Since we seek to approximate the agent's expectations, we perform the method using the stochastic process for y_t, *as it is perceived by the agent*. The method can be described in five steps.

Step 1. Generate the agent's information set using the true stochastic process as follows: Draw y_0 from the unconditional distribution for y_t, which is $N[0, \sigma^2/(1-\rho^2)]$, and produce $\{y_1, y_2, \ldots, y_t\}$ using equation (17).

Step 2. Using this information set, $I_t = \{y_t, y_{t-1}, \ldots, y_0\}$, compute the current state and construct the agent's posterior probability density[23] for ρ. We assume that the agent's prior belief about ρ is diffuse, and so, the posterior distribution for ρ is[24]

$$N\left[\hat{\rho}_t, \sigma^2\left(\sum_0^{t-1} y_i^2\right)^{-1}\right]. \tag{32}$$

In the next three steps the Monte Carlo procedure is performed from the point of view of the agent who does not know ρ but has observed I_t.

Step 3. Draw a value of ρ, ρ^A, from the agent's posterior distribution for ρ. If the draw of ρ^A exceeds the discount rate, R, discard that ρ^A and draw again. This truncates the normal distribution in (32) at the upper bound R and redistributes the probability mass to draws below R.[25] Set y_0^A equal to the y_0 obtained in Step 1 and use ρ^A to generate a time series of observations, $\{y_0^A, y_1^A, \ldots, y_t^A, y_{t+1}^A, \ldots, y_{t+\tau-1}^A, y_{t+\tau}^A\}$ according to $y_{t+1}^A = \rho^A y_t^A + \varepsilon_{t+1}$, where ε_{t+1} is i.i.d. $N(0, \sigma^2)$.

Step 4. Repeat Step 3 a large number, N_R, of times. (The results reported below use $N_R = 500$.) This gives an artificially generated sample of N_R observations that the agent can use to fix the $\boldsymbol{a}_j$ coefficients.

Step 5. Regress the N_R realizations of y_{t+1}^A, on the polynomials $\phi_i(s_t^A)$ to obtain the coefficient vector, $\boldsymbol{a}_1$. Repeat, regressing the N_R realizations of y_{t+j}^A on the polynomials $\phi_i(s_t^A)$ to obtain $\boldsymbol{a}_j$, for $j = 2, 3, \ldots, \tau$. Having determined these coefficient vectors, we can construct the polynomial approximations, which use the true state from Step 2:

$$\hat{E}_t y_{t+1} \approx \psi(a_1, \boldsymbol{s_t}), \hat{E}_t y_{t+2} \approx \psi(a_2, \boldsymbol{s_t}), \ldots, \hat{E}_t y_{t+\tau} \approx \psi(a_\tau, \boldsymbol{s_t}).$$

This method produces a set of expectations, $\hat{E}_t y_{t+1}$ through $\hat{E}_t y_{t+\tau}$, for the agent who has observed a single realization of $[y_t, \ldots, y_0]$. These expectations are then used in the decision rule, equation (15), to obtain C_t. Next, the true stochastic process for y_t, equation (17), is used to generate $\boldsymbol{s_{t+1}}$. The agent's expectations in period $t+1$ are obtained from

$$\hat{E}_{t+1}\, y_{t+1+j} \approx \psi(\boldsymbol{a}_j, \boldsymbol{s_{t+1}}), \quad j = 1, 2, \ldots, \tau - 1. \tag{33}$$

These expectations are used in the decision rule (15) to determine C_{t+1}.

TABLE 2. Regressions of simulated consumption growth under full learning on lagged income: Long-horizon cases

