Learning, Large Deviations and Rare Events

Jess Benhabib
NYU

Chetan Dave
NYU
The learning literature (e.g., Marcet and Sargent (1989), Woodford (1990) and Evans and Honkapohja (2001)) replaced expectations with regressions in dynamic stochastic models.

A main focus was to explore how ‘learning’ $\rightarrow$ convergence to rational expectations equilibria (REE) in dynamic stochastic economic models.
The idea that learning processes can generate volatility has been explored in the asset pricing literature (e.g. Timmerman (1993, 1996, 2007), Weitzman (2007), Adams, Marcet and Nicolini (2006)).

Sargent (1999) and Cho, Sargent and Williams (2002), Sargent and Williams (2005) introduced a new idea: Under recursive least squares constant gain (RLSCG) learning algorithms, uncertainty about estimated parameters can persist and fuel ‘escape’ dynamics in which a sequence of rare and unusual shocks propel agents away from the REE of a model (see also Williams (2009)): They suggested that this can be studied through "large deviations theory".

In the context of monetary policy with learning, Cho, Sargent and Williams (2002) showed via simulations that escape dynamics can fuel large deviations in inflation.
Sargent and Williams (2005) incorporated the expectation on the part of agents that the estimated parameters follow a random walk, so that uncertainty about parameters persists over time. (We will see that this is a natural assumption under adaptive learning) They then show that the generalized constant gain stochastic gradient (SGCG) algorithm is the optimal Bayesian estimator in that case. (See also Holmstrom (1999)). 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.

So "escapes" or "large deviations" in economic variables can take place when sequences of large shocks throw off the learning process from the rational expectations equilibrium.
Constant gains and optimal learning

- But how big should the constant gain be? It will depend on the underlying variances of the exogenous driving process (dividends), the perceived variance of the random walk, and the variance of the agent’s "perceived law of motion". Will the constant gain be self-confirming?
Introduction

- We characterize the limiting probabilities of escape dynamics and large deviations from a REE under Constant Gain Stochastic Gradient learning.

- **Context**: a univariate linear expectational difference equation
  - Encompasses asset pricing, overlapping generations models and others (see Evans and Honkapohja (1999)).

- **Finding**: Recursion for estimated coefficient can exhibit occasional large deviations (rare events) inducing a limiting ‘fat tailed’ power law distribution.
  - **Under Asset Pricing Interpretation**: Ratios of asset prices to dividends can exhibit volatility and deviate from their REE values.
Asset Prices
Figure 1. Monthly S & P 500, 1871.1 - 2010.12, (Source: Shiller).
Asset Prices

Figure 2. Quarterly CRSP, 1926.1 - 1998.4, (Source: Campbell (2003)).
The Model

Consider

\[ p_t = \delta E_t(p_{t+1}) + \gamma d_t, \quad \delta \in (0, 1). \tag{1} \]

with \( d_t \) an exogenous process

\[ d_t = \rho d_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad t = 1, 2... \tag{2} \]

where \( \varepsilon_t \sim i.i.d.(0, \sigma^2) \) with compact support \([-a, a], \ a > 0\).
Asset Pricing Interpretation: Consider a single asset in a Lucas (1978)-type economy, CRRA preferences parameterized by $\theta$, then

$$P_t = E_t \left\{ \delta \left( \frac{D_{t+1}}{D_t} \right)^{-\theta} (P_{t+1} + D_{t+1}) \right\}$$

yields

$$p_t = \delta E_t(p_{t+1}) + \gamma d_t, \quad \delta \in (0, 1), \quad \gamma \equiv (1 - \delta - \theta)\rho + \theta$$

for $p_t = \log(P_t) - \log(P)$ and $d_t = \log(D_t) - \log(D)$. 
The REE of the model in (1)-(2) is

$$p_t = \phi d_{t-1} + \eta_t, \quad \phi = \frac{\gamma \rho}{1 - \delta \rho}, \quad \forall \delta \rho \neq 1. \quad (3)$$

Under learning, agents assumed to be learning $\phi$. First they form a perceived law of motion (PLM)

$$p_t = \phi_{t-1} d_{t-1} + \xi_t \quad (4)$$

where $\xi_t$ is a regression error and estimates of $\phi_{t-1}$ will have a recursive form.

