NBER WORKING PAPER SERIES

LEARNING, LARGE DEVIATIONS AND RARE EVENTS

Jess Benhabib
Chetan Dave

Working Paper 16816
http://www.nber.org/papers/w16816

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2011

We thank Chryssi Giannitsarou, In-Koo Cho, John Duffy, George Evans, Boyan Jovanovic, Tomasz Sadzik, Benoite de Saporta, and Tom Sargent for helpful comments and suggestions. The usual disclaimer applies. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Learning, Large Deviations and Rare Events
Jess Benhabib and Chetan Dave
NBER Working Paper No. 16816
February 2011, Revised July 2011
JEL No. D83,D84

ABSTRACT

We examine the asymptotic distribution of the price-dividends ratio in a standard asset pricing model when agents learn adaptively using a constant gain stochastic gradient algorithm. The asymptotic distribution is characterized using techniques from linear recursions with multiplicative noise, and is shown to exhibit fat tails even though dividends follow a standard stationary AR(1) process with thin tails. We demonstrate our results theoretically and empirically match the model to S & P 500 and CRSP data.

Jess Benhabib
Department of Economics
New York University
19 West 4th Street, 6th Floor
New York, NY 10012
and NBER
jess.benhabib@nyu.edu

Chetan Dave
New York University
Department of Economics
19 W. 4th Street, 6FL
New York, NY, 10012
cdave@nyu.edu

1. Introduction

Figures 1-2 respectively plot annualized monthly data for the S & P 500 and quarterly CRSP data for aggregate stock prices and dividends in the U.S. The plots show that prices and dividends do move in tandem, as predicted by standard theory. However the price-dividend ratios, shown in the third panel of each Figure, exhibit large fluctuations, especially in the latter parts of the sample.[1] These large fluctuations in the price to dividend ratios are difficult to explain within the context of the standard rational expectations asset pricing model, for example that of Lucas (1978).

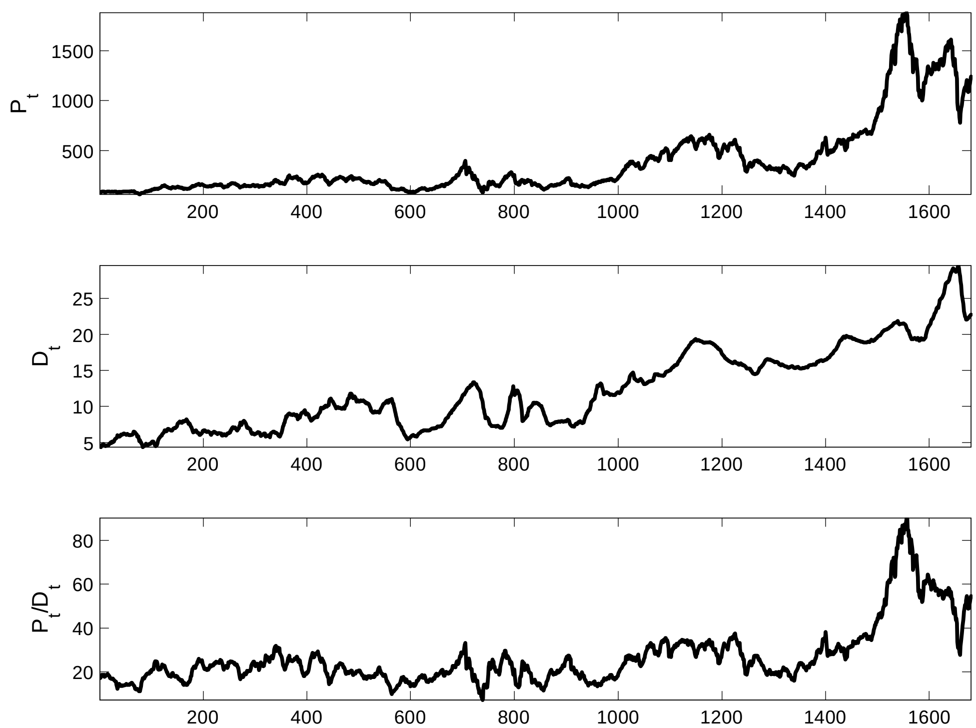

Figure 1. Monthly S & P 500 (Source: Shiller (2005)).

[1] The Data Appendix provides details on the series employed.

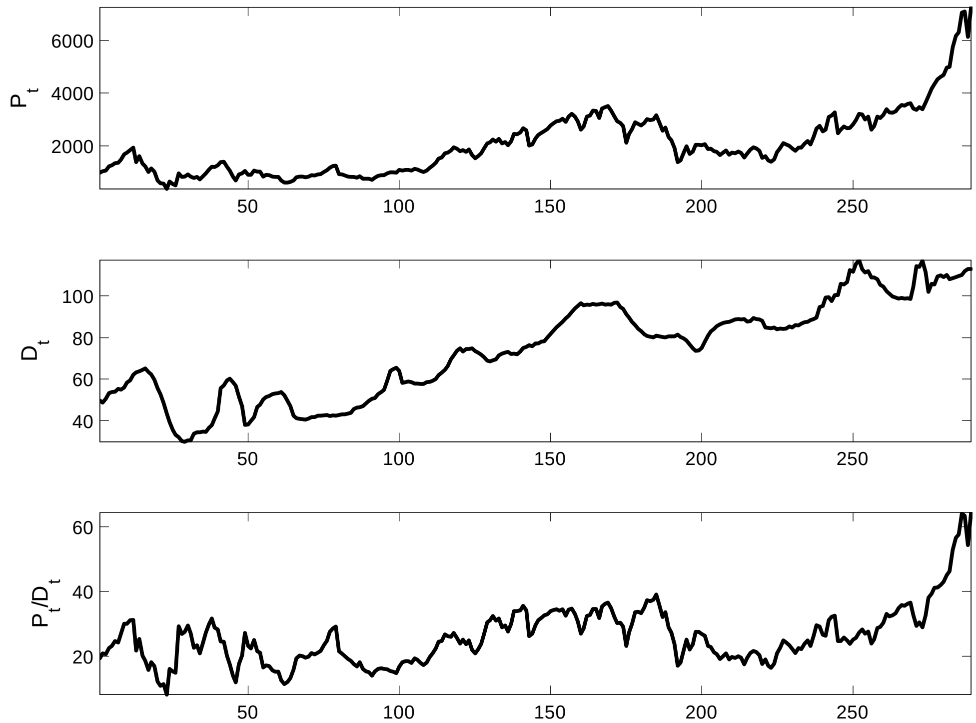

Figure 2. Quarterly CRSP (Source: Campbell (2003)).

Several modifications of the seminal model have been proposed to account for this (and other) departures of the data from model. A particular formulation replaces the rational expectations assumption with that of adaptive learning in which agents are assumed to estimate parameters of processes to be forecasted using recursive (adaptive) methods.[2] In this paper we demonstrate, theoretically and empirically, that one particular form of learning, stochastic gradient adaptive learning with constant gain, can account for these data features.

Theoretically, we demonstrate that under adaptive learning the tails of the stationary distribution of the price dividends ratio follow a power law. We show via simulations how the coefficient characterizing the power law varies as a function of the deep parameter values. Thus under learning a stationary dividends process generates a distribution for the price-

[2]See Marcet and Sargent (1989), Woodford (1990) and Evans and Honkapoja (1999, 2001).

dividends ratio that is not Normal. Rare large shocks to exogenous dividends throw off the persistent learning process and lead to large deviations from the rational expectations equilibrium of the price-dividends ratio.

