



A simple recursive forecasting model

William A. Branch ^a, George W. Evans ^{b,*}

^a *University of California, Irvine, United States*

^b *Department of Economics, 1285 University of Oregon, Eugene, OR 97403-1285, United States*

Received 8 February 2005; accepted 1 September 2005

Abstract

We compare the performance of alternative recursive forecasting models. A simple constant gain algorithm, used widely in the learning literature, both forecasts well out of sample and also provides the best fit to the Survey of Professional Forecasters.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Constant gain; Recursive learning; Expectations

JEL classification: E37; D84; D83

1. Introduction

There is now an extensive literature which, in place of rational expectations (RE), assumes that agents behave as econometricians when making forecasts (see [Evans and Honkapohja \(2001\)](#) for a systematic treatment and references). In particular, a growing number of theoretical and applied macro models replace RE with a linear forecasting rule, the parameters of which are updated by a constant gain version of least squares ([Sargent \(1999\)](#), [Evans and Honkapohja \(2001, 2005\)](#), [Cho, Williams, and Sargent \(2002\)](#), [Williams \(2002\)](#), [Kasa \(2004\)](#), [McGough \(in press\)](#), [Orphanides and Williams \(2003, 2005\)](#), [Bullard and Cho \(in press\)](#), [Cho and Kasa \(2002\)](#), [Bullard and Duffy 2004](#)). Although theoretically

* Corresponding author. Tel.: +1 541 346 4662; fax: +1 541 346 1243.

E-mail address: gevans@uoregon.edu (G.W. Evans).

motivated, an important empirical issue remains open in this literature: is constant gain learning a plausible bounded rationality assumption? To address this issue we answer two questions: do constant gain algorithms perform well in out-of-sample forecasting in comparison to other recursive forecasting models, and, does constant gain learning provide a good description of actual forecaster behavior?

We begin by setting out a general recursive forecasting model that allows the parameters to follow a random walk. As noted by Ljung and Soderstrom (1983) and Sargent (1999), depending on the covariance structure of the stochastic components of the parameter law of motion, specific adaptive learning algorithms arise as special cases, including constant gain least squares. Time-varying parameter models of the economy have been estimated extensively before, and recently by Cogley and Sargent (2005) and Sims and Zha (in press). Our principal contributions are two-fold. First, we estimate the optimal gains for the constant gain least squares model and show that it provides the best out of sample forecasts, compared to the alternatives, for inflation and GDP growth. Second, we show that a simple version of this model, namely constant gain least squares with equal gain for each variable, provides the best fit to the Survey of Professional Forecasters (SPF).

2. Forecasting model

We estimate forecasting models of inflation and output growth that are based on recursive parameter updating¹, and then evaluate their performance in out-of-sample forecasts. The importance of out-of-sample forecasts, to protect against data-mining and to provide a stringent test of alternative models, has been stressed, e.g., in Harvey (1990), Stock and Watson (1996) and Pesaran and Timmermann (1995). As a second way to assess the models, we compare the forecasts to those of the SPF.

2.1. A unified framework

We follow Ljung and Soderstrom (1983) and Sargent (1999) in specifying a general state space forecasting model that will nest our alternative models. Let $y_t = (y_{1,t}, \dots, y_{n,t})'$ denote the variables of interest. The economic law of motion takes the form

$$y_{j,t} = b_{j,t}' x_t + \varepsilon_{j,t}, \quad j = 1, \dots, n,$$

$$b_{j,t} = b_{j,t-1} + \theta_{j,t}$$

where $b_{j,t}$ is the $(K \times 1)$ parameter vector and x_t is the $(K \times 1)$ vector of explanatory variables. $\varepsilon_{j,t}$ and $\theta_{j,t}$ are assumed to be zero mean mutually independent random sequences with $E\varepsilon_{j,t}\varepsilon_{j',t}' = R_{2,t}^j$ and $E\theta_{j,t}\theta_{j',t}' = R_{1,t}^j$.

We are interested in Vector Autoregression (VAR) specifications so that $x_t = (1, y_{t-1})'$.² Conditional forecasts of $y_{j,t}$ are constructed as

$$y_{j,t|t-1} = \hat{b}_{j,t-1}' x_t.$$

¹ Some authors have argued that forecasts based on real-time data sets, rather than revised data, are preferable for matching survey data. To keep the analysis as close to the existing literature as possible we use historical data. For a discussion of real-time data in forecasting see Croushore and Stark (2002) and Kozicki and Tinsley (2002).