Given the PLM, agents assumed to form expectations as

$$E_t(p_{t+1}) = \phi_{t-1} d_t. \quad (5)$$
Next, insert $E_t(p_{t+1}) = \phi_{t-1} d_t$ into (1) to obtain actual law of motion (ALM)

$$p_t = (\delta \phi_{t-1} + \gamma) \rho d_{t-1} + (\delta \phi_{t-1} + \gamma) \epsilon_t,$$

(6)

Aside: Note how the fixed point of the map from PLM to the ALM ($\phi = (\delta \phi + \gamma) \rho = T(\phi)$) will deliver the REE (this is the $T$-map of Evans and Honkapohja (2001)).

Assumed timing within period $t$:

1. agent forms $E_t(p_{t+1})$ using $\phi_{t-1}$ and $d_t$
2. $p_t$ realized from ALM.
3. agent estimates $\phi_t$ using all data including $p_t$
The Model

- Recall the PLM

\[ p_t = \phi_{t-1} d_{t-1} + \xi_t, \quad \xi_t \sim iid(0, \sigma^2_{\xi}), \quad \sigma^2_{\xi} < +\infty \]  

(7)

- Given Sargent and Williams (2005) and Evans et al. (2010), we focus on a constant gain stochastic gradient learning algorithm to update \( \phi_t \) as

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

(8)

where \( g \) is the gain parameter. Note that under constant gains, from the perspective of agents \( \phi_t \) is a random walk and from the PLM \( (p_t - d_{t-1} \phi_{t-1}) \) is a zero mean forecast error.
The Model

What is $g$ if agents try to compute it? In the limit, as data accumulates, the optimal $g$ from the Kalman filter, if agents believe $\phi_{t-1}$ is a random walk

$$\phi_t = \phi_{t-1} + \Lambda_t, \quad \Lambda_t \sim iid(0, \sigma^2_{\Lambda}), \quad \sigma^2_{\Lambda} < +\infty,$$

then

$$g = g = \frac{\sigma_{\Lambda}\sigma_d}{\sigma_{\xi}}$$

where $\sigma_d$ is the variance of the stationary distribution of dividends $\{d\}$. We will come back to this point.
Usually, the ALM is inserted for $p_t$ in above to yield

$$
\phi_t = \phi_{t-1} + gH(\phi_{t-1}, d_{t-1})
$$

and then following Benveniste, Metivier and Priouret (1990) a stochastic approximation is conducted to show, as $g \to 0$, \{${\phi^g_t - \nu}$\} / $g^{0.5}$ converges to a Gaussian random variable (where $\nu$ is the stable point of the associated ODE, $\dot{\phi} = T(\phi) - \phi$ describing the ‘mean dynamics’ of (9)).

We instead cast the environment as one of random linear recursions, with associated techniques to characterize the tail of the stationary distribution of $\phi_t$ for fixed $g > 0$. 
Insert the ALM in place of $p_t$ in (8) to obtain

$$
\phi_{t+1} = \lambda_{t+1}\phi_t + \psi_{t+1}
$$

(10)

with

$$
\lambda_{t+1} = 1 - (1 - \rho \delta)gd_t^2 + \delta gd_t \varepsilon_{t+1} = 1 - gd_t^2 + g \delta d_{t+1}d_t
$$

and

$$
\psi_{t+1} = \gamma \rho gd_t^2 + \gamma gd_t \varepsilon_{t+1} = \gamma gd_{t+1}d_t.
$$

(11)

Now $\{\lambda_t\}$ generates multiplicative noise for $\{\phi_t\}$.

Can be source of large deviations and fat tails for stationary distribution of $\{\phi_t\}$.

Note: $\{\phi\}_t$ is not a random walk....
Consider

\[ y_{t+1} = \lambda y_t, \; \lambda \sim f(\lambda) \text{ w/ c.d.f } F(\lambda), \text{ over some } \mathbb{D}, \; y_t \geq y_0 \; \forall \; t. \]