Similar results have been obtained through simulations by Sargent (1999) and Cho, Williams and Sargent (2002) in monetary policy contexts. We build on that literature by demonstrating that for the asset pricing model, large deviations are possible under a stochastic gradient constant gain (SGCG) learning algorithm.[3] Empirically, we estimate the power law coefficient for the price-dividends ratio in the data. We then estimate the deep parameters, including the constant gain coefficient, by minimizing the squared deviation between the empirical and model predicted power law coefficients; a minimum distance structural estimation exercise. We find that not only does the data exhibit fat tails, but the model predicts this behavior for reasonable deep parameter values and an estimated constant gain parameter.

The literature that incorporates learning into asset pricing environments is large and rich with both theoretical and empirical results. Carceles-Poveda and Giannitsarou (2008) study the ability of asset pricing models to match data features under a variety of learning algorithms. They find that SGCG algorithms have difficulty matching data volatility under their calibrations. Adam, Marcet and Nicolini (2008) study a nonlinear environment in which the representative agent learns via recursive least squares to forecast the growth of

[3]Sargent (1999) and Cho et al. (2002) study recursive least squares constant gain (RLSCG) learning algorithms. Under RLSCG learning uncertainty about estimated parameters persists and fuels 'escape' dynamics in which a sequence of rare and unusual shocks propel agents away from the REE of a model (see also Williams (2009)). Sargent and Williams (2005) further incorporate random walk drift for estimated parameters, so that uncertainty about parameters persists over time. They then show that the generalized constant gain stochastic gradient (SGCG) algorithm is the optimal Bayesian estimator in that case. Evans et al. (2010) follow Sargent and Williams (2005) and show how a SGCG learning algorithm approximates an optimal (in a Bayesian sense) Kalman filter. We therefore restrict attention to SGCG learning algorithms.

asset prices. They find that the second moments of the data are well matched. In contrast we study an environment in which the agent learns by using a SGCG recursion about deviations from steady state, and we show that, for reasonable parameter values, the distribution of price-dividends ratio is a power law that has at most a few moments.

In an emerging literature, some authors have also incorporated the notion of rare disasters directly into learning models of asset pricing. For instance, Koulovatianos and Wieland (2011) adopt the notion of rare disasters studied by authors such as Barro (2009) in a Bayesian learning environment. They find that volatility issues are well addressed. We demonstrate that even with a regular stationary dividend process (such as a stationary $AR(1)$), the endogenous variable can still exhibit large deviations and fat tails. That is, under SGCG learning, the *equilibrium evolution* of the price-dividends ratio may follow a fat tailed distribution that matches the resulting model to data. This may briefly be described as a stochastic process with 'thin tails in, but fat tails out.' The reason is that under the SGCG algorithm, new dividend realizations can stochastically amplify forecast errors and throw off the price dividend ratio away from its rational expectations equilibrium.[4] Gabaix (2009) 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, Gopikrishnan, Plerou and Stanley (2006) suggest that large trades in illiquid asset markets on the part of institutional investors could generate extreme behavior in trading volumes (usually predicted to be zero in Lucas-type environments) and returns.

The paper is structured as follows. We first describe the single asset pricing version of Lucas (1978) under learning. We then prove in Section 3 that the model, written as a random

[4]Constant gain algorithms limit the horizon of learning of the agent (see Cho et al. (2002)).

linear recursion, predicts that the tails of the stationary distribution of the price-dividends ratio will follow a power law with coefficient κ. In Section 4 we provide estimates of the deep parameters that are consistent with the κ estimated from the price-dividends ratio plotted in Figures 1 and 2 above. Finally, we use simulations to study how κ varies with the deep parameters using as a baseline parametrization the estimates we obtain; Section 6 concludes.

2. Learning and Asset Pricing

Consider a discrete time single asset, endowment economy following Lucas (1978) with utility over consumption given by

$$u(C_t) = \frac{C_t^{1-\gamma}}{1-\gamma}, \ \gamma > 0. \tag{1}$$

Under a no-bubbles condition the nonlinear pricing equation is

$$P_t = E_t \left\{ \beta \left(\frac{D_{t+1}}{D_t} \right)^{-\gamma} (P_{t+1} + D_{t+1}) \right\} \tag{2}$$

where $\beta \in (0,1)$ is the usual exponential discount factor and (real) dividends (D_t) follow some exogenous stochastic process. Linearizing the above equation yields

$$p_t = \beta E_t(p_{t+1}) + (1 - \beta - \gamma) E_t(d_{t+1}) + \gamma d_t \tag{3}$$

where all lowercase variables denote log-deviations from the steady state $(\overline{P}, \overline{D}) = \left(\frac{\delta}{1-\delta}, 1\right)$. We assume that the exogenous dividends process follows

$$d_t = \rho d_{t-1} + \varepsilon_t, \ |\rho| < 1 \tag{4}$$

in which ε_t is an $iid(0, \sigma^2)$ random variable (such that $\sigma^2 < +\infty$) with compact support $[-a, a]$, $a > 0$, and a non-singular distribution function F.[5] Since $E_t(d_{t+1}) = \rho d_t$

$$p_t = \beta E_t(p_{t+1}) + \theta d_t, \ \theta \equiv (1 - \beta - \gamma)\rho + \gamma \tag{5}$$

is the fundamental expectational difference equation for prices. The rational expectations solution to (5) is

$$p_t = \phi^{REE} d_t, \ \phi^{REE} = \frac{\theta}{1 - \beta\rho} \tag{6}$$

for all $\beta\rho \neq 1$.

We follow Evans and Honkapohja (1999, 2001) and assume the perceived law of motion (PLM) on the part of the representative agent is

$$p_t = \phi_{t-1} d_{t-1} + \xi_t, \ \xi_t \sim i.i.d.(0, \sigma_\xi^2), \ \sigma_\xi^2 < +\infty \tag{7}$$

which in turn implies

$$E_t(p_{t+1}) = \phi_{t-1} d_t. \tag{8}$$

[5]The distribution function F is non-singular with respect to the Lebesque measure if there exists a function $f \in R_+$, $\int_R f(t)dt > 0$, such that $F(dt) \geq f(t)dt$.

where ϕ_{t-1} is the coefficient that agents estimate from the data to forecast p_t. Inserting the above into (5) yields the actual law of motion (ALM) under SGCG learning[6]:

$$\begin{aligned} p_t &= \beta\phi_{t-1}d_t + \theta d_t = (\beta\phi_{t-1} + \theta)d_t & (9) \\ &= (\beta\phi_{t-1} + \theta)\rho d_{t-1} + (\beta\phi_{t-1} + \theta)\varepsilon_t & (10) \end{aligned}$$

By contrast the ALM under rational expectations is

$$p_t = \phi d_t = \phi\rho d_{t-1} + \phi\varepsilon_t. \quad (11)$$

Under SGCG learning, ϕ_t evolves as[7]

$$\phi_t = \phi_{t-1} + g d_{t-1}(p_t - \phi_{t-1}d_{t-1}),\ g \in (0, 1) \quad (12)$$

Following the usual practice in the literature for analyzing learning asymptotics, we insert the ALM under learning in place of p_t in the recursion for ϕ_t in (12) to obtain

$$\begin{aligned} \phi_t &= \lambda_t\phi_{t-1} + \psi_t & (13) \\ \lambda_t &= 1 - (1 - \rho\beta)gd_{t-1}^2 + \beta g d_{t-1}\varepsilon_t = 1 - gd_{t-1}^2 + g\beta d_t d_{t-1} & (14) \\ \psi_t &= \theta\rho g d_{t-1}^2 + \theta g d_{t-1}\varepsilon_t = \theta g d_t d_{t-1}. & (15) \end{aligned}$$

[6] We note that in the asset pricing context, the ALM is linear in the 'belief' parameter (ϕ_t). In other contexts the ALM might be nonlinear in beliefs. However, the linear forces generating large deviations in the adaptive learning model may drive the dynamics in nonlinear contexts. For example in Cho et al. (2002) adaptive learning leads to non-neglible probablities for large deviations even in the prescence of nonlinearities for the true data generating process.