² For transparency, we focus on equation by equation estimation.

The forecasting problem is selection of an appropriate procedure for constructing the sequence $\{\hat{b}_t\}$. The parameter vectors can be estimated by the Kalman Filter recursion³,

$$\hat{b}_{j,t} = \hat{b}_{j,t-1} + L_{j,t}(y_{j,t} - \hat{b}'_{j,t-1} x_t) \quad (1)$$

$$L_{j,t} = \frac{P_{j,t-1} x_t}{R_{2,t}^j + x_t' P_{j,t-1} x_t} \quad (2)$$

$$P_{j,t} = P_{j,t-1} - \frac{P_{j,t-1} x_t x_t' P_{j,t-1}}{R_{2,t}^j + x_t' P_{j,t-1} x_t} + R_{1,t}^j, \quad (3)$$

where $\hat{b}_{j,t}$ has covariance matrix $P_{j,t} = E(b_{j,t} - \hat{b}_{j,t})(b_{j,t} - \hat{b}_{j,t})'$.

The properties of the estimator depend on the assumptions for $R_{1,t}^j$, $R_{2,t}^j$ and priors $\hat{b}_{j,0}$, $P_{j,0}$. Cogley and Sargent (2005) pursue a Bayesian estimation strategy where they characterize the posterior densities of b_j under very general assumptions on $R_{1,t}^j$. Our interest is in comparing the forecast performance of alternative assumptions on $R_{1,t}^j$ that yield simple recursive learning rules. We turn now to our specifications of $R_{1,t}^j$.

2.1.1. Recursive least squares

As noted by Ljung and Soderstrom (1983) and Sargent (1999), under the restrictions

$$R_{1,t}^j = 0 \text{ and } R_{2,t}^j = 1, \text{ for } j = 1, \dots, n,$$

the Kalman Filter (1)–(3) is equivalent to recursive least squares (RLS),

$$\hat{b}_{j,t} = \hat{b}_{j,t-1} + \gamma_{j,t} R_t^{-1} x_t (y_{j,t} - \hat{b}'_{j,t-1} x_t) \quad (4)$$

$$R_t = R_{t-1} + \gamma_{j,t} (x_t x_t' - R_{t-1}) \quad (5)$$

where $\gamma_{j,t} = t^{-1}$. RLS is simply a recursive formulation of ordinary least squares (e.g. Evans and Honkapohja (2001)).

2.1.2. Constant gain least squares

As an alternative to RLS, Sargent (1999) emphasizes the restrictions,

$$R_{1,t}^j = \frac{\gamma_j}{1 - \gamma_j} P_{j,t-1} \text{ and } R_{2,t}^j = 1 - \gamma_j,$$

and shows that it leads to the constant gain version of recursive least squares (4) and (5) in which $\gamma_{j,t} = \gamma_j$ where $0 < \gamma_j < 1$. The scalar γ_j , referred to as the “gain,” in this rule is constant over time in contrast to the “decreasing gain” t^{-1} of RLS. Constant gain least squares discounts past observations at geometric rate $1 - \gamma_j$ and is consequently more robust to structural change. We allow $\gamma_j \neq \gamma_{j'}$ for $j \neq j'$, which would be appropriate if variables have different rates of structural change.

³ The usual Kalman filter recursion is split into ‘prediction’ and ‘transition’ equations. It is straightforward to verify that for our assumed stochastic process the equations given represent the Kalman filter recursion.

2.1.3. Stationary time-varying parameters

The restrictions on $R_{1,t}^j$ restrict the conditional distributions for the parameter vector. For RLS the parameters are non-stochastic and under constant gain least squares they may be heteroskedastic. As a third alternative, we follow [Stock and Watson \(1996\)](#) in specifying a time-varying parameter model estimated by the Kalman filter, with homoskedastic parameter error structure:

$$R_{1,t}^j = \lambda_j^2 \sigma_j^2 Q \quad \text{and} \quad R_{2,t}^j = \sigma_j^2$$

where σ_j^2 is estimated as the variance from the residuals of an AR(1) model for y_j , $Q = (Ex_t x_t')^{-1}$, and λ_j is a parameter for controlling the drift of the parameter estimates, which we again allow to differ across j . This approach is similar to [Stock and Watson \(1996\)](#) and [Doan, Litterman, and Sims \(1984\)](#) as formulated by [Hamilton \(1994\)](#). For brevity we will refer to this as the “Kalman filter” learning rule.