Consider the density \( \pi(y_{t+1}) \): composed of \( \pi(y_t) \) and terms that would adjust for the stochastic multiplicative evolution. Overlooking the lower bound of \( y \), that is \( y_0 \), (which we should not, but this is a heuristic argument that will apply to the tail, for large \( \tau \), see below) we write

\[ \pi(y_{t+1}) = \pi(y_t) + \int_{\mathbb{D}} \pi \left( \frac{y_t}{\lambda} \right) dF(\lambda) - \int_{\mathbb{D}} \pi \left( y_t \right) dF(\lambda) \]

where the 2\textsuperscript{nd} term accounts for evolution to \( y_{t+1} \) from \( \frac{y_t}{\lambda} \) and the 3\textsuperscript{rd} term accounts for the evolution out of \( y_{t+1} \).
Recognizing that
\[
\int_{\mathcal{D}} dF(\lambda) = 1
\] (12)
implies
\[
\pi(y_{t+1}) = \int_{\mathcal{D}} \pi \left( \frac{y_t}{\lambda} \right) dF(\lambda).
\] (13)

Consider the stationary probability \( \pi(y_{t+1}) = \pi(y_t) \) and the task of extracting \( \pi(y) \), re-written without indices as
\[
\pi(y) = \int_{\mathcal{D}} \pi \left( y\lambda^{-1} \right) dF(\lambda).
\] (14)
Let \( \pi(y) = Ky^{-\kappa-1} \) with \((\kappa, K) \in \mathbb{R}_{++}\), and insert into the above:

\[
Ky^{-\kappa-1} = \int_{\mathcal{D}} Ky^{-\kappa-1} \lambda^{-(\kappa-1)} dF(\lambda) \quad (15)
\]

\[
Ky^{-\kappa-1} = Ky^{-\kappa-1} \int_{\mathcal{D}} \lambda^{-(\kappa-1)} dF(\lambda) \quad (16)
\]

\[
1 = \int_{\mathcal{D}} \lambda^{-(\kappa-1)} dF(\lambda) = \int_{\mathcal{D}} \lambda^{\kappa+1} dF(\lambda). \quad (17)
\]

Therefore, whatever the value of \( \kappa > 0 \) that solves

\[
E(\lambda^{\kappa+1}) \equiv \int_{\mathcal{D}} \lambda^{\kappa+1} dF(\lambda) = 1 \quad (18)
\]

will determine the value of \( \kappa \) in \( \pi(y) = Ky^{-\kappa-1} \).
Thus, for large $\tau$ at the stationary distribution of $\{y_t\}$, $P$, 

$$P(y > \tau) = \tau^{-\kappa} C > 0,$$

(19)

or

$$\tau^\kappa P(y > \tau) = C > 0.$$ 

(20)

Of course this is a heuristic argument that will work for the tail, as we ignored the lower bound $y_0$.

\[
\phi_{t+1} = \lambda_{t+1}\phi_t + \psi_{t+1} = \phi_1 \prod_{i=2}^{t+1} \lambda_i + \left( \psi_{t+1} + \sum_{j=1}^{t} \psi_j \prod_{i=j+1}^{t+1} \lambda_i \right)
\]

and assume \( \lambda_t > 0 \) has compact support with \( E(\lambda_t) < 1 \) and \( P(\lambda_t > 1) > 0 \) at its stationary dist. Then \( \exists \) a \( \kappa > 0 \) solving

\[
\Lambda(\beta) = \lim_{t \to \infty} \sup \frac{1}{t} \log E \left( \prod_{i=0}^{t} \lambda_i^\kappa \right) = 0
\]

\[
\Lambda(\beta) = \lim_{t \to \infty} \sup E \left( \prod_{i=0}^{t} \lambda_i^\kappa \right)^{\frac{1}{t}} = 1
\]

Then tails of the stationary distribution of \( \{\phi_t\} \) can be characterized by:

\[
\lim_{\tau \to \infty} \tau^\kappa P(\phi > \tau) = K_1(d_0) > 0 \quad \text{and} \quad \lim_{t \to \infty} \tau^\kappa P(\phi > -\tau) = K_{-1}(d_0) > 0
\]
As $\kappa \uparrow$, tail of the stationary distribution of $\{\phi_t\}_t$ thins. A (simulated) plot of $\Lambda(\beta)$ vs. $\beta$ is below, showing a particular $\kappa$

- At $\beta = 0$ we have the first moment ($\Lambda'(\beta)$) being $< 1$ ($E(\lambda_t) < 1$). We will show $\Lambda(\beta)$ crosses 1 at $\kappa < \infty$. 
Denote the stationary distribution of \( \{d_t\}_{t \in \mathbb{N}} \) by \( \pi \).