[7] See Carceles-Poveda and Giannitsarou (2007, 2008) for details and derivations under a variety of learning algorithms.

The equation in (13) takes the form of a linear recursion with both multiplicative (λ_t in (14)) and additive (ψ_t in (15)) noise. We show that the tail of the stationary distribution of ϕ_t follows a power law and can have fat tails. We characterize the tail of the distribution and show that under learning the the price-dividend ratio can exhibit large deviations from its rational expectations equilibrium value with non-negligible probabilities. We provide that theoretical analysis in the next section.

3. Large Deviations and Rare Events

We begin by noting that λ_t is a random variable, generating multiplicative noise, and can be the source of large deviations and fat tails for the stationary distribution of ϕ_t. We use results from large deviation theory (see Hollander (2000)) together with the work of Saporta (2005), Roitershtein (2007) and Collamore (2009) to characterize the tail of the distribution of ϕ_t.[8]

Let $\mathbb{N} = 0, 1, 2...$ We first note that the stationary $AR(1)$ Markov chain $\{d_t\}_{t\in Z}$ given by (4) is uniformly recurrent, and has compact support $\left[\frac{-a}{1-\rho}, \frac{a}{1-\rho}\right]$ (see Nummelin (1984), p. 93). We denote the stationary distribution of $\{d_t\}_{t\in\mathbb{N}}$ by π. Since $\{d_t\}_{t\in\mathbb{N}}$ and ε_t for $t = 1, 2...$ are bounded, so are $\{\lambda_t\}_{t\in\mathbb{N}}$ and $\{\psi_t\}_{t\in\mathbb{N}}$. In fact, following the first definition of Roitershtein (2007), $\{\lambda_t, \psi_t\}_{t\in\mathbb{N}}$ constitutes a Markov Modulated Process (MMP): conditional on d_t, the

[8]For an application of these techniques to the distribution of wealth see Benhabib et al. (2011) and to regime switching, Benhabib (2010).

evolution of the random variables $\lambda_{t+1}(d_t, d_{t-1})$ and $\psi_{t+1}(d_t, d_{t-1})$ are given by

$$
\begin{aligned}
P(d_t \in A, (\lambda_t, \psi_t) \in B) &= \int_A K(d, dy)\, G(d, y, B)\,|_{d=d_{t-1}}, &(16)\\
G(d, y, \cdot) &= P((\lambda_t, \psi_t) \in \cdot) \mid d_{t-1} = d,\ d_t = y), &(17)
\end{aligned}
$$

where $K(d, dy)$ is the transition kernel of the Markov chain $\{d_t\}_{t\in\mathbb{N}}$.

Next we seek restrictions on the support of the *iid* noise $\varepsilon_t \in [-a, a]$ to assure that $E|\lambda_\infty| < 1$ where, from equation (14), λ_∞ is the random variable associated with the stationary distribution of d_t . We assume:

$$
a < \left(\frac{6(1-\rho^2)}{g(1-\beta\rho)}\right)^{0.5}. \tag{18}
$$

Note that

$$
\begin{aligned}
E(\lambda_t) &= E\left(1 - g(d_{t-1})^2 + g\beta(d_{t-1}(\rho d_{t-1} + \varepsilon_t))\right)\\
E(\lambda_t) &= 1 - gE(d_{t-1})^2 + g\beta\rho E(d_{t-1})^2\\
E(\lambda_\infty) &= \left(1 - gE(d_{t-1})^2 (1-\beta\rho)\right)_{t\to\infty}
\end{aligned}
$$

Since ε_t is *iid* and is uniform with variance σ^2,

$$
\begin{aligned}
E(\lambda_\infty) &= 1 - g\frac{\sigma^2}{1-\rho^2}(1-\beta\rho) &(19)\\
E(\lambda_\infty) &= 1 - g\frac{\frac{1}{12}(2a)^2}{1-\rho^2}(1-\beta\rho) &(20)
\end{aligned}
$$

From equation (20) it follows that $E(\lambda_\infty) < 1$, and solving for a such that $E(\lambda_\infty) > -1$, we obtain the restriction (18) to guarantee that $E\left|\lambda_\infty\right| < 1$.

Let $S_n = \sum_{t=1}^n \log|\lambda_t|$. Following Roitershtein (2007) and Collamore (2009)[9] the tail of the stationary distribution of $\{\phi_t\}_t$ depends on the limit[10]

$$\Lambda(\delta) = \lim_{n\to\infty} \sup \frac{1}{n} \log E \prod_{t=1}^{n} |\lambda_t|^\delta = \lim_{n\to\infty} \sup \frac{1}{n} \log E[\exp(\delta S_n)] \ \forall\ \delta \in \mathbb{R}. \tag{21}$$

Using results in Roitershtein (2007), we can now prove the following about the tails of the stationary distribution of $\{\phi_t\}_{t\in\mathbb{N}}$:

Proposition 1 *For π-almost every $d_0 \in [-a, a]$, there is a unique positive $\kappa < \infty$ that solves $\Lambda(\delta) = 0$, such that*

$$K_1(d_0) = \lim_{\tau\to\infty} \tau^\kappa P(\phi > \tau|d_0) \text{ and } K_{-1}(d_0) = \lim_{\tau\to\infty} \tau^\kappa P(\phi < -\tau|d_0). \tag{22}$$

and $K_1(d_0)$ and $K_{-1}(d_0)$ are not both zero.[11]

Proof. The results follow directly from Roitershtein (2007), Theorem 1.6 if we show the

[9] For results on processes driven by finite state Markov chains see Saporta (2005).

[10] $\lim_{n\to\infty} \sup \frac{1}{n} \log E[\exp(\delta S_n)]$ is the Gartner Ellis limit that also appears in Large Deviation theory. For an exposition see Hollander (2000).