Household's years of previous observations (t)	Horizon $\tau = 5$: $\hat{\alpha}_1$ (std. err.)	Horizon $\tau = 5$: t-statistic	Horizon $\tau = 10$: $\hat{\alpha}_1$ (std. err.)	Horizon $\tau = 10$: t-statistic	Horizon $\tau = 15$: $\hat{\alpha}_1$ (std. err.)	Horizon $\tau = 15$: t-statistic
(A) $\rho = 0.7$						
15	−0.072 (0.007)	−10.6	−0.102 (0.006)	−16.1	−0.130 (0.008)	−15.3
30	−0.027 (0.005)	−5.8	−0.033 (0.004)	−8.1	−0.038 (0.004)	−9.8
60	−0.013 (0.004)	−3.0	−0.016 (0.003)	−4.8	−0.015 (0.003)	−5.58
(B) $\rho = 0.9$						
15	−0.070 (0.005)	−14.9	−0.107 (0.006)	−17.9	−0.152 (0.008)	−18.8
30	−0.036 (0.004)	−8.8	−0.047 (0.004)	−10.6	−0.059 (0.005)	−12.3
60	−0.016 (0.004)	−4.2	−0.024 (0.004)	−6.1	−0.030 (0.004)	−7.7
(C) $\rho = 0.95$						
15	−0.043 (0.004)	−12.3	−0.071 (0.005)	−13.8	−0.102 (0.006)	−15.8
30	−0.030 (0.003)	−9.0	−0.039 (0.004)	−10.3	−0.049 (0.004)	−11.6
60	−0.010 (0.003)	−3.4	−0.017 (0.003)	−5.0	−0.022 (0.004)	−6.3
(D) $\rho = 0.99$						
15	−0.018 (0.002)	−9.5	−0.027 (0.003)	−8.1	−0.037 (0.005)	−7.7
30	−0.014 (0.002)	−8.8	−0.018 (0.002)	−9.4	−0.022 (0.002)	−9.9
60	−0.007 (0.002)	−4.8	−0.009 (0.002)	−4.9	−0.010 (0.002)	−5.2

For a given τ, the foregoing procedure yields a single observation of C_{t+1}, C_t, and Y_t. We repeat this procedure 5,000 times to create a simulated cross-sectional data set.[26]

Results are presented in Table 2 for $\tau = 5$, $\tau = 10$, and $\tau = 15$.[27] The results are qualitatively similar to those in the simulations of the short-horizon cases

https://doi.org/10.1017/S1365100501010069 Published online by Cambridge University Press

presented in Table 1. There is strong evidence of excess sensitivity: All of the estimated coefficients on lagged income are negative and significantly different from zero. As expected, the magnitude of the covariance effect (as measured by the coefficient on lagged income) declines as agents have more observations on which to base their estimate of ρ. If anything, however, the evidence of excess sensitivity is stronger for the long-horizon cases presented here than for the short-horizon cases in Table 1. As a result, there is significant evidence of excess sensitivity even when agents have 60 years of observations of their income process.

5. COVARIANCE EFFECT IN OTHER MODELS

We have shown that, under learning, expectations will include covariance terms that can help to explain the behavior of consumption. Here we show that our results have a broad range of possible applications. Specifically, we show that our results apply to three very widely used macroeconomic models: the Cagan model of money demand, a monetary model of exchange-rate dynamics, and a quadratic-linear model of investment under uncertainty.

The following form is common to the three examples that follow: an equilibrium price or quantity, z_t, obeys

$$z_t = a\hat{E}_t z_{t+1} + bx_t, \tag{34}$$

where

$$x_{t+1} = \rho x_t + \upsilon_{t+1}. \tag{35}$$

Here a, b, and ρ are constants (possibly matrices) and υ_{t+1} is i.i.d. $N(0, \sigma_\upsilon^2)$. Repeated substitution for z_{t+j} in (34) gives

$$z_t = bx_t + ba\hat{E}_t x_{t+1} + ba^2\hat{E}_t x_{t+2} + \cdots \tag{36}$$

Thus, z_t is determined by expectations of x_{t+j}. If ρ is estimated, the expectations $\hat{E}_t x_{t+j}$, for $j = 2, 3, \ldots$, will include covariance terms that do not appear under rational expectations, and z_t will display a correlation with the forcing variable that appears excessive under the assumption of rational expectations.

Consider first the Cagan money demand function,

$$\frac{M_t^d}{P_t} = \exp\left[-\alpha\left(\frac{\hat{E}_t P_{t+1} - P_t}{P_t}\right)\right]. \tag{37}$$

Let lowercase letters denote logs. Use $\hat{E}_t p_{t+1} - p_t \approx (\hat{E}_t P_{t+1} - P_t)/P_t$ and $m_t^d = m_t^s$ to obtain $m_t^s - p_t = -\alpha(\hat{E}_t p_{t+1} - p_t)$. Rearranging, we have[28]

$$p_t = a\hat{E}_t p_{t+1} + bm_t^s, \tag{38}$$

where $a \equiv \alpha/(1+\alpha)$ and $b \equiv 1/(1+\alpha)$. If m_t^s follows an AR(1) process,

$$m_{t+1}^s = \rho m_t^s + \varepsilon_{t+1} \tag{39}$$

then (38) and (39) are of the same form as (34) and (35).