2.1.4. A simple model

Finally, we also consider the special case of constant gain least squares in which we impose a common constant gain $\gamma_j = \gamma_j = \gamma$ across variables. This model has the attraction of relative simplicity, which is, of course, one of the motivations of bounded rationality. More generally, parsimony is often recommended as one criterion for selecting between forecasting models ([Harvey, 1990](#)).

3. Results

We focus on low order VARs. Although there are more sophisticated (and likely more accurate) ways to model the economy – for example, [Cogley and Sargent \(2005\)](#) – we are interested in simple, parsimonious forecasting models that are plausible descriptions of boundedly rational agents. We therefore examine the four formulations just presented using a bivariate VAR(1) in output growth and inflation.⁴ Output growth is measured in terms of quarterly real GDP calculated as $\log(\text{GDP}_t / \text{GDP}_{t-1})$. Inflation is measured similarly using the quarterly GDP deflator. These variables are chosen to stay close to the SPF, which asks a sample of professional forecasters, among other questions, for their expectation of one step ahead GDP and inflation.

Our approach is straightforward. We split the sample into three parts: the first, a pre-forecasting period, is the period in which prior beliefs are formed by estimating the VAR; the second part is an in-sample period in which the optimal parameters $\gamma_j, \gamma_j, j = 1, 2$ are determined (for RLS the gain sequence continues to be updated as t^{-1}); finally, the remainder is the out-of-sample forecasting period. We segment these samples so that the out-of-sample forecasting period corresponds to the SPF: 1981:3–2002:4. We choose a fairly long pre-forecasting period, 1947:2–1969:4, to avoid over-sensitivity of the initial in-sample estimates. The in-sample period is thus 1970:1–1981:2.

We envision professionals as forecasting throughout the sample. When they begin reporting their forecasts in 1981:3, their forecasting model has been shaped by experience, with the free parameters chosen optimally for fitting the in-sample period. Our metrics for whether a forecast method is a ‘reasonable’ model are mean square forecast error in out-of-sample forecasting and mean square deviations from the survey responses in the SPF.

⁴ Our key qualitative results hold also for a VAR(4).

3.1. Optimal in-sample parameter choices

To measure forecast success, we calculate the mean square forecast error

$$\text{MSE}(y_j) = \frac{1}{T} \sum_{t=t_0}^T (y_{j,t} - \hat{y}_{j,t})^2$$

where y_j is the j th component of y and $\hat{y}_{j,t}$ is the forecast of the j th component based on $t-1$ information. For constant gain least squares we determine the gain j which minimizes the in-sample MSE by searching over all $\gamma_j \in (0, 1), j=1, 2$, with $t_0=1970.1$ and $T=1982.2$. If the degree of structural change differs across series then we would expect the optimal gain to be different. Evidence of different structural changes in inflation and GDP growth is provided by Cogley and Sargent (2005) and Branch and Evans (2005). We choose the in-sample estimates of λ_j , for the Kalman filter rule, in the same way. The results for the “in-sample” period 1970:1–1981:2 are as follows:

1970:1–1981:2	γ	λ
GDP growth	0.007	0.001
Inflation	0.062	0.021

The estimates of both γ and λ are larger for inflation than for GDP growth, presumably because the degree of structural change in GDP growth during the 1970s was considerably less than for inflation (the theoretical link between degree of structural change and equilibrium gain in a self-referential model is investigated in Evans and Ramey (in press)). An estimate of the optimal gain γ was also given in Orphanides and Williams (2003). However, their estimate is based on calibration of a structural model. Our focus is on out-of-sample forecasting performance, to which we now turn.

