1. The Model

Consider an (endowment oriented) N-asset pricing environment following Lucas (1978) given by

$$\max_{\{C_t\}_{t=0}^{\infty}} \quad \Gamma = E_0 \sum_{t=0}^{\infty} \delta^t u(C_t) \tag{1}$$

$$s.t. \quad C_t + \sum_{i=1}^{N} P_{it} S_{it+1} = \sum_{i=1}^{N} S_{it} \left(P_{it} + D_{it}\right) \tag{2}$$

which yields the following well known pricing equation under a no-bubbles condition

$$P_{it} = E_t \left\{ \delta \frac{u'(C_{t+1})}{u'(C_t)} \left(P_{it+1} + D_{it+1}\right) \right\}. \tag{3}$$

Assuming the existence of single asset and noting that in this single asset case the competitive equilibrium condition

$$C_t = \sum_{i=1}^{N} D_{it} \tag{4}$$

reduces to $C_t = D_t$, we have

$$P_t = E_t \left\{ \delta \frac{u'(D_{t+1})}{u'(D_t)} \left(P_{t+1} + D_{t+1}\right) \right\} \tag{5}$$

$$= E_t \left\{ \delta M_{t+1} \left(P_{t+1} + D_{t+1}\right) \right\}. \tag{6}$$

For the CRRA utility functional form

$$u(C_t) = \frac{C_t^{1-\theta}}{1-\theta}, \;\; \theta > 0 \tag{7}$$

the pricing equation is

$$P_t = E_t \left\{ \delta \left(\frac{D_{t+1}}{D_t} \right)^{-\theta} \left(P_{t+1} + D_{t+1}\right) \right\}. \tag{8}$$

For steady state computation let D_t have steady state value $\overline{D}$ then the above equation reduces to

$$\overline{P} = \delta \left(\frac{\overline{D}}{\overline{D}} \right)^{-\theta} \left(\overline{P} + \overline{D}\right) \tag{9}$$

$$= \delta \left(\overline{P} + \overline{D}\right) \tag{10}$$

$$\rightarrow \;\; \overline{P} = \delta \overline{P} + \delta \overline{D} \tag{11}$$

$$\rightarrow \;\; \overline{P} = \frac{\delta}{1-\delta} \overline{D} \tag{12}$$

Next, rewrite the pricing equation as

$$0 = E_t \left\{ \delta \left(\frac{e^{\tilde{D}_{t+1}}}{e^{\tilde{D}_t}} \right)^{-\theta} \left(e^{\tilde{P}_{t+1}} + e^{\tilde{D}_{t+1}} \right) \right\} - e^{\tilde{P}_t} \tag{13}$$

$$= f(\tilde{D}_t, \tilde{D}_{t+1}, \tilde{P}_t, \tilde{P}_{t+1}) \tag{14}$$

where variables are written as $X_t = \exp\left(\log X_t\right) = e^{\tilde{X}_t}$. Now, take a first order Taylor series approximation of the function f around $(\overline{P},\overline{D})$ as follows. First the relevant derivatives are

$$\frac{\partial f}{\partial \tilde{D}_t} \quad |_{(\overline{P},\overline{D})} = -\theta\delta\left(\frac{e^{\tilde{D}_{t+1}}}{e^{\tilde{D}_t}}\right)^{-\theta-1}\left(e^{\tilde{P}_{t+1}}+e^{\tilde{D}_{t+1}}\right)\left(-e^{-\tilde{D}_t}\right) = \frac{\theta\delta(\overline{P}+\overline{D})}{\overline{D}} \tag{15}$$

$$\frac{\partial f}{\partial \tilde{D}_{t+1}} \quad |_{(\overline{P},\overline{D})} = -\theta\delta\left(\frac{e^{\tilde{D}_{t+1}}}{e^{\tilde{D}_t}}\right)^{-\theta-1} e^{\tilde{D}_{t+1}}\left(e^{\tilde{P}_{t+1}}+e^{\tilde{D}_{t+1}}\right)+\delta\left(\frac{e^{\tilde{D}_{t+1}}}{e^{\tilde{D}_t}}\right)^{-\theta}\left(e^{\tilde{D}_{t+1}}\right) \tag{16}$$

$$= -\theta\delta\overline{D}(\overline{P}+\overline{D})+\delta\overline{D} \tag{17}$$

$$\frac{\partial f}{\partial \tilde{P}_t} \quad |_{(\overline{P},\overline{D})} = -e^{\tilde{P}_t} = -\overline{P} \tag{18}$$

$$\frac{\partial f}{\partial \tilde{P}_{t+1}} \quad |_{(\overline{P},\overline{D})} = \delta\left(\frac{e^{\tilde{D}_{t+1}}}{e^{\tilde{D}_t}}\right)^{-\theta} e^{\tilde{P}_{t+1}} = \delta\overline{P} \tag{19}$$