[11] *We can also show that $\pi(K_1(d_0) = K_{-1}(d_0)) = 1$ if a is large enough. This follows from Condition G given by Roitershtein (2007): Conditon G holds if there does not exist a partition of the irreducible set $D = \left\{d \subset \left(\frac{-a}{1-\rho}, \frac{a}{1-\rho}\right)\right\}$ into two disjoint sets D_{-1} and D_1 such that:*

$$\begin{aligned} P(d &\in D_{-1}, \rho d + \varepsilon \in D_1, \lambda < 0) \\ &= P(d \in D_{-1}, \rho d + \varepsilon \in D_{-1}, \lambda > 0) = 0 \end{aligned}$$

where $\varepsilon \subset [-a, a]$ and $\rho \subset (0, 1)$. (See Roiterstheinʼs Definition 1.7 and subsequent discussion, and his Proposition 4.1.) Suppose in fact that $P(d \in D_{-1}, \rho d + \varepsilon \in D_1, \lambda > 0) = 0$ for D_{-1} with minimal element d_0 and maximal element d_1. Then $P(d \in D_{-1}, \rho d + \varepsilon \in D_{-1}, \lambda > 0) = 1$. Then it must be true, since d_1 is the maximum element of D_{-1}, that $\rho d_1 + a \leq d_1$ and so $\frac{a}{1-\rho} \leq d_1$, implying $d_1 = \frac{a}{1-\rho}$. Similarly, it must be true that $\rho d_0 - a \geq d_0$ so that $\frac{-a}{1-\rho} \geq d_0$, implying $\frac{-a}{1-\rho} \geq d_0$. Thus $D_{-1} = D$, that is the whole set. Now we following:

(i) There exists a δ_0 such that $\Lambda(\delta_0) < 0$. First we note that $\Lambda(0) = 0$ for all n. Note also that

$$\begin{aligned}
\Lambda'(0) &= \lim_{n\to\infty} \sup \frac{1}{n} \frac{d \log E \prod_{t=1}^{n} |\lambda_t|^{\delta}}{d\delta} |_{\delta=0} \\
&= \lim_{n\to\infty} \sup \frac{1}{n} \left(E \prod_{t=1}^{n} |\lambda_t|^{\delta} \right)^{-1} E \left(\prod_{t=1}^{n} |\lambda_t|^{\delta} \log \prod_{t=1}^{n} |\lambda_t| \right) |_{\delta=0} \\
&= \lim_{n\to\infty} \sup \frac{1}{n} E \log \prod_{t=1}^{n} |\lambda_t|
\end{aligned}$$

For large n, as $\{\lambda_t\}_t$ converges to its stationary distribution ω, we have

$$\Lambda'(0) = \lim_{n\to\infty} \sup \frac{1}{n} \log E \prod_{t=1}^{n} |\lambda_t| = E_\omega \left(\log |\lambda_\infty| \right)$$

From equations (18)-(20) we have $E_\omega |\lambda_\infty| < 1$. Therefore $\Lambda'(0) = E_\omega \log(|\lambda_\infty|) < 0$, and there exists $\delta_0 > 0$ such that $\Lambda(\delta_0) < 0$.

(ii) There exists a δ_1 such that $\Lambda(\delta_1) > 0$. As in (i) above, we can evaluate, using Jensen's

can show that for a large enough, $P(d \in D, \rho d + \varepsilon \in D, \lambda > 0) = 1$ cannot hold. Since

$$\lambda = 1 - g(d_0)^2 + g\beta d_0 (\rho d_0 + \varepsilon) = 1 - g\left(d_0^2\right)(1 - \rho\beta) + g\beta d_0 \varepsilon,$$

we attain the smallest possible λ if we set $d_0 = \frac{a}{1-\rho}$ and $\varepsilon = -a$, or equivalently $d_0 = \frac{-a}{1-\rho}$ and $\varepsilon = a$. Then $\lambda \geq 0$ with probability 1 if and only if $a \leq \bar{a} = \frac{(1-\rho)}{(g(1+\beta(1-2\rho)))^{0.5}}$. If $a > \bar{a}$ with positive probability, then $P(\lambda < 0) > 0$, which contradicts $P(d \in D_{-1}, \rho d + \varepsilon \in D_{-1}, \lambda > 0) = 1$. Note also that $\lambda = 1$ for $d_0 = 0$ so it also follows that the $P(\lambda > 0) > 0$.

inequality,

$$\Lambda(\delta) = \lim_{n\to\infty}\sup\frac{1}{n}\log E\prod_{t=1}^{n}|\lambda_t|^{\delta} = \lim_{n\to\infty}\sup\frac{1}{n}\log E[\exp(\delta S_n)] \tag{23}$$

$$= \lim_{n\to\infty}\sup\log\left(E[\exp(\delta S_n)]\right)^{\frac{1}{n}} \geq \lim_{n\to\infty}\sup\log\left(E[\exp(\delta\frac{S_n}{n})]\right) \tag{24}$$

so that at the stationary distribution of $\{\lambda_t\}_{t\in\mathbb{N}}$

$$\Lambda(\delta) \geq \log E_{\omega}[\exp(\delta\log|\lambda_{\infty}|)] = \log\int_{\lambda}[\exp(\delta\log|\lambda_{\infty}|)]d\omega(\lambda)\,. \tag{25}$$

As $\delta\to\infty$ for $\log|\lambda| < 0$ we have $[\exp(\delta\log|\lambda_t|)] \to 0$, but if $P_{\omega}(\log|\lambda| > 0) > 0$ at the stationary distribution of $\{\lambda_t\}_t$, then $\lim_{\delta\to\infty}\ \Lambda(\delta) = \log\int_{\lambda}[\exp(\delta\log|\lambda_t|)]d\omega(\lambda) \to \infty$. Therefore if we can show that $P_{\omega}(\log|\lambda_t| > 0) > 0$, it follows that there exists a δ_1 for which $\Lambda(\delta_1) > 0$. Since $\Lambda(\delta)$ is convex[12], it follows that there exists a unique κ for which $\Lambda(\kappa) = 0$. To show that $P_{\omega}(|\lambda| > 1) > 0$, define $A = \left\{d \in \left(0, \frac{\mu a\beta}{1-\rho\beta}\right)\right\}$, $\mu \in (0,1)$ so that $\frac{\mu a\beta}{1-\rho\beta} < \frac{a}{1-\rho}$. At its stationary distribution $\{d_t\}_{t\in\mathbb{N}}$ is uniformly recurrent over $\left[\frac{-a}{1-\rho}, \frac{a}{1-\rho}\right]$ which implies that $P_{\pi}(d_{t-1}\in A) > 0$. We have $\lambda_t = 1-\beta g d_{t-1}\left(\beta^{-1}(1-\rho\beta)d_{t-1}-\varepsilon_t\right)$, so for $d_{t-1}\in A$ and $\varepsilon_t \in (\mu a, a]$, it follows that $\lambda_t > 1$. Thus $P_{\omega}(|\lambda_t| > 1) = P_{\pi}(d_{t-1}\in A_t)\,P(\varepsilon_t \in (\mu a, a]) > 0$.

(iii) The non-arithmeticity assumption required by Roitershtein (2007) (p. 574, (A7)) holds[13]: There does not exist an $\alpha > 0$ and a function $G : \mathcal{R}\times\{-1,1\} \to R$ such that

$$P\left(\log|\lambda_t| \in G(d_{t-1},\eta) - G(d_t, \eta\cdot sign(\lambda_t)) + \alpha\mathbb{N}\right) = 1. \tag{26}$$

[12] This follows since the moments of nonnegative random variables are log convex (in δ); see Loeve (1977, p. 158).

[13] See also Alsmeyer (1997). In other settings $\{\lambda_t\}_t$ may contain additional *iid* noise independent of the Markov Process $\{d_t\}_t$, in which case the non-aritmeticity is much more easily satisfied.

