Review of Economic Dynamics 17 (2014) 367–382

Contents lists available at ScienceDirect

Review of Economic Dynamics

www.elsevier.com/locate/red

Learning, large deviations and rare events ☆

Jess Benhabib [a,*,1], Chetan Dave [b]

[a] New York University, Department of Economics, 19 W. 4th Street, 6FL, New York, NY 10012, USA
[b] New York University (Abu Dhabi), PO Box 903, New York, NY 10276, USA

ARTICLE INFO

Article history:
Received 26 January 2012
Received in revised form 12 September 2013
Available online 20 September 2013

JEL classification:
D80
D83
D84

Keywords:
Adaptive learning
Large deviations
Fat tails
Asset prices

ABSTRACT

We examine the role of generalized stochastic gradient constant gain (SGCG) learning in generating large deviations of an endogenous variable from its rational expectations value. We show analytically that these large deviations can occur with a frequency associated with a fat-tailed distribution even though the model is driven by thin-tailed exogenous stochastic processes. We characterize these large deviations, driven by sequences of consistently low or consistently high shocks and then apply our model to the canonical asset pricing framework. We demonstrate that the tails of the stationary distribution of the price–dividend ratio will follow a power law.

1. Introduction

Dynamic stochastic models have at times difficulty matching some features of macroeconomic data.[2] One route to reconcile differences between data and theory has been to replace the assumption of rational expectations with that of adaptive learning, in which agents are assumed to estimate the underlying parameters of a model via recursive least squares. For example, if the monetary authority adaptively learns the underlying Phillips curve via decreasing gain least squares regressions, then the high inflation (Nash) outcome is the one deemed stable (see Evans and Honkapohja, 2001). Still, the U.S. economy escaped the high inflation of the 1970's predicted by the standard decreasing gain model. To provide an explanation Sargent (1999) and Cho et al. (2002) assume instead that a monetary authority estimates a misspecified Phillips curve using a constant gain algorithm that places more weight on recent observations. This assumption allows the possibility of escape from a Nash outcome to a low inflation (Ramsey) outcome. In particular, within the context of their endogenous tracking model, a sequence of otherwise rare shocks can cause frequent large deviations from a high inflation self-confirming equilibrium. Indeed Sargent et al. (2006) take this endogenous tracking model to the data and account for the behavior of inflation in the U.S.

☆ We thank Chryssi Giannitsarou, In-Koo Cho, John Duffy, George Evans, Boyan Jovanovic, Tomasz Sadzik, Benoite de Saporta, Tom Sargent and two anonymous referees for helpful comments and suggestions. The usual disclaimer applies.

* Corresponding author.
E-mail addresses: benhabib@nyu.edu (J. Benhabib), cdave@nyu.edu (C. Dave).
[1] Fax: +1 212 995 4186.

[2] For example, empirical evaluations of consumption based asset pricing models lead to numerous asset pricing puzzles, and evaluations of real business cycle models cannot typically account for the pattern of hours worked without appealing to labor supply elasticities that are often at odds with microeconometric evidence.

1094-2025/$ – see front matter
http://dx.doi.org/10.1016/j.red.2013.09.004

Our analysis also focuses on the role of large deviations theory and its interplay with constant gain learning dynamics. Specifically, working within the adaptive learning tradition set out by Sargent and Williams (2005), Evans et al. (2010) and others, we examine the role of a generalized stochastic gradient constant gain (SGCG) learning algorithm in generating large deviations of an endogenous variable from its rational expectations value. We show analytically that these large deviations can occur with a frequency associated with a fat-tailed distribution even though the model is driven by thin-tailed exogenous stochastic processes. Using some new techniques in the analysis of stochastic processes and linear recursions with multiplicative noise,[3] we characterize these large deviations that occur under adaptive learning, driven by sequences of consistently low or consistently high shocks. Such sequences are rare in that the average of realizations in the sequences can significantly diverge from the population mean of the shocks. We then apply our model to the single asset version of the canonical model of Lucas (1978) that has been studied extensively by Carceles-Poveda and Giannitsarou (2007, 2008) who look at the ability of learning models to approximate the behavior of aggregate stock market data.