Next consider a monetary model of exchange-rate dynamics.[29] Meese (1986) and Lewis (1989) study a version of the monetary model[30] which uses the following money demand function:

$$m_t - p_t = \delta_1 y_t - \delta_2\left(i_t^u - i_t^f\right). \tag{40}$$

Here, m_t, p_t, and y_t denote the logs of relative (U.S. to foreign) money supplies, price levels, and national incomes (so that $p_t \equiv p_t^u - p_t^f$, etc.). Letting e_t denote the log of the dollar price of foreign exchange, purchasing power parity[31] gives $e_t = p_t$. Use this together with interest parity, $i_t^u - i_t^f = \hat{E}_t(e_{t+1}) - e_t$, in (40) to derive $e_t = m_t - \delta_1 y_t + \delta_2(\hat{E}_t e_{t+1} - e_t)$. Rearranging gives

$$e_t = a\hat{E}_t e_{t+1} + bw_t, \tag{41}$$

where $a = \delta^2/(1+\delta_2)$, $b = [1/(1+\delta_2)\ \ -\delta_1/(1+\delta_2)]$, and $w_t = [m_t\ \ y_t]^T$. Thus, repeated substitution for the exchange rate on the right-hand side of (41) would give e_t as a function of expectations of future values of the money supply and income. If the money supply and income are autoregressive processes,

$$\begin{bmatrix} m_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} \rho_m & 0 \\ 0 & \rho_y \end{bmatrix} \begin{bmatrix} m_t \\ y_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{mt+1} \\ \varepsilon_{yt+1} \end{bmatrix},$$

where ε_{mt+1} and ε_{yt+1} are uncorrelated white-noise shocks, we again have a model of the same form as (34) and (35). If agents are learning ρ_m and/or ρ_y, the equilibrium exchange rate will depend on terms like $\widehat{\text{Cov}}_t(\hat{\rho}_{mt+1}, m_{t+1})$ and $\widehat{\text{Cov}}_t(\hat{\rho}_{yt+1}, y_{t+1})$.

Finally, in their paper on the convergence of learning, Marcet and Sargent (1989b) study a quadratic-linear model of investment under uncertainty.[32] In that model, the representative firm chooses a sequence of capital inputs, $\{k_j\}_{j=0}^{\infty}$, to maximize

$$\hat{E}_0 \sum_{t=0}^{\infty} \beta^t \left[p_t f k_t - \gamma_t k_t - \left(\frac{d_1}{2}\right)(k_t - k_{t-1})^2 \right]. \tag{42}$$

Here $0 < \beta < 1$, f and d_1 are positive constants, γ_t is the price of capital inputs, and p_t is the price of output. Because it faces an adjustment cost [represented by the quadratic term in (42)], the optimizing firm must forecast future values of the output price. Marcet and Sargent (1989b) use $p_t = -d_2 f K_t$, where K_t denotes the aggregate or industrywide stock of capital and d_2 is a positive constant.[33] Thus, to forecast the output price the firm must forecast K_{t+j}. Marcet and Sargent (1989b) assume the firm believes that K_t follows

$$K_{t+1} = \rho K_t + \omega_{t+1} \quad \omega_{t+1} \text{ is i.i.d.} \left(0, \sigma_\omega^2\right) \tag{43}$$

and that it estimates ρ using OLS.[34] They show that, when the model converges to rational expectations, the belief in (43) is correct.

To simplify, set $\gamma_t \equiv 0$ for all t. The Euler equation for the problem of maximizing (42) is

$$\beta \hat{E}_t k_{t+1} - (\beta + 1)k_t + k_{t-1} = \frac{f^2 d_2}{d_1} K_t. \quad \textbf{(44)}$$

Using $\Delta k_t \equiv k_t - k_{t-1}$, (44) becomes

$$\Delta k_t = \beta \hat{E}_t \Delta k_{t+1} + b K_t, \quad \textbf{(45)}$$

where

$$b = -\left(\frac{f^2 d_2}{d_1}\right),$$

and the model (45) and (43) is in the form of (34) and (35).