We have

$$\log|\lambda_t| = \log\left|(1 - gd_{t-1}^2 + g\beta d_t d_{t-1})\right| = \log\left|\left(1-(1-\rho\beta)gd_{t-1}^2 + \beta g d_{t-1}\varepsilon_t\right)\right| = F\left(d_{t-1},\varepsilon_t\right), \tag{27}$$

which contains the cross-partial term $d_t d_{t-1}$. Therefore in general $F\left(d_{t-1},\varepsilon_t\right)$ cannot be represented in separable form as $R\left(d_{t-1},\eta\right) - R\left(d_t,\eta\right) + \alpha\mathbb{N}\ \ \forall\left(d_{t-1},d_t\right)$ where $d_t = \rho d_{t-1} + \varepsilon_t$. Suppose to the contrary that there is a small rectangle $[D, D^*] \times [E, E^*]$ in the space of (d,ε), over which λ remains of constant sign, say positive, such that $F(d,\varepsilon) = R(d) - R(\rho d+\varepsilon)$, d is in the interior of $[D, D^*]$, and ε is in the interior of $[E, E^*]$, up to a constant from the discrete set $\alpha\mathbb{N}$, which we can ignore for variations in $[D, D*] \times [E, E*]$ that are small enough. Now fix d, d' close to one another in the interior of $[D, D^*]$. We must have, for $\varepsilon \in [E+\rho|d-d'|,\ E^* - \rho|d-d'|]$, that

$$\begin{aligned} F(d,\varepsilon) - R(d) &= -R(\rho d+\varepsilon) = -R(\rho d' + \varepsilon + \rho(d-d')) && (28)\\ &= F(d', \varepsilon + \rho(d-d')) - R(d'), && (29) \end{aligned}$$

or $F(d,\varepsilon) - F(d',\varepsilon+\rho(d-d')) = R(d) - R(d')$. However the latter cannot hold since the cross-partial term $d_{t-1}\varepsilon_t$ in $F(d_{t-1},\varepsilon_t) = 1-(1-\rho\beta)gd_{t-1}^2 + \beta g d_{t-1}\varepsilon_t$ is non-zero except of a set of zero measure where d or ε are zero.[14,15]

(iv) To show that $K_1\left(d_0\right) = \lim_{\tau\to\infty}\ \tau^\kappa P(\phi > \tau|d_0)$ and $K_{-1}\left(d_0\right) = \lim_{\tau\to\infty}\ \tau^\kappa P(\phi < -\tau|d_0)$ are not both zero, we have to assure, since ψ_t and λ_t are not assumed to be inde-

[14] We thank Tomasz Sadzik for suggesting this proof for (iii).

[15] We can avoid possible degeneracies that may occur if λ_t and ψ_t have a specific form of dependence so that

$$P\left(\phi|\lambda_t\phi + \psi_t = \phi\right) = 1.$$

pendent, that ϕ is not a deterministic function of the initial d_{-1}. We invoke (a) and (c) of Proposition 8.1 in Roitersthein (2007): Condition 1.6, $\pi\left(K_1\left(d_0\right)+K_{-1}\left(d_0\right)=0\right)=1$, holds if and only if there exists there exists a measurable function $\Gamma:\left[\frac{-a}{1-\rho},\frac{a}{1-\rho}\right]\to R$ such that

$$P\left(\psi_0+\lambda_0\Gamma\left(\rho d_{-1}+\varepsilon_0\right)=\Gamma\left(d_{-1}\right)\right)=1.$$

However

$$\psi_0+\lambda_0\Gamma\left(\rho d_{-1}+\varepsilon_0\right)=\theta g d_{-1}\rho d_{-1}+\theta g d_{-1}\varepsilon_0+\left(1-g d_{-1}^2+g\beta d_{-1}\left(\rho d_{-1}+\varepsilon_0\right)\right)\Gamma\left(\rho d_{-1}+\varepsilon_0\right)$$

is a random variable that depends on ε_0 while $\Gamma\left(d_{-1}\right)$ is a constant, so

$$P\left(\psi_0+\lambda_0\Gamma\left(\rho d_{-1}+\varepsilon_0\right)=\Gamma\left(d_{-1}\right)\right)<1$$

and Condition 1.6 in Roitersthein (2007) cannot hold. Then from Roitersthein (2007) Proposition 1.8 (c), $K_1\left(d_0\right)$ and $K_{-1}\left(d_0\right)$ are not both zero.[16] ■

The Proposition above characterizes the tail of the stationary distribution of ϕ as a

Note

$$\begin{aligned}\phi &= \frac{\psi_t}{1-\lambda_t}=\frac{\theta\rho g d_t^2+\theta g d_t\varepsilon_{t+1}}{1-(1-\rho\beta)g d_t^2+\beta g d_t\varepsilon_{t+1}}\\ &= \frac{\theta}{\beta}\frac{\beta\rho g d_t^2+gbgd_t\varepsilon_{t+1}}{1-(1-\rho\beta)g d_t^2+\beta g d_t\varepsilon_{t+1}}\end{aligned}$$

Differentiating wrt ε_t, the right side is zero only if $\beta\rho g d_t^2=1-(1-\rho\beta)g d_t^2$, or $\beta\rho g=1-g+g\rho\beta$. This holds only if $g=1$. So in general, for any d_t, there exists a constant ϕ such that $P\left(\phi|\lambda_t\phi+\psi_t=\phi\right)=1$ only if $g=1$, which we ruled out by assumption.

[16] In models where the driving stochastic process is *iid* or is a finite stationary Markov chain, the exponent κ can be analytically derived using the results of Kesten (1973) and Saporta (2005). In the case where λ is *iid* in equation (13), κ solves $E\left(\lambda^\kappa\right)=1$. In the finite markov chain case, under appropriate assumptions, κ solves $\varsigma\left(PA^\kappa\right)=1$ where P is the transition matrix, A is a diagonal matrix of the states of the Markov chain assumed to be non-negative, and $\varsigma\left(PA^\kappa\right)$ is the dominant root of PA^κ.

power tail with exponent κ. It follows that the distribution of ϕ has moments only up to the highest integer less than κ, and is a 'fat tailed' distribution rather than a Normal. The results are driven by the fact that the stationary distribution of $\{\lambda_t\}_{t\in\mathbb{N}}$ has a mean less than one, which tends to induce a contraction towards zero, but also has support above 1 with positive probability, which tends to generate divergence towards infinity. The stationary distribution arises out of a balance between these two forces. Then large deviations as strings of realizations of λ_t above one, even though they may be rare events, can produce fat tails.

In the asset price model ϕ relates the dividends to assets prices. Under adaptive learning, the results above show how the probability distribution of large deviations, or 'escapes' of ϕ from its REE value is characterized by a fat tailed distribution, and will occur with higher likelihood than under a Normal.[17]

We now briefly discuss the case where $\{d_t\}_t$ is an $MA(1)$ process. Proposition 1 still applies and we obtain similar results to the $AR(1)$ case. Let

$$d_t = \varepsilon_t + \zeta\varepsilon_{t-1}, \ |\zeta| < 1, \ t = 1, 2... \tag{30}$$

Then at its stationary distribution $d_t \in [-a(1+\zeta), a(1+\zeta)]$. Under the PLM

$$p_t = \phi_{0t}\varepsilon_t + \phi_{1t}\varepsilon_{t-1}, \tag{31}$$

[17] In the model of Cho et al. (2002), the monetary authority has a misspecified Philips curve and sets inflation policy to optimize a quadratic target. The learning algorithm using a constant gain however is not linear in the recursively estimated parameters (the natural rate and the slope of the Philips curve).

after observing ε_t at time t but not ϕ_{1t+1}, the agents expect

$$E_t(p_{t+1}) = \phi_{0t}E_t(\varepsilon_{t+1}) + \phi_{1t}E_t(\varepsilon_t) = \phi_{1t}\varepsilon_t \tag{32}$$