3.2. Forecast comparison

We now compare how the forecasting models perform in out-of-sample forecasting over 1981:3–2002:4. The top panel of Table 1 presents out-of-sample MSE results for RLS, constant gain and the Kalman filter, where the learning rule parameters have been set at their in-sample 1970:1–1981:2 estimates.⁵ Table 1 demonstrates that the optimal constant gain algorithm outperforms both RLS and the Kalman filter over the period 1981:3–2002:4. One possible explanation is that both parameter drift and conditional heteroskedasticity are present in the data, and that constant gain least squares captures both of these features. The top panel of Table 1 also shows that the forecasting gains are non-trivial, from about 2.9–4.3% in the case of GDP growth and 5.6–12.5% for inflation forecasting.

These results have two implications for theoretical and applied models. First, constant gain least squares, with each gain optimally tuned to track the variable of interest, does appear plausible as a model of bounded rationality given its satisfactory out-of-sample performance. Second, in line with Cogley and Sargent (2005), the results suggest that both parameter drift and stochastic volatility are components of the stochastic structure of the economy.

⁵ The Stock and Watson (1996) approach sets $R'_{1,t} = \lambda^2 \sigma^2 Q$ where $Q = (Ex_t x'_t)^{-1}$. Stock and Watson compare $\lambda = .0025, .0075, .015, .01, .02$. We instead set λ_j according to in-sample forecasting success.

Table 1
Comparisons of fit across forecasting models

Out-of-sample period: 1981:3–2002:4		
	GDP growth	Inflation
<i>Mean square forecast error</i>		
RLS	.3507	.0532
Constant gain	.3408	.0503
Kalman filter	.3558	.0570
log(RLS/Constant gain)	2.86%	5.61%
log(Kalman filter/Constant gain)	4.31%	12.50%
<i>Mean square SPF error</i>		
RLS	.1354	.0316
Constant gain	.1204	.0266
Kalman filter	.1258	.0259
Simple Model	.1181	.0250
log(RLS/Simple Model)	13.67%	23.43%
log(Constant gain/Simple Model)	1.93%	6.20%
log(Kalman filter/Simple Model)	6.32%	3.54%

Learning rule parameters have been set at their 1970:1–1981:2 in-sample estimates, with the Simple Model gain set at .0345. The top panel of the table is the mean square forecast error in out-of-sample forecasting of actual inflation and GDP. The bottom panel is the mean square deviations of the forecasting models from the median response of the SPF.

3.3. Comparisons to the survey of professional forecasters

We now compare the forecasts from the optimal constant gain, RLS, Kalman filter, and simple constant gain models to those from the SPF. The learning literature often motivates bounded rationality on the basis of assuming that forecasters behave like econometricians, and thus one test of the plausibility of constant gain models of adaptive learning is their ability to fit professionals' forecasts.

For comparison with the SPF we calculate the Mean Square Comparison Error (MSCE)

$$\text{MSCE}(y_j) = \frac{1}{T} \sum_{t=1}^T \left(y_{j,t}^{\text{SPF}} - \hat{y}_{j,t} \right)^2$$

where y_j^{SPF} is the median survey forecast of the j th component of y . We take the same optimal parameters $\gamma_j, \lambda_j, j=1, 2$ estimated for the in-sample period 1970:1–1981:2, and compare whether the forecasts over 1981:3–2002:4 are consistent with the SPF over this period. We also assess the simple model in which a common constant gain is imposed for both variables. For the simple constant gain model we choose γ to be average of the optimal gains for the two variables in the in-sample period, i.e. $\gamma = .0345$.

The bottom panel of Table 1 reports the results. RLS is clearly dominated by the other forecast methods that allow for parameter drift. For GDP growth the optimal constant gain method fits more closely than the Kalman filter model, whereas the opposite is true for inflation forecasts. Most strikingly, however, the simple model with common gain $\gamma = .0345$ fits best across all models for both variables. This provides significant support for constant gain learning as a model of actual forecasting.