The problem that we address in this paper is not limited to models of the same general form as (34) and (35) and, even within that form, the three examples we present are certainly not exhaustive. Still, they suggest the broad range of models in which equilibrium depends on the expectation of future values of an autoregressive variable and in which the covariance effect will therefore appear.

6. CONCLUSION

Previous research treats the economic parameters on which agents base their expectations as nonstochastic. This can be justified in two ways. Under rational expectations, the strong assumption is made that economic agents know all the relevant parameters. In previous work on learning, in order to resolve expectations it is assumed, explicitly or implicitly, that the parameter estimates are not random variables. It is understandable why previous research has avoided fully treating parameters as random variables. Doing so is difficult, so difficult that it has the potential to make some problems intractable.

This paper directly tackles the problem of treating a relevant parameter as a random variable and makes some progress: We draw attention to some of the rules that govern the expectation operator in this environment, derive three potentially useful propositions, and develop a numerical method for approximating the agent's subjective expectations.

A key point that emerges from our analysis is what we refer to as the covariance effect. We illustrate the potential importance of the covariance effect using a consumption example. In the case of consumption, the covariance effect can explain several previously documented empirical results. We present four other examples (one in Section 2 and three in Section 5) of how the covariance effect enters a variety of economic models—asset pricing, money demand, exchange rate, and fixed investment. These examples suggest that there is much room for further research to explore whether the covariance effect can explain empirical anomalies in models in which the relevant economic parameters are assumed to be known or nonstochastic.

NOTES

1. See, for example, Timmermann (1996, p. 528), who points out that if agents treated the parameter that they are learning "as a stochastic variable, analysis of the effect of learning on stock price volatility would be intractable."

2. Timmermann (1996) considers the contribution of a covariance term to the volatility of stock prices under learning. To assess the contribution of learning to excess volatility, he [in his eq. (13)] writes the variance of the stock price as the sum of three terms: the variance of the rational expectations stock price, the variance of the part reflecting learning, and twice their covariance. This last term, the covariance term, incorporates the covariance between the parameter estimate and the current realization of the exogenous dividend. However, since the agents in his model treat their parameter estimates as known constants, their dividend forecasts do not include the covariance effect that is the subject of this paper.

3. However, note that the covariance effect does not alter the fundamental convergence results of the learning literature as presented, for example, by Evans and Honkapohja (2000).

4. Goodfriend (1992) and Pischke (1995) study how a different type of learning can explain the correlation of aggregate consumption changes with lagged aggregate income. In their models, agents cannot (or do not) immediately distinguish aggregate from idiosyncratic shocks to income, but do learn the true nature of shocks over time. In these models, household consumption will follow a random walk but measures of aggregate consumption will suffer from an "information aggregation bias," which causes lagged aggregate income to predict changes in aggregate consumption.

5. For an extensive review of micro empirical evidence on excess sensitivity, see Browning and Lusardi (1996).

6. To simplify notation, we abstract from any constant terms.

7. It is easiest to see how the covariance term arises by working from the iterated expectations term, $\hat{E}_t \hat{E}_{t+1} D_{t+2}$. Still, the law of iterated expectations applies here so that $\hat{E}_t \hat{E}_{t+1} D_{t+2} = \hat{E}_t D_{t+2}$, and the result in (10) can be obtained from a consideration of $\hat{E}_t D_{t+2}$ instead of $\hat{E}_t \hat{E}_{t+1} D_{t+2}$.The somewhat more complicated derivation that begins from $\hat{E}_t D_{t+2}$ is available from the authors on request.

8. We thank an anonymous referee for bringing this interesting paper to our attention and the authors for providing us with a preliminary draft.

9. Blanchard and Mankiw (1988).

10. A brief derivation of (14) and (15) is given in the Appendix.

11. Browning and Lusardi (1996) list six studies that use this specification as well as several other studies that use the closely related specification in which consumption growth is regressed on (instrumented) earnings growth.

12. In addition to the reasons mentioned earlier, we focus on the AR(1) specification because a single-parameter stochastic process helps to make the effects of learning more transparent than does a stochastic process with more parameters.

13. We focus here on the implications of parameter estimation uncertainty when the correct specification of the data generating process (DGP) is known. Pesaran and Timmermann (1995) look at the interesting question of what happens when there is also uncertainty about the DGP. They study the predictability of stock market returns when investors use a predefined selection criterion to reevaluate the specification of the DGP in each period. They find an economically significant improvement in performance when the forecasting model is allowed to change in response to historically available information.