Then the ALM is

$$p_t = \beta\phi_{1t}\varepsilon_t + \gamma(\varepsilon_t + \zeta\varepsilon_{t-1}) = [\beta\phi_{1t} + \gamma]\,\varepsilon_t + \gamma\zeta\varepsilon_{t-1}$$

and the REE is given by

$$\phi_0 = \gamma(1 + \beta\zeta) \tag{33}$$

$$\phi_1 = \gamma\zeta. \tag{34}$$

Under the learning algorithm in equation (**??**) we obtain

$$\phi_{1t} = \phi_{1t-1} + gd_{t-1}(p_t - \phi_{1t-1}d_{t-1}) \tag{35}$$

$$\phi_{1t+1} = \lambda_{t+1}\phi_{1t} + \psi_{t+1} \tag{36}$$

$$\lambda_{t+1} = 1 - gd_t^2 + g\beta\varepsilon_{t+1}d_t \tag{37}$$

$$\psi_{t+1} = g\gamma\varepsilon_{t+1}d_t + \gamma\zeta g d_t\varepsilon_t \tag{38}$$

It is straightforward to show that at the stationary distribution of $\{\lambda_t\}_t$, $E(\lambda_t) < 1$, and that $P(\lambda_t > 1) > 0$. It is also easy to check that $\lambda_t > 0$ if $a < ((1+\zeta)(1+\zeta-\beta))^{-0.5}$. With the latter restriction, it is easy to check that the other conditions in the proof of Proposition

1 are satisfied.

4. Empirics

We first check whether real world data on price-dividend ratios have fat tails. We use a maximum likelihood procedure following Clauset et al. (2009) to estimate κ associated with P_t/D_t for both S&P and CRSP dividend series plotted in Figures 1 and 2 above. The results provided in Table 1 below show fairly small values of κ for both series, suggesting that only the first few moments of P_t/D_t exist irrespective of the data source. Table 1 also reports the estimated persistence ρ under an $AR(1)$ specification for the two linearly detrended dividends series, alongside the average price-dividends ratio (P_t/D_t) and its standard deviation.

Table 1. Data Characteristics

	Monthly S & P 500	Quarterly CRSP
$\widehat{\kappa}$	3.5753	6.9779
$s.e.(\widehat{\kappa})$	0.1620	1.2008
$\widehat{\rho}$	0.9966	0.9747
$s.e.(\widehat{\rho})$	0.0021	0.0150
Mean (P_t/D_t)	26.5901	26.0271
Std. Dev (P_t/D_t)	13.7530	8.7663

Next we feed the actual S&P and CRSP dividend series into our learning model and estimate the parameters, $\vartheta = [g\ \gamma\ \beta\ \rho]$ by minimizing the squared difference between the empirical κ's reported in Table 1 and those generated by our model. That is, we implement a simulated minimum distance method to estimate ϑ as[18]

$$\min_{\vartheta} \, [\kappa - \kappa(\vartheta)]^2. \tag{39}$$

The minimization procedure proceeds as follows. For candidate parameterizations of ϑ we employ the S&P or CRSP series dividends d_t to calculate ϕ_t as per (13)-(15). The ALM (9) then produces a corresponding p_t series which in turn delivers a price-dividend ratio P_t/D_t. We then estimate the κ associated with the 'simulated' P_t/D_t, using the methods of Clauset et al. (2009) to produce the $\kappa(\vartheta)$. The minimization procedure searches over the parameter space of ϑ to implement (39). Table 2 below reports the estimates $(\widehat{\vartheta})$ and associated standard errors $(s.e.(\widehat{\vartheta}))$ for each of the the S&P or CRSP dividend series, as well as the κ associated with the estimated parameters.[19]

Table 2. Parameter Estimates

	Monthly S & P 500		Quarterly CRSP	
Parameter	$\widehat{\vartheta}$	$s.e.(\widehat{\vartheta})$	$\widehat{\vartheta}$	$s.e.(\widehat{\vartheta})$
g	0.4750	1.7273	0.6708	0.5281
γ	2.6504	0.6244	2.3720	2.2971
β	0.9816	24.3008	0.9801	0.5099
ρ	0.9792	0.0033	0.9788	7.4603
Associated κ	4.2376		5.9676	

[18] Minimization was conducted using a simplex method and standard errors were computed using a standard inverse Hessian method.

[19] Starting values for the minimization procedure were $\vartheta_0 = [0.5 \ 2.5 \ 0.95 \ 0.95]$.

The point estimates of g are plausible, although the standard errors are quite large in the case of the Monthly S & P 500 dataset. Carceles-Poveda and Giannitsaraou (2008) discuss plausible values of g, where under constant gains the decay in weights on past observations dating i periods back is given by $(1-g)^{i-1}$. For example, for quarterly observations a $g = 0.46$ corresponds to 15 years of learning, with periods beyond 15 years having practically zero weight. If we want learning to go back 20 years, then g becomes 0.37. By contrast, looking at standard deviations of the price-dividend ratios for the Lucas asset pricing model, Carceles-Poveda and Giannitsaraou (2008) report that the standard deviations generated by the rational expectations or the learning models are smaller than the standard deviations in the actual data by factors of about 20 to 50.[20]

5. Model Simulations and Comparative Statics

To explore how κ is related to the underlying parameters of our model, we can simulate the learning algorithm that updates ϕ, and then estimate κ using the tail index estimator employed in the previous section. We can then explore how our estimate of κ from simulated series varies as we vary parameters. We simulate 100 series (each of length 1680 as in the S&P series) for ϕ_t under the $AR(1)$ assumption for dividends with *iid* uniform shocks. We then feed the simulated series into the model to produce $\{p_t\}$ and $\{P_t/D_t\}$. We estimate κ for each simulation and produce an average κ.

Escapes or large deviations in prices will take place when sequences of large shocks to

[20] Note that our estimates match the parameter values used by Carceles-Poveda and Giannitsaraou (2008) in their simulations except for γ, the CRRA parameter, which they set equal to 1 while we have it at $\gamma = 2.5$. Note also that for the simulations of Figure 3 in the next section, κ drops dramatically with γ.

dividends throw off the learning process from the rational expectations equilibrium. Such escapes will be more likely if dividend shocks can produce values of λ_t above 1, as we can see from equations (35-38). We expect lower κ, or fatter tails, as the support of λ_t that lies above 1 gets larger.

In the $AR(1)$ case for dividends we have $\lambda_{t+1} = 1 - (1 - \rho\beta) g d_t^2 + \beta g d_t \varepsilon_{t+1}$. Given the stationary distribution of $\{d_t\}_t$ and that of $\{\varepsilon_t\}_t$, the support of λ_t above 1 unambiguously increases if β increases. Increasing ρ however can have an ambiguous effect: while the term $(1 - \beta\rho)$ declines and tends to raise λ_t, the support of the stationary distribution of $\{d_t\}_t$ gets bigger with higher ρ. This increases $(1 - \rho\beta) g d_t^2$ and reduces the support of λ that is above 1 for large realizations of d_t^2. Finally in our simulations we expect that decreasing g tends to shrink the support of λ_t that is above 1 so that κ increases with g: as the gain parameter decreases, the tails of the stationary distribution of $\{\phi_t\}$ get thinner.[21]