A particular issue in the modification of standard rational expectations models to better account for features of the data by introducing adaptive learning is the choice of the learning algorithm itself. Typically, in replacing the rational expectations assumption with that of adaptive learning, agents are assumed to estimate parameters of processes to be forecasted using recursive (adaptive) methods.[4] A particular strain of this literature demonstrates the consistency of this approach with Bayes' Law. In a stationary model with optimal learning, estimated parameters ultimately converge to their rational expectations equilibrium. In recent work however, Sargent and Williams (2005) introduce a model in which agents expect a random walk drift in estimated parameters. They then show that the SGCG algorithm, that assigns more weight to recent observations on account of the expected underlying drift in the estimated parameters, is asymptotically the optimal Bayesian estimator. Evans et al. (2010) follow Sargent and Williams (2005) and show how an SGCG learning algorithm approximates an optimal (in a Bayesian sense) Kalman filter. Under such adaptive SGCG learning, uncertainty about estimated parameters persists over time and can fuel escape dynamics in which a sequence of consistently high or consistently low shocks propel agents away from the Rational expectations Equilibrium (REE) of a model.[5] In an asset pricing context Weitzman (2007) also shows that if recent observations are given more weight under Bayesian learning of the variance of the consumption growth rate, agents will forecast returns and asset prices using thick-tailed distributions for consumption growth.[6] It is for this reason that we focus on an asset pricing context to analytically demonstrate how SGCG learning, consistent with optimal Bayesian learning, can account for the data features and fat-tailed distributions of the price–dividend ratio.

Theoretically, we demonstrate that under adaptive learning of the asset prices, the tails of the stationary distribution of the price–dividend ratio will follow a power law, even though the dividend process has thin tails and is specified as a stationary $AR(1)$ process. The tail index or power-law coefficient of the price–dividend ratio can be expressed as a function of model parameters, and in particular of the optimal gain parameter that assigns decaying weights to older observations. In fact, as demonstrated by Sargent and Williams (2005) and more recently by Evans et al. (2010), the optimal gain depends on the variance of the underlying drift in the estimated parameters: the higher the variance of the drift parameter, the higher the gain, and the thicker the tail of the distribution of the price–dividend ratio. We characterize how the power law tail index of the long-run stationary distribution of the price–dividend ratio varies as a function of the gain parameter and of the other deep parameters of the model. Under our adaptive learning scheme that approximates optimal Bayesian learning, stationary dividend processes generate distributions for the price–dividend ratio that are not Normal. Thus, large deviations of the price–dividend ratio from the rational expectations equilibrium are possible with a frequency higher than that associated with a Normal distribution even though the dividend process is thin tailed.

Our analysis and simulations indicate that under standard parameter calibrations, to match either the empirical tail index or the variance of the quarterly "fat-tailed" price–dividend ratio, we require a gain parameter around 0.1–0.3, significantly higher than what is typically used in the adaptive learning literature (0.01–0.04) in, for instance, the context of New Keynesian models. Carceles-Poveda and Giannitsarou (2008) also employ large values for the gain in asset pricing contexts, as do Branch and Evans (2010). In order to get an empirical handle on the parameters of our model, including the gain parameter, we estimate them by two separate methods. The first is a structural minimum distance estimation for the tail index and the first two moments of the price–dividends ratio. This method puts higher weight on the empirically observed tail of the price–dividend ratio, and produces a gain estimate in the range of 0.1–0.3. The second method computes the gain as Bayesian agents expecting drifting parameters would, using a Kalman filter on the data. This yields a gain parameter in the range of 0.3–0.45, assigning decaying weights on past observations that take the parameter drift into account. Therefore agents who use this gain parameter would indeed have their expectations confirmed by the data.

[3] See Kesten (1973), Saporta (2005) and Roitershtein (2007).

[4] In asset pricing contexts, see for example: Adam et al. (2008), Adam and Marcet (2011), Branch and Evans (2010), Brennan and Xia (2001), Bullard and Duffy (2001), Carceles-Poveda and Giannitsarou (2008), Cogley and Sargent (2008), and Timmermann (1993, 1996).

[5] See also Holmstrom (1999) for an application to managerial incentives of learning with an underlying drift in parameters.

[6] See also Koulovatianos and Wieland (2011). They adopt the notion of rare disasters studied by Barro (2009) in a Bayesian learning environment. They find that volatility issues are well addressed. Similarly Chevillon and Mavroeidis (2011) find that giving more weight to recent observations under learning can generate low frequency variability observed in the data. See also Gabaix (2009) who provides an excellent summary of instances in which economic data follow power laws and suggests a number of causes of such laws for financial returns. In particular, Gabaix et al. (2006) suggest that large trades in illiquid asset markets on the part of institutional investors could generate extreme behavior in trading volumes and returns (usually predicted to be zero in Lucas-type environments).