thus the approximation is

$$0 \approx \frac{\partial f}{\partial \tilde{D}_t}|_{(\overline{P},\overline{D})}\, d_t + \frac{\partial f}{\partial \tilde{D}_{t+1}}|_{(\overline{P},\overline{D})}\, E_t(d_{t+1}) + \frac{\partial f}{\partial \tilde{P}_t}|_{(\overline{P},\overline{D})}\, p_t + \frac{\partial f}{\partial \tilde{P}_{t+1}}|_{(\overline{P},\overline{D})}\, E_t(p_{t+1}) \tag{20}$$

$$0 \approx \frac{\theta\delta(\overline{P}+\overline{D})}{\overline{D}}d_t + \left[-\theta\delta\overline{D}(\overline{P}+\overline{D})+\delta\overline{D}\right]E_t(d_{t+1}) + \left[-\overline{P}\right]p_t + \left[\delta\overline{P}\right]E_t(p_{t+1}) \tag{21}$$

$$\left[\overline{P}\right]p_t = \left[\delta\overline{P}\right]E_t(p_{t+1}) + \left[-\theta\delta\overline{D}(\overline{P}+\overline{D})+\delta\overline{D}\right]E_t(d_{t+1}) + \frac{\theta\delta(\overline{P}+\overline{D})}{\overline{D}}d_t \tag{22}$$

$$p_t = \delta E_t(p_{t+1}) + \left[\frac{\delta\overline{D}-\theta\delta\overline{D}(\overline{P}+\overline{D})}{\overline{P}}\right]E_t(d_{t+1}) + \frac{\theta\delta(\overline{P}+\overline{D})}{\overline{P}\overline{D}}d_t \tag{23}$$

$$= \delta E_t(p_{t+1}) + \left[\frac{\delta\overline{D}-\theta\delta\overline{D}\left(\frac{\delta}{1-\delta}\overline{D}+\overline{D}\right)}{\frac{\delta}{1-\delta}\overline{D}}\right]E_t(d_{t+1}) + \frac{\theta\delta\left(\frac{\delta}{1-\delta}\overline{D}+\overline{D}\right)}{\frac{\delta}{1-\delta}\overline{D}\overline{D}}d_t \tag{24}$$

$$= \delta E_t(p_{t+1}) + \left[\frac{\delta-\theta\delta\left(\frac{\delta}{1-\delta}\overline{D}+\overline{D}\right)}{\frac{\delta}{1-\delta}}\right]E_t(d_{t+1}) + \frac{\theta\delta\overline{D}\left(\frac{\delta}{1-\delta}+1\right)}{\frac{\delta}{1-\delta}\overline{D}\overline{D}}d_t \tag{25}$$

$$= \delta E_t(p_{t+1}) + \left[\frac{\delta-\theta\delta\overline{D}\left(\frac{\delta+1-\delta}{1-\delta}\right)}{\frac{\delta}{1-\delta}}\right]E_t(d_{t+1}) + \frac{\theta\delta\left(\frac{\delta+1-\delta}{1-\delta}\right)}{\frac{\delta}{1-\delta}\overline{D}}d_t \tag{26}$$

$$= \delta E_t(p_{t+1}) + \left[\frac{\delta-\theta\overline{D}\left(\frac{\delta}{1-\delta}\right)}{\frac{\delta}{1-\delta}}\right]E_t(d_{t+1}) + \frac{\theta\left(\frac{\delta}{1-\delta}\right)}{\frac{\delta}{1-\delta}\overline{D}}d_t \tag{27}$$

$$p_t = \delta E_t(p_{t+1}) + \left[1-\delta-\theta\overline{D}\right]E_t(d_{t+1}) + \frac{\theta}{\overline{D}}d_t \tag{28}$$

Now let $\overline{D}=1$ to obtain the linearized pricing equation around its' non-stochastic steady state

$$p_t = \delta E_t(p_{t+1}) + (1-\delta-\theta)E_t(d_{t+1}) + \theta d_t, \quad \delta\in(0,1), \quad \theta\in\mathbb{R} \tag{29}$$

where all lowercase variables denote log-deviations from steady state $(\overline{P},\overline{D}) = \left(\frac{\delta}{1-\delta},1\right)$.

Assume that the exogenous dividends process follows

$$d_t = \rho d_{t-1} + \varepsilon_t, \quad |\rho|<1, \quad \varepsilon_t \sim i.i.d.(0,\sigma_\varepsilon^2), \quad \sigma_\varepsilon^2 < +\infty \tag{30}$$

where $d_t = \log(D_t)$ then from this process for dividends we know

$$E_t(d_{t+1}) = \rho d_t \tag{31}$$

and so

$$p_t = \delta E_t(p_{t+1}) + \gamma d_t, \quad \gamma \equiv (1-\delta-\theta)\rho + \theta \tag{32}$$

is the fundamental expectational difference equation corresponding to the assumed form for the exogenous driving process.