To ensure \( E|\lambda_\infty| < 1 \), we restrict support of \( \varepsilon_t \in [-a, a] \) by assuming

\[
a < \left( \frac{6 (1 - \rho^2)}{g (1 - \beta \rho)} \right)^{0.5}
\]

Let \( S_t = \sum_{i=1}^{t} \log \lambda_i \), and

\[
\Lambda(\beta) = \lim_{t \to \infty} \sup \frac{1}{t} \log E \prod_{i=1}^{t} \lambda_i^\beta
\]

\[
= \lim_{t \to \infty} \sup \frac{1}{t} \log E[\exp(\beta S_t)] \forall \beta \in \mathbb{R}.
\]
Proposition

For $\pi$-almost every $d_0 \in [-a, a]$, there is a unique positive $\beta < \infty$ that solves $\Lambda(\beta) = 0$, and the following limits exist and are positive:

$$K_1(d_0) = \lim_{\tau \to \infty} \tau^\beta P(\phi > \tau | d_0) \quad \text{and}$$

$$K_{-1}(d_0) = \lim_{\tau \to \infty} \tau^\beta P(\phi < -\tau | d_0).$$
Part (i): ∃ a $\beta_0$ s.t. $\Lambda(\beta_0) < 0$. Note that $\Lambda(0) = 0$ and as $
abla \lambda \nabla$ converges to its stationary distribution $\omega$, we can derive

$$\Lambda'(0) = \lim_{t \to \infty} \sup \frac{1}{E} \prod_{i=1}^{t} \lambda_i = E_{\omega}(\log \lambda_{\infty})$$

But we can show that

$$E_{\omega}(\lambda_{\infty}) = 1 - g \frac{\sigma^2}{1 - \rho^2} (1 - \delta \rho) < 1$$

$\therefore \Lambda'(0) = E_{\omega} \log(\lambda_{\infty}) < 0$ by Jensen’s inequality and $\exists$ a $\beta_0 > 0$ s.t. $\Lambda(\beta_0) < 0$. 
Part (ii): \( \exists \beta_1 \) s.t. \( \Lambda(\beta_1) > 0 \). As in (i) above, using Jensen’s inequality,

\[
\Lambda(\beta) = \lim_{t \to \infty} \sup \frac{1}{t} \log E \prod_{i=1}^{t} \lambda_i^\beta
\]

\[
= \lim_{t \to \infty} \sup \frac{1}{t} \log E[\exp(\beta S_t)]
\]

\[
= \lim_{t \to \infty} \sup \log \left( E[\exp(\beta S_t)] \right)^{\frac{1}{t}}
\]

\[
\geq \lim_{t \to \infty} \sup \log \left( E[\exp(\beta \frac{S_t}{t})] \right)
\]
At the stationary distribution of $\{\lambda_t\}_{t \in \mathbb{N}}$: $\Lambda(\beta) \geq \log E_{\omega}[\exp(\beta \log \lambda_\infty)] = \log \int_{\lambda}[\exp(\beta \log \lambda_\infty)]d\omega(\lambda)$.

As $\beta \to \infty$ for $\log \lambda_t < 0$, $\exp(\beta \log \lambda_\infty) \to 0$, but if $P_{\omega}(\log \lambda_\infty > 0) > 0$ at the stationary distribution of $\{\lambda_t\}_t$, then as $\beta \to \infty$,

$$\Lambda(\beta) = \log \int_{\lambda}[\exp(\beta \log \lambda_\infty)]d\omega(\lambda) \to \infty$$

Therefore if we can show that $P_{\omega}(\log \lambda_\infty > 0) > 0 \longrightarrow \exists$ a $\beta_1$ for which $\Lambda(\beta_1) > 0$.

Note: $\Lambda(\beta)$ convex, since moments of nonnegative random variables are log convex (in $\beta$) (Loeve (1977, p. 158)).

$\therefore \Lambda(\beta)$ convex $\longrightarrow \exists$ a unique $\kappa$ for which $\Lambda(\kappa) = 0$. 
To show that $P_\omega (\lambda_\infty > 1) > 0$, define $A = \left\{ d \in \left(0, \frac{\mu a \delta}{1-\rho \delta}\right) \right\}$, $\mu \in (0, 1)$ so that $\frac{\mu a \delta}{1-\rho \delta} < \frac{a}{1-\rho}$.