Given the parameter estimates in the previous section, we use the baseline parameterization, $(\rho, g, \beta, \gamma) = (0.98, 0.5, 0.95, 2.5)$. The restriction given by equation (18) implies a maximum value of $a = \hat{a} = 2.6243$, so for the baseline parametrization we set the baseline value of $a = 0.225$. We find that the average κ is 4.9172, the average (P_t/D_t) is 20.4989 and the average Std. Dev (P_t/D_t) is 12.6142. We then vary each element of $(\rho, g, \beta, \gamma, \alpha)$ while keeping the others at their baseline values. The results of varying each parameter around the baseline values are plotted in Figures 3 and 4 below.[22]

[21] This of course is in accord with the Theorem 7.9 in Evans and Honkapohja (2001). As the gain parameter $g \to 0$ and $tg \to \infty$, $\{\phi_t^g - \varkappa\}/g^{0.5}$ converges to a Gaussian variable where $\varkappa$ is the globally stable point of the associated ODE describing the mean dynamics. More generally, as $g \to 0$, the estimated coefficient under learning with gain parameter g, ϕ_t^g, converges in probability (but not uniformly) to $\varkappa$ for $t \to \infty$. However, there will always exist arbitrarily large values of t with ϕ_t^g taking values remote from $\varkappa$ (See Benveniste, Métivier and Priouret (1980), pp. 42-45). Note however that our characterization of the tail of the stationary distribution of $\{\phi_t\}_t$ and of κ is obtained for fixed $g > 0$.

[22] For all parameter values used to produce Tables 3 and 4, the restriction given by (18) is easily satisfied.

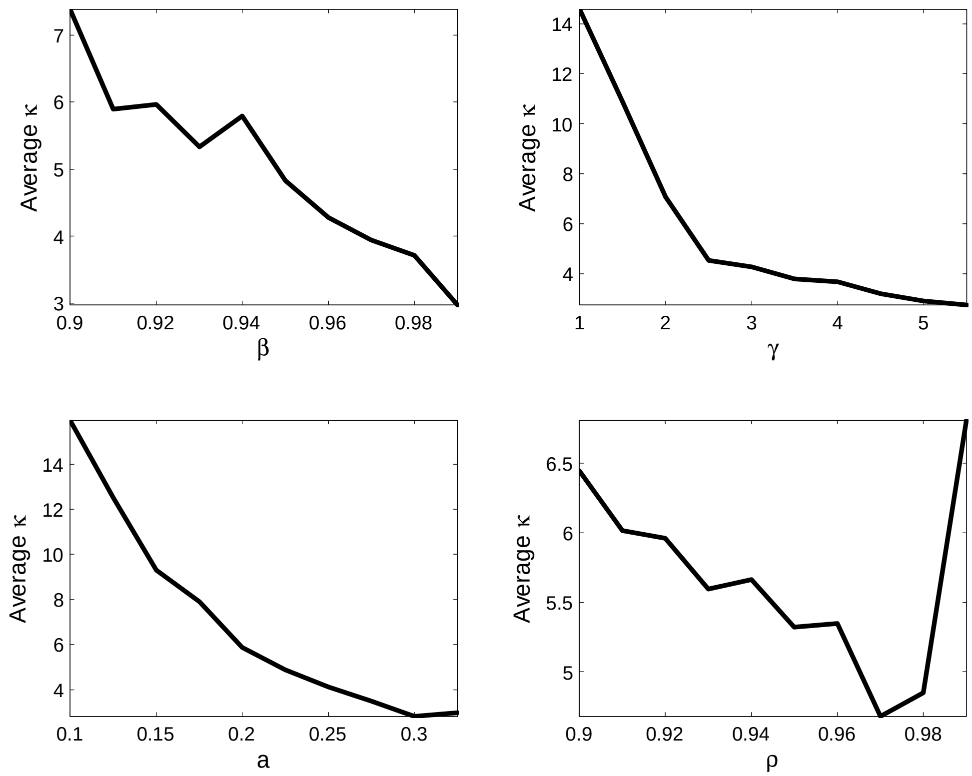

Figure 3. Simulation Results.

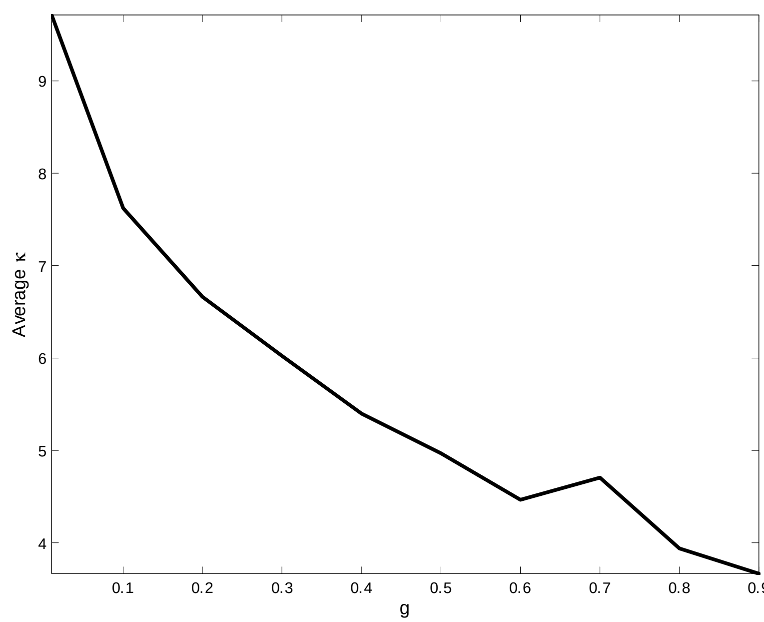

Figure 4. Simulation Results (cont'd.).

The simulation results confirm the notion that the average κ's should decline with β, γ

and a. The results with respect to ρ are non-monotonic, as expected. Figure 4 plots the results of the critical learning parameter g; it clearly demonstrates that as the learning gain falls, that is, the horizon for learning increases, average κ rises. However, for empirically plausible values of g the average κ is small. In summary, constant gains stochastic gradient learning leads to large deviations of (P_t/D_t) from its rational expectations value in response to rare large shocks in dividends.

6. Conclusion

An important and growing literature replaces expectations in dynamic stochastic models not with realizations and unforecastable errors, but with regressions where agents 'learn' the rational expectations equilibria. When such agents employ constant gain learning algorithms that put heavier emphasis on recent observations, escape dynamics can propel estimated coefficients away from the REE values. In an asset pricing framework 'bubbles,' or asset price to dividend ratios that exhibit large deviations from their REE values, can occur with a frequency associated with a fat tailed power law. The techniques used in our paper generalize to higher dimensions and to finite state Markov chains under certain assumptions,[23] and can be applied to other more general economic models.

References

[1] Adam, K., Marcet, A., and J. P. Nicolini. 2008. Stock Market Volatility and Learning, *European Central Bank Working Paper Series*, No. 862.

[23] See for example Saporta (2005) and Ghosh et al. (2010).

[2] Alsmeyer, G. 1997. The Markov Renewal Theorem and Related Results. *Markov Process Related Fields* 3 103–127.

[3] Barro, R. J. 2009. Rare Disasters, Asset Prices, and Welfare Costs, *American Economic Review*, 99:1, 243–264.

[4] Benhabib, J. Bisin, A. and S. Zhu, 2011. The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents. *Econometrica* 79, 123-158.

[5] Benhabib, J., 2010. A Note Regime Switching, Monetary Policy and Multiple Equilibria, *NBER Working Paper* No. 14770.

[6] Benveniste, A., Métivier, M. and P. Priouret, 1980. *Adaptive Algorithms and Stochasic Approximations,* Springer-Verlag, New York.