14. Hamilton (1994, pp. 74–76).

15. Hamilton (1994, pp. 334, 358).

16. The derivation of (20) is given in the Appendix. Timmermann [1996, eq. (14)] derives a similar expression in an asset pricing model by linearizing the excess rate of return. His result gives the excess rate of return as a function of the change in parameter estimates and of the deviation of those estimates from their true value.

17. The Taylor-series approximation of the present value in (15) may be less accurate for large values of τ. This is because long-horizon present values are sensitive to the persistence parameter ρ.

This sensitivity may not be fully captured by the Taylor-series expansion. We discuss a second approach, which relies on numerical methods, later.

18. To derive (25), follow the procedure used to obtain (20), which is given in the Appendix.

19. For comparison, Zeldes (1989) has 9,362 degrees of freedom in the regression reported in the first column of his Table 2.

20. Variation in σ and the ratio of assets to mean income made little difference, and so, the results are not reported separately.

21. The learning system considered in this paper satisfies the conditions of the Benveniste et al. (1990) theorem of the convergence rate, as cited and applied by Marcet and Sargent (1995). Regarding that theorem, Marcet and Sargent (1995, p. 205) state that "the assertion of the theorem of root-t convergence seems to be nearly true in samples of about 10,000. It is remarkable, though, that in samples of smaller size the rate of convergence can be very low."

22. Judd (1998, p. 394).

23. Guidolin and Timmermann (2000), who study learning when agents explicitly treat their parameter estimates as random variables, also use a Bayesian approach to evaluate the representative agent's beliefs.

24. Hamilton (1994, pp. 354, 358).

25. When we perform the simulations without imposing $\rho^A \leq \mathrm{R}$, there is no qualitative change in our results.

26. For $t = 15$, the numerical method occasionally (never more than once in 5,000 repetitions) generated a negative value of consumption. We discarded these observations and generated new observations to replace them.

27. To check robustness, we conducted simulations that used $N_R = 1{,}000$ in Step 4 and that generated a cross section of 10,000 observations. Each of these changes significantly increased the computational burden with little or no effect on our results.

28. This presentation of the Cagan model follows Blanchard and Fischer (1989, pp. 216–217), where it is shown that the OLG model with money has an equilibrium condition similar to (38).

29. See, for example, Mussa (1982).

30. Here we modify Lewis's (1989) model in order to simplify our exposition. In her model, the first differences of m_t^s and y_t follow AR (1) processes, and deviations from purchasing power parity follow a random walk.

31. Our statement of purchasing power parity assumes a normalization of units such that $e_t + p_t^f - p_t^u = 0$.

32. Sargent (1987, pp. 407–411) uses a similar model to study endogenous growth.

33. They express all variables as deviations from means to dispense with constant terms.

34. To resolve expectations, they treat $\hat{\rho}_t$ as a known constant.

REFERENCES

Benveniste, A., M. Metivier & P. Priouret (1990) *Adaptive Algorithms and Stochastic Approximations*. Berlin: Springer.

Blanchard, O. & S. Fischer (1989) *Lectures on Macroeconomics*. Cambridge, MA: MIT Press.

Blanchard, O. & N.G. Mankiw (1988) Consumption: Beyond certainty equivalence. *American Economic Review* 78, 173–177.

Browning, M. & A. Lusardi (1996) Household saving: Micro theories and micro facts. *Journal of Economic Literature* 34, 1797–1855.

Evans, G. & S. Honkapohja (2001) *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press.

Goodfriend, M. (1992) Information aggregation bias. *American Economic Review* 82, 506–519.