For the REE computation we first conjecture following Uhlig's (1999) undetermined coefficients approach that

$$\begin{aligned} p_t &= \phi d_t \qquad (33) \\ &\rightarrow E_t(p_{t+1}) = \phi\rho d_t \qquad (34) \end{aligned}$$

where ϕ is the unknown REE coefficient. Insert the above into (32) to get

$$p_t = (\delta\phi\rho + \gamma)d_t \tag{35}$$

and compare coefficients across (35) and (33) to get

$$\begin{aligned} \phi &= \delta\phi\rho + \gamma \qquad (36) \\ &\rightarrow \phi^{REE} = \frac{\gamma}{1-\delta\rho} \qquad (37) \end{aligned}$$

which will be unique and finite for all $\delta\rho \neq 1$ so we assume that condition.

In summary, the minimal state variable rational expectations solution to

$$\begin{aligned} p_t &= \delta E_t(p_{t+1}) + \gamma d_t, \quad \gamma \equiv (1-\delta-\theta)\rho + \theta \qquad (38) \\ d_t &= \rho d_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2), \quad \sigma_\varepsilon^2 < +\infty \qquad (39) \end{aligned}$$

is

$$p_t = \phi^{REE} d_t, \ \phi^{REE} = \frac{\gamma}{1-\delta\rho}. \tag{40}$$

With respect to data, Shiller (2005) provides a "Real Price" (P_t) and "Real Dividend" (D_t) series for the S & P 500 at a monthly frequency that covers 1871.1 through 2010.12. We therefore need to construct linearly detrended series that are the model equivalents to p_t and d_t. The steady state value $\overline{P}/\overline{D} = \frac{\delta}{1-\delta}$ and the assumption of trend stationarity for dividends implies a set of restricted trajectories for both prices and dividends that depend on the deep parameters $[\delta, \theta]$ which we set to $[0.95, 2.5]$ respectively throughout this paper. Empirically, we implement this detrending via a restricted least squares regression (see DeJong and Dave (2011)). Figure 1 below plots the original data in levels (that is, P_t, D_t and P_t/D_t) whereas Figure 2 plots the linearly detrended data (that is, p_t, d_t and p_t/d_t). Note that d_t is specified as an $AR(1)$ in the model and our estimate for ρ is 0.99885 implying that ϕ^{REE} equals 1.0337; the mean (standard deviation) of p_t/d_t in the data is 0.94881 (25.495) whereas the mean (standard deviation) of P_t/D_t is 26.59 (13.753).

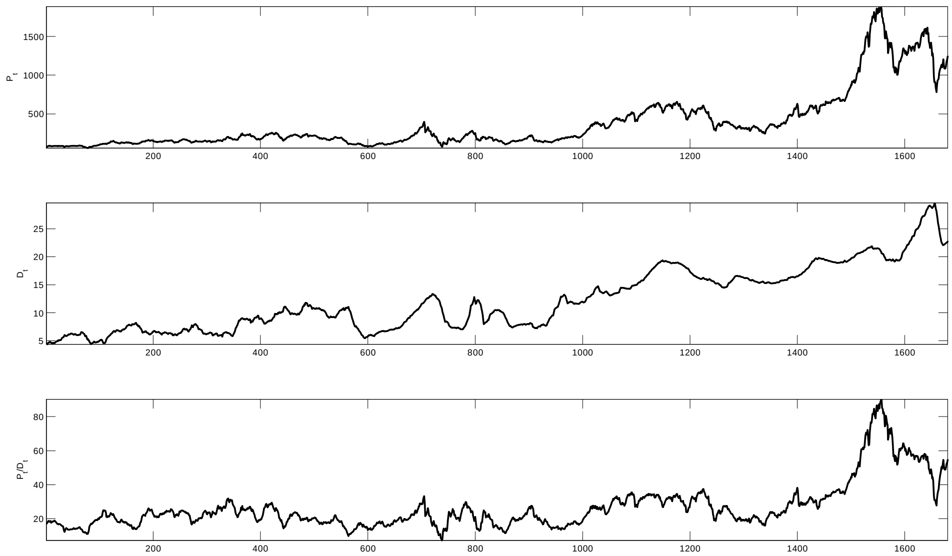

Figure 1. Monthly S & P 500 Prices, Dividends and Price to Dividend Ratio (1871.1-2010.12).