[7] Campbell, J. Y., 2003. Consumption-Based Asset Pricing, *Handbook of the Economics of Finance*, George Constantinides, Milton Harris, and Rene Stulz eds., North-Holland, Amsterdam.

[8] Carceles-Poveda, E., Giannitsarou, C., 2007. Adaptive Learning in Practice. *Journal of Economic Dynamics and Control* 31, 2659-2697.

[9] Carceles-Poveda, E., Giannitsarou, C., 2008. Asset pricing with adaptive learning. *Review of Economic Dynamics* 11 629–651.

[10] Cho, I-K., Sargent, T. J., Williams, N., 2002. Escaping Nash Inflation. *Review of Economic Studies* 69, 1-40.

[11] Clauset, A., Shalizi, C. R. and M. E. J. Newman, 2009. Power-law distributions in empirical data, *SIAM Review* 51(4), 661-703

[12] Collamore, J. F. , 2009. Random Recurrence Equations and Ruin in a Markov-Dependent Stochastic Economic Environment, *Annals of Applied Probability* 19, 1404–1458.

[13] Evans, G., Honkapohja, S., 1999. Learning Dynamics. *Handbook of Macroeconomics*, Vol.1, eds. J. Taylor and M. Woodford, 1999, North-Holland, pp.449-542.

[14] Evans, G., Honkapohja, S., 2001. *Learning and Expectations in Macroeconomics*. Princeton University Press.

[15] Evans, G., Honkapohja, S., and N. Williams, 2010. Generalized Stochastic Gradient Learning, *International Economic Review*, 51, 237-262.

[16] Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H. E. 2006. Institutional Investors and Stock Market Volatility, *Quarterly Journal of Economics*, 121 (2), p. 461-504.

[17] Gabaix, X., 2009. Power Laws in Economics and Finance, *Annual Review of Economics*, 1, p. 255-93.

[18] Goldie, C. M., 1991. Implicit Renewal Theory and Tails of Solutions of Random Equations, *Annals of Applied Probability*, 1, 126–166.

[19] Ghosh, A.P, Haya, D., Hirpara, H., Rastegar, R., Roitershtein, A.,Schulteis, A., and Suhe, J, (2010), Random Linear Recursions with Dependent Coefficients, *Statistics and Probability Letters* 80, 1597 1605

[20] Hollander, F. den, (2000), *Large Deviations*, Fields Institute monographs,American Mathematical Society, Providence, Rhode Island.

[21] Kesten, H., 1973. Random Difference Equations and Renewal Theory for Products of Random Matrices, *Acta Mathematica*. 131 207–248.

[22] Koulovatianos, C. and V. Wieland, 2011. Asset Pricing under Rational Learning about Rare Disasters, Manuscript.

[23] Loeve, M. 1977. Probability Theory, 4^{th} Ed., Springer, New York.

[24] Lucas, R. E. Jr., 1978. Asset Prices in an Exchange Economy, *Econometrica*, Vol. 46, No. 6. (Nov., 1978), pp. 1429-1445.

[25] Marcet, A, and Sargent, T. J., 1989. Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models, *Journal of Economic Theory*, 48, 337-368.

[26] Nummelin, E., 1984. *General irreducible Markov chains and non-negative operators.* Cambridge Tracts in Mathematics 83, Cambridge University Press.

[27] Roitershtein, A., 2007. One-Dimensional Linear Recursions with Markov-Dependent Coefficients, *The Annals of Applied Probability*, 17(2), 572-608.

[28] Saporta, B., 2005. Tail of the stationary solution of the stochastic equation $Y_{n+1} = a_n Y_n + \gamma_n$ with *Markovian Coefficients, Stochastic Processes and their Applications*, 115(12), 1954-1978.

[29] Sargent, T. J., 1999. *The Conquest of American Inflation.* Princeton University Press.

[30] Sargent, T. J. and Williams, N., 2005. Impacts of Priors on Convergence and Escape from Nash Inflation, *Review of Economic Dynamics,* 8(2), 360-391.

[31] Shiller, R. J., 2005. *Irrational Exuberance*, 2^{nd} edition, Broadway Books.

[32] Williams, N. 2009. Escape Dynamics, Manuscript.

[33] Woodford, M., 1990. Learning to Believe in Sunspots, *Econometrica*, 58(2), 277-307.

7. Data Appendix

1. Monthly S&P 500 Dataset

   (a) Download monthly data from http://www.econ.yale.edu/~shiller/data.htm, the Excel file is titled ie_data.xls

   (b) The following monthly time series are extracted/constructed for 1871.1 through 2010.12 from the above Excel file (note that $t = 1, \ldots T$ where $T = 2010.12$):

      i. Extract S & P Comp ($\widetilde{P}(t)$)

      ii. Extract Dividend ($\widetilde{D}(t)$)

      iii. Extract Consumer Price Index ($CPI(t)$)

      iv. Construct Real Price ($P(t)$) as $P(t) = [\widetilde{P}(t) \times \; CPI(T)]/CPI(t)$

      v. Construct Real Dividend ($D(t)$) as $D(t) = [\widetilde{D}(t) \times \; CPI(T)]/CPI(t)$

      vi. Construct the Price to Dividends Ratio (ratio) as $P(t)/D(t)$

2. Quarterly CRSP Dataset

(a) Download the quarterly data from http://scholar.harvard.edu/campbell/data, where the particular data being used is associated with "Replication Data for: Consumption Based Asset Pricing". The relevant file is titled USAQE.ASC note that this is effectively a CRSP dataset with the relevant variables being VWRETD and VWRETX. The text below is an extract from the explanations for this dataset on the above website.

(b) The following quarterly time series are extracted/constructed for 1926.1 through 1998.4 from the above dataset (note that $t = 1, \dots T$ where $T = 1998.4$):

i. Extract Col. 2: $\widetilde{P}(t)$. For each month, the price index is calculated as $\widetilde{P}(t) = (VWRETX(t) + 1) \times \widetilde{P}(t-1)$. (Note that time t in this equation is in months.) The price index for a quarter, as reported in this column, is the price index for the last month of the quarter. The original data, which goes up to 1996.4 was not altered. The new data, which goes up to 1998.4, was created as described here starting from 1997.1.

ii. Extract Col. 3: $\widetilde{D}(t)$. Dividend in local currency, calculated as follows. The dividend yield for each month is calculated as $\widetilde{DY}(t) = [1+VWRETD(t)]/[1+VWRETX(t)]-1$. Note that if the return index is calculated from $VWRETD$ as above, then this formula agrees with the formula for the dividend yield given earlier. As before, the dividend for each month is calculated as $\widetilde{D}(t) = \widetilde{DY}(t) \times \widetilde{P}(t)$. The dividend for a quarter, as reported in this column, is the sum of the dividends for the three months comprising the quarter.

iii. Extract the Consumer Price Index from Shiller's Data ($CPI(t)$) which is monthly and compute a quarterly average. Use that quarterly average for $CPI(t)$ in Campbell's data.

iv. Construct Real Price $(P(t))$ as $P(t) = [\widetilde{P}(t) \times \ CPI(T)]/CPI(t)$

v. Construct Real Dividend $(D(t))$ as $[\widetilde{D}(t) \times \ CPI(T)]/CPI(t)$ and then take running sums to get $D(t)$

vi. Construct the Price to Dividends Ratio (ratio) as $P(t)/D(t)$