Guidolin, M. & A. Timmermann (2000) Asset Prices on Learning Paths. Mimeo, University of California at San Diego.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton, NJ: Princeton University Press.
Honkapohja, S. & K. Mitra (2000) Learning with Bounded Memory in Stochastic Models. RUESG and Department of Economics working paper, University of Helsinki.
Judd, K. (1998) *Numerical Methods in Economics*. Cambridge, MA: MIT Press.
Lewis, K. (1989) Can learning affect exchange rate behavior? The case of the dollar in the early 1980's. *Journal of Monetary Economics* 23, 79–100.
Lucas, R.E., Jr. (1978) Asset prices in an exchange economy. *Econometrica* 46, 1426–1445.
Marcet, A. & T.J. Sargent (1989a) Convergence of least squares learning in environments with hidden state variables and private information. *Journal of Political Economy* 97, 1306–1322.
Marcet, A. & T.J. Sargent (1989b) Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic Theory* 48, 337–368.
Marcet, A. & T.J. Sargent (1995) Speed of convergence of recursive least squares: Learning with autoregressive moving average perceptions. In A. Kirman & M. Salmon (eds.), *Learning and Rationality in Economics*, pp. 179–215. Oxford: Blackwell.
Meese, R. (1986) Testing for bubbles in exchange markets: A case of sparkling rates? *Journal of Political Economy* 94, 345–372.
Mussa, M. (1982) A model of exchange rate dynamics. *Journal of Political Economy* 90, 74–104.
Nankervis, J.C. & N.E. Savin (1988) The exact moments for the least-squares estimator for the autoregressive model: Corrections and extensions. *Journal of Econometrics* 37, 381–388.
Pesaran, M.H. & A. Timmermann (1995) Predictability of stock returns: Robustness and economic significance. *Journal of Finance* 50, 1201–1228.
Pischke, J.-S. (1995) Individual income, incomplete information, and aggregate consumption. *Econometrica* 63, 805–840.
Sargent, T. (1987) *Macroeconomic Theory*, 2nd ed. San Diego, CA: Academic Press.
Timmermann, A. (1996) Excess volatility and predictability of stock prices in autoregressive dividend models with learning. *Review of Economic Studies* 63, 523–557.
Zeldes, S.P. (1989) Consumption and liquidity constraints: An empirical investigation. *Journal of Political Economy* 97, 305–346.

APPENDIX

Derivation of equations (14) and (15). For quadratic utility with $\beta = 1/R$, the solution to (11) implies that

$$C_{t+j} = \hat{E}_{t+j} C_{t+j+1} \quad \text{for} \quad j = 0, 1, 2, \ldots, \tau - 1.$$

Use this condition together with (12) to derive (14) from (13). Repeat recursively to obtain (15).

Derivation of equation (20). For period $t+1$ with $\tau = 2$, equation (14) gives

$$C_{t+1} = \frac{R}{1+R}\left[A_{t+1} + Y_{t+1} + \frac{1}{R}\hat{E}_{t+1}Y_{t+2}\right],$$

which we can rewrite as

$$C_{t+1} = \bar{y} + \frac{R}{1+R}\left[A_{t+1} + y_{t+1} + \frac{1}{R}\hat{\rho}_{t+1}y_{t+1}\right]. \quad \textbf{(A.1)}$$

https://doi.org/10.1017/S1365100501010069 Published online by Cambridge University Press

Additionally, for period t with $\tau = 2$, equation (15) gives

$$C_t = \frac{R^2}{1+R+R^2}\left[A_t + Y_t + \frac{1}{R}\hat{E}_t Y_{t+1} + \frac{1}{R^2}\hat{E}_t Y_{t+2}\right],$$

which we can rewrite, using (7), (10), and the definition (21), as

$$C_t = \bar{y} + \frac{R^2}{1+R+R^2}\left[A_t + y_t + \frac{1}{R}\hat{\rho}_t y_t + \frac{1}{R^2}\hat{\rho}_t^2 y_t + \frac{1}{R^2}\hat{\sigma}_{\rho y}(t)\right]. \quad \textbf{(A.2)}$$

Use (12) in (A.1) to write

$$C_{t+1} = \bar{y} + \frac{R}{1+R}\left[R(A_t + y_t + \bar{y} - C_t) + \left(1 + \frac{1}{R}\hat{\rho}_{t+1}\right) y_{t+1}\right]. \quad \textbf{(A.3)}$$

Note from (A.2) that

$$A_t + y_t = \left(\frac{1+R+R^2}{R^2}\right)(C_t - \bar{y}) - \left(\frac{1}{R}\hat{\rho}_t + \frac{1}{R^2}\hat{\rho}_t^2\right) y_t - \frac{1}{R^2}\hat{\sigma}_{\rho y}(t).$$

Substituting into (A.3) then gives

$$C_{t+1} = C_t + \frac{R}{1+R}\left[\left(1 + \frac{1}{R}\hat{\rho}_{t+1}\right) y_{t+1} - \left(1 + \frac{1}{R}\hat{\rho}_t\right)\hat{\rho}_t y_t\right] - \left(\frac{1}{1+R}\right)\hat{\sigma}_{\rho y}(t).$$

Use (17) to substitute for y_{t+1} to obtain (20).