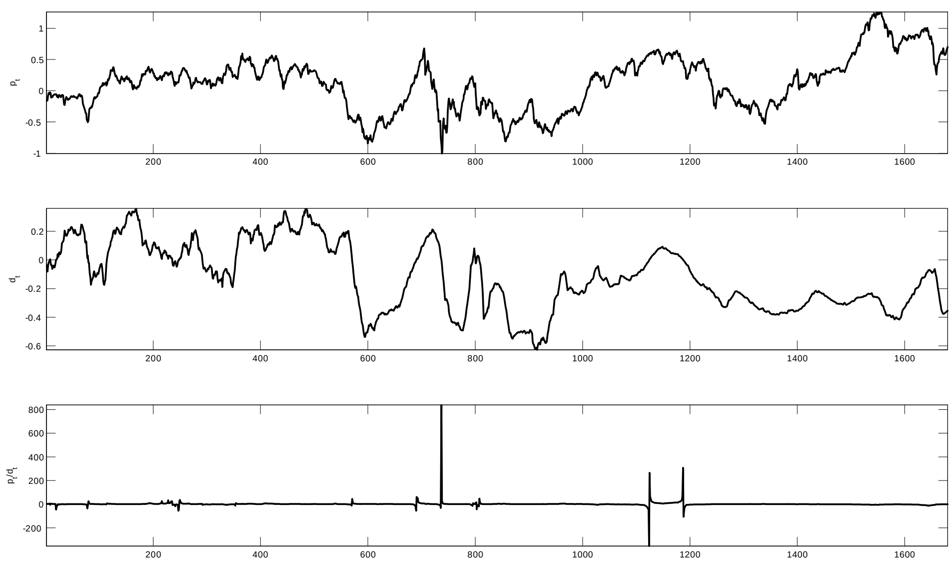

Figure 2. Linearly Detrended Data ($[\delta, \theta, \rho] = [0.95, 2.5, 0.99885]$).

Next, given a time series, say x_t, Clauset, Shalizi and Newman (2007) provide a full information method to estimate the κ and $\underline{x}$ in

$$p(x_t) \sim x^{-\kappa} \; \forall \; x \geq \underline{x}. \tag{41}$$

We estimate the power law coefficients associated with P_t/D_t and p_t/d_t as 3.5779 and 2.0007 respectively. These price to dividends ratios and associated power law coefficients form the basis of our theoretical and empirical modeling in the following sections.

2. Constant Gain Stochastic Gradient Learning

For learning we first conjecture

$$p_t = \phi_{t-1} d_{t-1} + \xi_t, \quad \xi_t \sim i.i.d.(0, \sigma_\xi^2), \quad \sigma_\xi^2 < +\infty \tag{42}$$

which in turn implies

$$E_t(p_{t+1}) = \phi_{t-1} d_t \tag{43}$$

which when inserted into (32) yields

$$\begin{aligned} p_t &= \delta \phi_{t-1} d_t + \gamma d_t && (44) \\ &= (\delta \phi_{t-1} + \gamma) \rho d_{t-1} + (\delta \phi_{t-1} + \gamma) \varepsilon_t && (45) \end{aligned}$$

and hence the actual law of motion under adaptive learning is

$$p_t = (\delta \phi_{t-1} + \gamma) \rho d_{t-1} + (\delta \phi_{t-1} + \gamma) \varepsilon_t, \tag{46}$$

which can be contrasted against the actual law of motion under rational expectations which is

$$p_t = \phi d_t = \phi \rho d_{t-1} + \phi \varepsilon_t. \tag{47}$$

Now return to the learning case and assume that ϕ_t evolves as per

$$\phi_t = \phi_{t-1} + g d_{t-1} (p_t - \phi_{t-1} d_{t-1}), \quad g \in (0, 1) \tag{48}$$

and insert the actual law of motion under learning in place of p_t; this would yield

$$\begin{aligned} \phi_t &= \lambda_t \phi_{t-1} + \psi_t && (49) \\ \lambda_t &= 1 - (1 - \rho\delta) g d_{t-1}^2 + \delta g d_{t+1} \varepsilon_t = 1 - g d_{t-1}^2 + g \delta d_t d_{t-1} && (50) \\ \psi_t &= \gamma \rho g d_{t-1}^2 + \gamma g d_{t-1} \varepsilon_t = \gamma g d_t d_{t-1}. && (51) \end{aligned}$$

The equation in (49) takes the form of a linear recursion with both multiplicative (λ_t in (50)) and additive (ψ_t in (51)) noise. For such a recursion, in order to characterize the number of moments of the stationary distribution ϕ_t we apply the results in Roitershtein (2007). This application is not direct, what is required is to prove a theorem using Roitershtein (2007) that characterizes the tail of the stationary distribution of ϕ_t (the price to dividends ratio as it evolves under constant gain stochastic gradient learning).