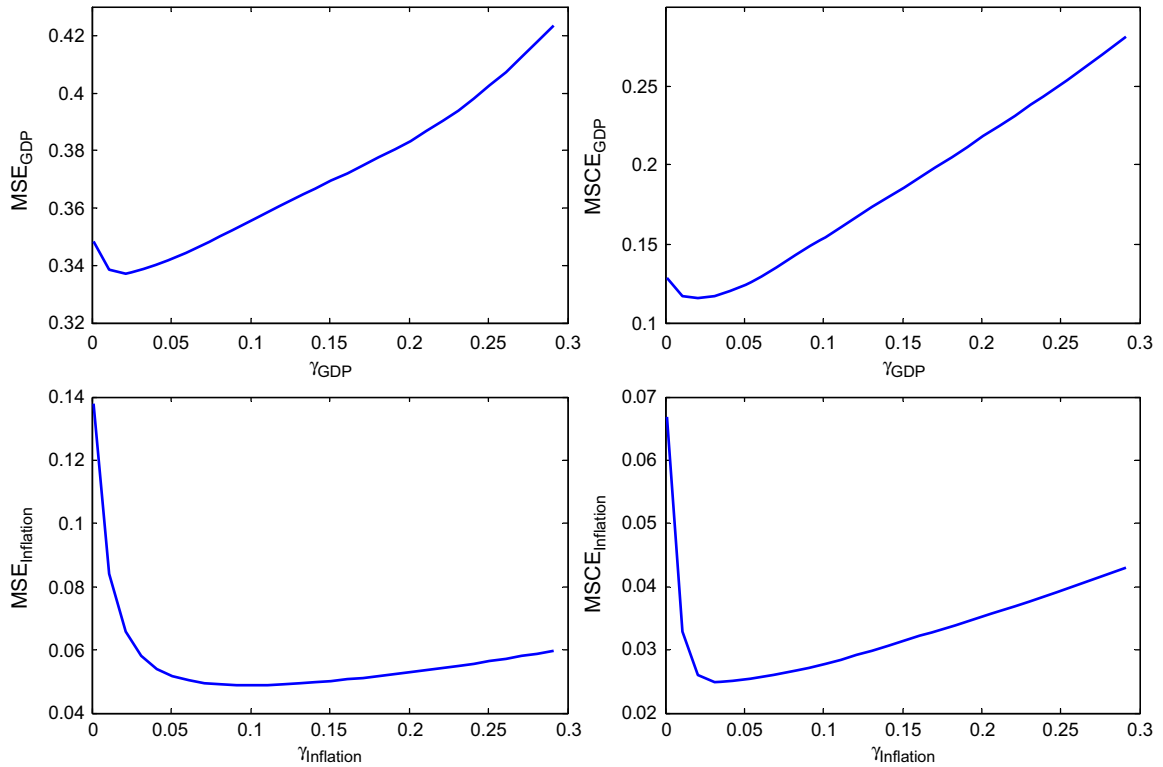


Fig. 1. Mean square comparison and forecast errors as a function of the gain. Left panels forecast out-of-sample 1983:3–2002:4; right panels forecast the SPF.

3.4. Discussion

Fig. 1 provides additional information on the performance of constant gain least squares in the 1981:3–2002:4 period. The left two curves show the MSE, for constant gain forecasts of GDP growth and inflation, and the right two curves show the corresponding MSCE, measuring the closeness of fit of constant gain forecasts to the SPF forecasts. In each case the curves plot mean square error as a function of different values of the gain.

Several features stand out. First, the optimal gains for forecasting both variables, given by the gains at which the curves are minimized, are higher during 1981:3–2002:4 than for the 1970:1–1981:2 period, and the optimal gain in the later as well as the earlier period is again higher for inflation than for GDP growth. However, given the flatness of the MSE for inflation forecasts, a common gain of $\gamma=0.05$ produces reasonable forecasts.⁶ Second, for explaining SPF forecasts, the best fitting gain for inflation is substantially below those for fitting actual data in the period, and a common constant gain in the range $\gamma=.03$ to $\gamma=.04$ provide the best fit for the SPF.

Overall our results provide empirical support for the recent “learning theory” approach in which agents are modeled as employing constant gain least squares forecasting rules, with the gain chosen to

⁶ If the two MSEs are weighted equally the optimal common gain for this period is $\gamma=.047$.

reflect the degree of structural change. Our results complement those of [Orphanides and Williams \(2003\)](#), who use a calibrated structural macroeconomic model to fit and analyze the role of stabilization policy in the US, and of [Milani \(2004\)](#) who estimates econometrically a DSGE model with learning and provides evidence that constant gain learning is a major source of persistence.