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

$$\lambda_\infty = 1 - \delta g d_{t-1} \left(\delta^{-1} (1 - \rho \delta) d_{t-1} - \varepsilon_t\right),$$

so for $d_{t-1} \in A$ and $\varepsilon_t \in (\mu a, a]$, it follows that $\lambda_t > 1$.

Thus $P_\omega (\lambda_\infty > 1) = P_\pi (d_{t-1} \in A) P (\varepsilon_t \in (\mu a, a]) > 0$.

Parts (iii) and (iv) establish additional technical conditions to satisfy remainder of Theorem 1.6 of Roitershtein (2007).
Proposition characterizes the tail of the stationary distribution of $\phi$ as a power tail with exponent $\kappa$.

Distribution of $\phi$ has moments only up to the highest integer less than $\kappa$.

Therefore is a ‘fat tailed’ distribution rather than a Normal distribution.

Results driven by: the stationary distribution of $\{\lambda_t\}_{t \in \mathbb{N}}$ has a mean less than 1 but has support above 1 with positive probability.

Thus large deviations are strings of realizations of $\lambda_t$ above one, even though they are rare events, and can produce fat tails.
In the asset pricing model $\phi$ relates the dividends to assets prices.

Under adaptive learning, the results above show how the probability of large deviations of $\phi$ from its REE value is characterized by a fat tailed distribution, and will occur with higher likelihood than under a Normal distribution.
What if $d_t$ is an $MA(1)$ process? Proposition still applies.

Let

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

Then under the PLM

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

after observing $\varepsilon_t$ at time $t$ but not $\phi_{1t+1}$, agents expect

$$E_t(p_{t+1}) = \phi_{0t} E_t(\varepsilon_{t+1}) + \phi_{1t} E_t(\varepsilon_t) = \phi_{1t} \varepsilon_t$$
The ALM is

\[ p_t = \delta \phi_{1t} \epsilon_t + \theta (\epsilon_t + \zeta \epsilon_{t-1}) = [\delta \phi_{1t} + \theta] \epsilon_t + \theta \zeta \epsilon_{t-1} \]

and the REE is

\[ \phi_0 = \theta (1 + \delta \zeta), \quad \phi_1 = \theta \zeta. \]

Under the learning algorithm in (8) we obtain

\[
\begin{align*}
\phi_{1t} & = \phi_{1t-1} + gd_{t-1} (p_t - \phi_{1t-1} d_{t-1}) \\
\phi_{1t+1} & = \lambda_{t+1} \phi_{1t} + \psi_{t+1} \\
\lambda_{t+1} & = 1 - gd_t^2 + g \delta \epsilon_{t+1} d_t \\
\psi_{t+1} & = g \theta \epsilon_{t+1} d_t + \theta \zeta gd_t \epsilon_t
\end{align*}
\]
To explore how $\kappa$ is related to the underlying parameters of our model, we can simulate the learning algorithm that updates $\phi$, and then estimate $\kappa$. We can then explore how our estimate of $\kappa$ from simulated series varies as we vary parameters.

We simulate 100 series for $\phi_t$ 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 $\kappa$ for each simulation and produce an average $\kappa$. 
Model Simulations and Comparative Statics

- Escapes or large deviations in prices take place when sequences of large shocks to dividends throw off the learning process. Such escapes are more likely if dividend shocks produce values of $\lambda_t$ above 1. We expect lower $\kappa$, or fatter tails, as the support of $\lambda_t$ that lies above 1 gets larger.
Comparative Statics

- For simulations, given the parameter estimates in the next section, we use the baseline parameterization, \((\rho, g, \beta, \gamma) = (0.98, 0.5, 0.95, 2.5)\). The restriction given by equation on \(a\) 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 \(\kappa\) is 4.9172, the average \((P_t / D_t)\) is 20.4989 and the average Std. Dev \((P_t / D_t)\) is 12.6142. This are quite close to the data characteristics in Table 1.
Comparative Statics

Figure 3. Simulation Results.
As the learning gain falls, that is, the horizon for learning increases, average $\kappa$ rises. However, for empirically plausible values of $g$ the average $\kappa$ is small.
Comparative Statics

- This of course is in accord with the Theorem 7.9 in Evans and Honkapohja (2001). As the gain parameter \( g \to 0 \) and \( tg \to \infty \), \( \{\phi_t^g - \kappa\} / g^{0.5} \) converges to a Gaussian variable where \( \kappa \) is the globally stable point of the associated ODE \( \dot{\phi} = T(\phi) - \phi \) describing the mean dynamics.