Proof of Proposition 1.

$$\hat{\sigma}_{\rho y}(t) = \hat{E}_t\{[\hat{\rho}_{t+1} - \hat{E}_t\hat{\rho}_{t+1}][y_{t+1} - \hat{E}_t y_{t+1}]\}.$$

Since $\hat{E}_t\hat{\rho}_{t+1} = \hat{\rho}_t$ and $\hat{E}_t y_{t+1} = \hat{\rho}_t y_t$, we have

$$\hat{\sigma}_{\rho y}(t) = \hat{E}_t\{[\hat{\rho}_{t+1} - \hat{\rho}_t][y_{t+1} - \hat{\rho}_t y_t]\}. \quad \textbf{(A.4)}$$

Using equation (29) to substitute for $[\hat{\rho}_{t+1} - \hat{\rho}_t]$ in (A.4) gives

$$\hat{\sigma}_{\rho y}(t) = \frac{1}{t+1} M_t^{-1} y_t \hat{E}_t [y_{t+1} - \hat{\rho}_t y_t]^2. \quad \textbf{(A.5)}$$

Note that $(y_{t+1} - \hat{\rho}_t y_t) = [(\rho - \hat{\rho}_t) y_t + \varepsilon_{t+1}]$. Since

$$\hat{\rho}_t = \rho + \left(\frac{\sum_0^{t-1} y_i \varepsilon_{i+1}}{\sum_0^{t-1} y_i^2}\right),$$

we have

$$\hat{E}_t(y_{t+1} - \hat{\rho}_t y_t)^2 = \hat{E}_t\left[\left(\frac{-\sum_0^{t-1} y_i \varepsilon_{i+1}}{\sum_0^{t-1} y_i^2}\right)^2 y_t^2 + \varepsilon_{t+1}^2 - 2\left(\frac{\sum_0^{t-1} y_i \varepsilon_{i+1}}{\sum_0^{t-1} y_i^2}\right) y_t \varepsilon_{t+1}\right].$$

Since the ε_i's are known to be i.i.d., the expectation of the last term on the right-hand side is zero and the first term on the right-hand side is

$$\hat{E}_t\left[\left(\frac{-\sum_0^{t-1} y_i\varepsilon_{i+1}}{\sum_0^{t-1} y_i^2}\right)^2 y_t^2\right] = \frac{\left(\sum_0^{t-1} y_i^2\right)\sigma^2}{\left(\sum_0^{t-1} y_i^2\right)^2} y_t^2.$$

Collecting results and using $\hat{E}_t\varepsilon_{t+1}^2 = \sigma^2$ gives

$$\hat{E}_t(y_{t+1} - \hat{\rho}_t y_t)^2 = \left[\frac{\left(\sum_0^{t-1} y_i^2\right)\sigma^2}{\left(\sum_0^{t-1} y_i^2\right)^2} y_t^2 + \sigma^2\right]$$

$$= \left[\frac{y_t^2}{\sum_0^{t-1} y_i^2} + 1\right]\sigma^2 = \left[\frac{y_t^2 + \sum_0^{t-1} y_i^2}{\sum_0^{t-1} y_i^2}\right]\sigma^2,$$

which gives

$$\hat{E}_t(y_{t+1} - \hat{\rho}_t y_t)^2 = \left[\frac{\sum_0^t y_i^2}{\sum_0^{t-1} y_i^2}\right]\sigma^2. \qquad \textbf{(A.6)}$$

Use (A.5), (A.6), and the definition of M_t,

$$M_t = \frac{1}{t+1}\sum_{i=0}^{t} y_i^2,$$

to write

$$\hat{\sigma}_{\rho y}(t) = \left(\frac{1}{\sum_0^t y_i^2}\right) y_t \left(\frac{\sum_0^t y_i^2}{\sum_0^{t-1} y_i^2}\right)\sigma^2 = \left(\frac{\sigma^2}{\sum_0^{t-1} y_i^2}\right) y_t. \qquad \blacksquare$$