4. Conclusion

Constant gain least squares is becoming widely used in learning theory approaches to macroeconomics that replace RE with econometric forecasting rules. Using constant gain is a simple way of modifying least squares to adapt to ongoing structural change. We have examined the empirical performance of this learning rule, compared to recursive least squares and a more conventional time-varying parameter model, in forecasting US GDP growth and inflation. Our results are very encouraging. Constant gain least squares, with the gain optimally chosen in-sample for each variable, performs best in out-of-sample forecasting. An even simpler version, in which a common constant gain is used for both variables, provides the best fit for the Survey of Professional Forecasters.

While the most accurate forecasting model will depend on the stochastic nature of the economy, constant gain least squares does appear plausible as a model of adaptive learning for economic agents. Since the degree of structural change itself seems to change over time, it would be valuable in future work to explore empirical models in which agents gradually adapt the gain over time in response to observed data.⁷

References

- Branch, W.A., Evans, G.W., 2005. Model Uncertainty and Endogenous Volatility, mimeo.
- Bullard, J., Cho, I.-K., in press. Escapist Policy Rules, Working Paper 20002-022C, FRB St. Louis. *Journal of Economic Dynamics and Control*.
- Bullard, J., Duffy, J., 2004. Learning and Structural Change in Macroeconomic Data, Working Paper 2004-016A. Federal Reserve Bank of St. Louis.
- Cho, I.-K., Kasa, K., 2002. Learning Dynamics and Endogenous Currency Crises, mimeo, University of Illinois.
- Cho, I.-K., Williams, N., Sargent, T.J., 2002. Escaping Nash inflation. *Review of Economic Studies* 69, 1–40.
- Cogley, T., Sargent, T.J., 2005. Drifts and volatilities: monetary policy and outcomes in the post WWII U.S. *Review of Economic Dynamics* 8, 262–302.
- Croushore, D., Stark, T., 2002. Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics* 24, 507–531.
- Doan, T., Litterman, R.B., Sims, C.A., 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3, 1–100.
- Evans, G.W., Honkapohja, S., 2001. *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton, NJ.
- Evans, G.W., Honkapohja, S., 2005. Policy interaction, expectations and the liquidity trap. *Review of Economic Dynamics* 8, 303–323.
- Evans, G.W., Ramey, G., in press. Adaptive Expectations, Underparameterization and the Lucas Critique, University of Oregon Working Paper No. 237, *Journal of Monetary Economics*.
- Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton, NJ.
- Harvey, A., 1990. *The Econometric Analysis of Time Series*, 2nd ed. Philip Allan, London.

⁷ Evans and Ramey (in press) provide a suitable algorithm for updating the gain in a univariate setting.

- Kasa, K., 2004. Learning, large deviations, and recurrent currency crises. *International Economic Review* 45, 141–173.
- Ljung, L., Soderstrom, T., 1983. *Theory and Practice of Recursive Identification*. MIT Press.
- Kozicki, S., Tinsley, P.A., 2002. Comments on 'Forecasting with a real-time data set for macroeconomists'. *Journal of Macroeconomics* 24, 541–557.
- McGough, B., in press. *Shocking Escapes*, mimeo, Oregon State University, Economic Journal.
- Milani, F., 2004. *Expectations, Learning and Macroeconomic Persistence*, mimeo, Princeton University.
- Orphanides, A., Williams, J.C., 2003. The Decline of Activist Stabilization Policy: Natural Rate Misperceptions, Learning, and Expectations, FRBSF 2003-24.
- Orphanides, A., Williams, J.C., 2005. Imperfect knowledge, inflation expectations, and monetary policy. In: Bernanke, Ben S., Woodford, Michael (Eds.), Chapter 5 in *Inflation Targeting*. NBER and University of Chicago Press, Chicago.
- Pesaran, M.H., Timmermann, A., 1995. Predictability of stock returns: robustness and economic significance. *Journal of Finance* 50 (4), 1201–1228.
- Sargent, T.J., 1999. *The Conquest of American Inflation*. Princeton University Press, Princeton, NJ.
- Sims, C.A., Zha, T., in press. Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*.
- Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* 14, 11–30.
- Williams, N., 2002. *Escape Dynamics in Learning Models*, mimeo, Princeton University.