- More generally, as \( g \to 0 \), the estimated coefficient under learning with gain parameter \( g \), \( \phi_t^g \), converges in probability (but not uniformly) to \( \kappa \) for \( t \to \infty \). However, there will always exist arbitrarily large values of \( t \) with \( \phi_t^g \) taking values remote from \( \kappa \) (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 \( \kappa \) is obtained for fixed \( g > 0 \).
We use a maximum likelihood procedure following Clauset et al. (2009) to estimate \( \kappa \) associated with \( P_t/D_t \) for both S&P and CRSP dividend series.

\( \kappa \) is small for both series, suggesting that only the first few moments of \( P_t/D_t \) exist irrespective of the data source.

Table 1 below also reports the estimated persistence \( \rho \) 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>
<thead>
<tr>
<th></th>
<th>S &amp; P 500</th>
<th>CRSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\kappa}$</td>
<td>3.6914</td>
<td>5.5214</td>
</tr>
<tr>
<td>s.e.$(\hat{\kappa})$</td>
<td>0.3828</td>
<td>2.6046</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>0.7891</td>
<td>0.7519</td>
</tr>
<tr>
<td>s.e.$(\hat{\rho})$</td>
<td>0.0523</td>
<td>0.0777</td>
</tr>
<tr>
<td>Mean $(P_t / D_t)$</td>
<td>25.5211</td>
<td>26.1805</td>
</tr>
<tr>
<td>Std. Dev. $(P_t / D_t)$</td>
<td>13.1758</td>
<td>9.3298</td>
</tr>
</tbody>
</table>
We use two separate approaches to get estimates for the gain parameter \( g \).

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

\[
\min_{\Theta} [\kappa - \kappa(\Theta)]^2.
\]

Minimization was conducted using a simplex method and standard errors were computed using a standard inverse Hessian method.
Empirics

Parameter Estimates

- The minimization procedure proceeds as follows. For candidate parameterizations of $\vartheta$ we employ the S&P or CRSP series dividends $d_t$ to calculate $\phi_t$. The ALM then produces a corresponding $p_t$ series which in turn delivers a price-dividend ratio $P_t/D_t$.

- We then estimate the $\kappa$ 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 $\vartheta$.

- Table 2 below reports the estimates ($\hat{\vartheta}$) and associated standard errors ($s.e.(\hat{\vartheta})$) for each of the the S&P or CRSP dividend series, as well as the $\kappa$ associated with the estimated parameters.
### Table 2. Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$g$</td>
<td>0.3468</td>
<td>2.7158</td>
<td>0.5257</td>
<td>0.4722</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2.6503</td>
<td>1.7481</td>
<td>2.4598</td>
<td>0.6259</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9615</td>
<td>0.3870</td>
<td>0.8984</td>
<td>0.4576</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.8729</td>
<td>0.0552</td>
<td>0.7959</td>
<td>0.1355</td>
</tr>
<tr>
<td>Associated $\kappa$</td>
<td>2.4128</td>
<td></td>
<td></td>
<td>5.5214</td>
</tr>
</tbody>
</table>
Empirics

Parameter Estimates

- Point estimates of $g$ plausible, although standard errors are quite large. Carceles-Poveda and Giannitsaraou (2008) discuss plausible values of $g$, where under constant gain the decay in weights on past observations dating $i$ periods back is $(1 - g)^{i-1}$.

- For quarterly observations, $g = 0.46$ corresponds to 15 years of learning, with periods beyond 15 years having practically zero weight. For learning going back 20 years, $g = 0.37$.

- Carceles-Poveda and Giannitsaraou (2008) report that the standard deviations of price-dividend ratios for the Lucas asset pricing model under rational expectations or learning are smaller than the standard deviations in the data by factors of about 20 to 50.
Note that our estimates match the parameter values used by Carceles-Poveda and Giannitsaraou (2008) in their simulations except for $\gamma$, the CRRA parameter, which they set equal to 1 while we have it at $\gamma = 2.5$. Note also from comparative statics above that $\kappa$ drops dramatically with $\gamma$.