Proof of Proposition 2. Define $f(\hat{\rho}_{t+1}, y_{t+1}) \equiv \hat{\rho}_{t+1}^j y_{t+1}$. The expectation of the second-order Taylor-series expansion of this function around $(\hat{\rho}_t, \hat{\rho}_t y_t)$ gives

$$\hat{E}_t\left(\hat{\rho}_{t+1}^j y_{t+1}\right) \approx \hat{E}_t f(\hat{\rho}_t, \hat{\rho}_t y_t) + \frac{\partial f(\cdot,\cdot)}{\partial\hat{\rho}_{t+1}}\hat{E}_t(\hat{\rho}_{t+1} - \hat{\rho}_t) + \frac{\partial f(\cdot,\cdot)}{\partial y_{t+1}}\hat{E}_t(y_{t+1} - \hat{\rho}_t y_t)$$

$$+ \frac{\partial^2 f(\cdot,\cdot)}{\partial\rho_{t+1}\partial y_{t+1}}\hat{E}_t[(y_{t+1} - \hat{\rho}_t y_t)(\hat{\rho}_{t+1} - \hat{\rho}_t)] + \frac{1}{2}\frac{\partial^2 f(\cdot,\cdot)}{(\partial\hat{\rho}_{t+1})^2}\hat{E}_t(\hat{\rho}_{t+1} - \hat{\rho}_t)^2$$

$$+ \frac{1}{2}\frac{\partial^2 f(\cdot,\cdot)}{(\partial y_{t+1})^2}\hat{E}_t(y_{t+1} - \hat{\rho}_t y_t)^2 = f(\hat{\rho}_t\hat{\rho}_t y_t) + \frac{\partial^2 f(\cdot,\cdot)}{\partial\hat{\rho}_{t+1}\partial y_{t+1}}\hat{\sigma}_{\rho y}^2(t)$$

$$+ \frac{1}{2}\frac{\partial^2 f(\cdot,\cdot)}{(\partial\hat{\rho}_{t+1})^2}\hat{\sigma}_{\rho}^2(t) + \frac{1}{2}\frac{\partial^2 f(\cdot,\cdot)}{(\partial y_{t+1})^2}\widehat{\mathrm{Var}}_t(y_{t+1}).$$

Note that $f(\hat{\rho}_t, \hat{\rho}_t y_t) = \hat{\rho}_t^{j+1} y_t$. Also, evaluating the derivatives at $\hat{\rho}_{t+1} = \hat{\rho}_t$, $y_{t+1} = \hat{\rho}_t y_t$ gives

$$\frac{\partial^2 f(\cdot,\cdot)}{(\partial\hat{\rho}_{t+1})^2} = j(j-1)\hat{\rho}_t^{j-1} y_t, \quad \frac{\partial^2 f(\cdot,\cdot)}{(\partial y_{t+1})^2} = 0, \text{ and } \frac{\partial^2 f(\cdot,\cdot)}{\partial\hat{\rho}_{t+1}\partial y_{t+1}} = j\hat{\rho}_t^{j-1}.$$

Collecting, we have $\hat{E}_t(\hat{\rho}_{t+1}^{j} y_{t+1}) \approx \hat{\rho}_t^{j+1} y_t + j\hat{\rho}_t^{j-1}\hat{\sigma}_{\rho y}(t) + \frac{1}{2} j(j-1)\hat{\rho}_t^{j-1} y_t \hat{\sigma}_\rho^2(t)$. ■

Proof of Proposition 3. $\hat{\sigma}_\rho^2(t) = \hat{E}_t[\hat{\rho}_{t+1} - \hat{E}_t\hat{\rho}_t]^2$. Using $\hat{E}_t\hat{\rho}_{t+1} = \hat{\rho}_t$ and equation (27) gives

$$\hat{\sigma}_\rho^2(t) = \left[\left(\frac{1}{t+1}\right) M_t^{-1} y_t\right]^2 \hat{E}_t[y_{t+1} - \hat{\rho}_t y_t]^2.$$

Using (A.5) then gives

$$\hat{\sigma}_\rho^2(t) = \left[\left(\frac{1}{t+1}\right) M_t^{-1} y_t\right] \hat{\sigma}_{\rho y}(t).$$

Since

$$\left(\frac{1}{t+1}\right) M_t^{-1} = \left(\frac{1}{\sum_0^t y_i^2}\right),$$

this gives

$$\hat{\sigma}_\rho^2(t) = \frac{y_t}{\sum_0^t y_i^2} \sigma_{\rho y}(t).$$ ■

https://doi.org/10.1017/S1365100501010069 Published online by Cambridge University Press