### Parameter Estimates

<table>
<thead>
<tr>
<th></th>
<th>S &amp; P 500</th>
<th></th>
<th>CRSP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$g$</td>
<td>0.3468</td>
<td>2.7158</td>
<td>0.5257</td>
<td>0.4722</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2.6503</td>
<td>1.7481</td>
<td>2.4598</td>
<td>0.6259</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9615</td>
<td>0.3870</td>
<td>0.8984</td>
<td>0.4576</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.8729</td>
<td>0.0552</td>
<td>0.7959</td>
<td>0.1355</td>
</tr>
<tr>
<td>Associated $\kappa$</td>
<td>2.4128</td>
<td></td>
<td>5.5214</td>
<td></td>
</tr>
</tbody>
</table>
Empirics
Parameter Estimates-Second Approach

- The PLM is

\[ p_t = \phi_{t-1} d_{t-1} + \xi_t, \quad \xi_t \sim iid(0, \sigma^2_{\xi}), \quad \sigma^2_{\xi} < +\infty \]

- \( \phi \) follows a random walk:

\[ \phi_t = \phi_{t-1} + \Lambda_t, \quad \Lambda_t \sim iid(0, \sigma^2_{\Lambda}), \quad \sigma^2_{\Lambda} < +\infty \]

- The Bayesian agent would use these to estimate \( \sigma_{\Lambda}, \sigma_d \) and \( \sigma_{\xi} \) to set an optimal estimate of the gain in the limit as

\[ g = \frac{\sigma_{\Lambda} \sigma_d}{\sigma_{\xi}} \]
Empirics

To compute $g$ an estimate of $\sigma_d$ is of course readily obtained from the actual dividend data. However we need to specify a method for the agents to compute estimates of $\sigma_\Lambda$ and $\sigma_\xi$. If we recognize the system above as being analogous to a time varying parameter formulation, then using actual price data, and employing the methods laid out in Kim and Nelson (1999) we can obtain estimates of $\sigma_\Lambda$ and $\sigma_\xi$. 
Table 3. Model Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>S &amp; P 500</th>
<th>CRSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_\Lambda$</td>
<td>0.8122</td>
<td>0.7718</td>
</tr>
<tr>
<td>$\sigma_\zeta$</td>
<td>0.3157</td>
<td>0.0230</td>
</tr>
<tr>
<td>log $L$</td>
<td>-61.4102</td>
<td>-17.5256</td>
</tr>
<tr>
<td>$\sigma_d$</td>
<td>0.1892</td>
<td></td>
</tr>
<tr>
<td>Associated $g$</td>
<td>0.4866</td>
<td></td>
</tr>
</tbody>
</table>

$g$ is much larger than what is usually assumed in the literature. In Figure 4, a value of $g = 0.4866$ yields a tail $\kappa$ of about 4.9 while $g = 0.5455$ yields a $\kappa$ of about 4.75, compared to $\kappa$ in the data ranging from 3.7 to 5.5 in Table 1. We also simulated the model with baseline parameters and gains of 0.4866 and 0.5455. These resulted in average price-dividend ratios of 20.6324 and 20.6965 with standard deviation values of 10.0051 and 10.5870.
Finally, instead of using actual $P$ and $D$ data series, we generate data by simulating our model with our benchmark values $(\rho, g, \beta, \gamma) = (0.80, 0.4, 0.95, 2.5)$, and then compute $g$ using the methods in Kim and Nelson (1999). The average $g$ is 0.3826, which is quite close to and confirms the benchmark value of $g = 0.4$ that is used in generating the simulated data.

Question: Are there clever econometric methods for the agent to test whether $\{\phi\}_t$ is not a random walk? They can test the mean of $\{\phi_t - \phi_{t-1}\}$ to see if they can reject if is zero (They cannot reject)? They can check if its variance matches the forecast error in the PLM? All these also depend on the process for dividends...How smart are they?
Conclusion

- Asks whether agents ‘learn’ REE.
- Constant gain learning: heavier emphasis on recent observations, can lead to escape dynamics.
- Escape dynamics can propel estimated coefficients away from the REE values.
- We show (under asset pricing interpretation), ‘bubbles,’ or asset prices that exhibit large deviations from their REE ratios to dividends can occur with a frequency associated with a fat tailed power law.
- Techniques can be generalized to higher dimensions and to finite state Markov chains, can be applied to other more general economic models.