Economics Letters 124 (2014) 329–334

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Recursive preferences, learning and large deviations

Chetan Dave [a,*], Kwok Ping Tsang [b]

[a] New York University (Abu Dhabi), PO Box 903, New York, NY 10276, USA

[b] Department of Economics, Virginia Tech, Blacksburg, VA, 24061, USA

HIGHLIGHTS

- An asset pricing model with recursive preferences is specified.
- The model is estimated under the assumption of adaptive learning.
- Both of these sources of volatility account for fluctuations in liquid stock markets.
- However, only risk aversion matters for illiquid housing markets.

ARTICLE INFO

Article history:
Received 10 May 2014
Received in revised form
10 June 2014
Accepted 15 June 2014
Available online 24 June 2014

JEL classification:
D80
D83
D84

Keywords:
Recursive preferences
Adaptive learning
Large deviations
Fat tails
Asset prices

ABSTRACT

We estimate the relative contribution of recursive preferences versus adaptive learning in accounting for the tail thickness of price–dividends/rents ratios. We find that both of these sources of volatility account for volatility in liquid (stocks) but not illiquid (housing) assets.

1. Introduction

In analyzing sources of volatility in asset pricing models two mechanisms stand out: learning and the adoption of recursive preferences. The former attributes volatility in asset prices to processes via which an agent comes to know of underlying fundamentals. The latter requires the agent to care about *when* uncertainty is resolved, determined entirely by preferences for risk and intertemporal substitution. Benhabib and Dave (2014) suggest that, in a single asset version of Lucas (1978), adaptive learning via a constant gain stochastic gradient (CGSG) algorithm causes the stationary distribution of the price–dividends ratio (PDR) to exhibit fat tails despite dividends being modeled as a thin-tailed process. These fat tails are shown to vary as a function of the deep parameters of the model, and the large deviations of the PDR from its rational expectations equilibrium value are able to account for the volatility in stock indices. Here we investigate the empirical contribution of a recursive preference formulation, following Epstein and Zin (1989, 1991), to the learning and large deviations model for both the stock and housing markets.

Why this particular formulation? The estimates provided in Benhabib and Dave (2014) of the CRRA coefficient are high. High estimates of the CRRA coefficient are common in empirical consumption asset pricing and warrant further investigation. Further, Benhabib and Dave (2014) show that the higher the CRRA coefficient, the thicker the tails of the stationary distribution of the PDR; it would be useful to have low CRRA estimates and still account for the thick tails due to a learning algorithm. Given the elegant manner in which recursive preferences separate out the effects of risk aversion and intertemporal substitution, perhaps

* Corresponding author.
E-mail addresses: cdave@nyu.edu (C. Dave), byront@vt.edu (K.P. Tsang).

http://dx.doi.org/10.1016/j.econlet.2014.06.014

their adoption is sufficient to "break out" a high CRRA coefficient estimate into its constituent components whilst still retaining a role for learning. Further, simulations can indicate whether the tails of the stationary distribution of the PDR thicken as the parameterized preferences for risk *and* the desire to smooth consumption intertemporally increase. Finally, by comparing estimates from S&P 500 data versus an illiquid housing market, one can further focus on the three forces that could affect the thickness of the tail of a PDR or price–rent ratio (PRR) series: risk aversion, intertemporal substitution and learning.

Our investigation does deliver the sought-after results. Allowing for recursive preferences with S&P 500 data suggests an estimate for the CRRA coefficient of around 2.5 and an estimate for the inverse elasticity of intertemporal substitution of around 1.2. Further, as these parameters increase, the tail index of the PDR that governs the thickness of its tails does fall. The higher the CRRA coefficient or the inverse elasticity of intertemporal substitution, the smaller is the number of moments associated with the tail of the stationary distribution of the PDR. Finally, the estimate of the learning gain does not change much relative to Benhabib and Dave (2014), indicating that learning continues to play a strong role in determining volatility in the S&P 500. With housing data, we find that the estimate of the gain parameter, which in part governs the effect of learning on the tail of the stationary distribution of the PRR, comes with a very large standard error. However, the relative risk aversion parameter is precisely estimated. This suggests that for the housing market a main driver for the thickness of tail of the stationary PRR is risk aversion, an expected result given the illiquid nature of housing as an asset.

In the next section we specify the model under recursive preferences, and then provide minimum distance estimates of its parameters in Section 3. We discuss simulated "comparative statics" results in Section 4 and conclude in Section 5 with a description of how the formulation of models as linear recursions with multiplicative noise can assist in characterizing data (and models) that exhibit fat tails and thus the possibility of rare disasters.

2. The model

Under recursive preferences the representative agent optimizes

$$\left[(1-\beta)C_t^{1-\gamma}+\beta U_{t+1}^{*1-\gamma}\right]^{\frac{1}{1-\gamma}}, \quad \beta\in(0,1) \tag{1}$$

subject to the usual constraints, where β is the discount factor. The inverse of the parameter $\gamma > 0$ is the intertemporal elasticity of substitution. Certainty equivalent future utility is given by

$$U_{t+1}^{*}=\left[E_t\left(U_{t+1}^{1-\alpha}\right)\right]^{\frac{1}{1-\alpha}} \tag{2}$$

where the parameter $\alpha > 0$ is the relative risk aversion coefficient.[1] The stochastic discount factor is

$$m_{t,t+1}=\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\left(\frac{V_{t+1}}{U_{t+1}^{*}}\right)^{\gamma-\alpha}. \tag{3}$$

Defining wealth as the present value of consumption

$$W_t=C_t+E_t\left(m_{t,t+1}W_{t+1}\right) \tag{4}$$

the return to wealth is

$$R_{t,t+1}=\frac{W_{t+1}}{W_t-C_t}=\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\left(\frac{V_{t+1}}{U_{t+1}^{*}}\right)^{\gamma-1}\right]^{-1}, \tag{5}$$

yielding the stochastic discount factor

$$m_{t,t+1}=\beta\left(\frac{C_{t+1}}{C_t}\right)^{\frac{-\gamma(1-\alpha)}{1-\gamma}}R_{t,t+1}^{\frac{\gamma-\alpha}{1-\gamma}}. \tag{6}$$

If stocks are the only wealth, and consumption is the same as dividends (D_t) in equilibrium, we have the single asset pricing equation

$$1=\beta E_t\left[\left(\frac{D_{t+1}}{D_t}\right)^{\frac{-\gamma(1-\alpha)}{1-\gamma}}\left(\frac{P_{t+1}+D_{t+1}}{P_t}\right)^{\frac{1-\alpha}{1-\gamma}}\right]. \tag{7}$$

Let $\kappa=\frac{1-\alpha}{1-\gamma}$, then in the steady state

$$\overline{P}=\frac{\beta^{\frac{1}{\kappa}}}{1-\beta^{\frac{1}{\kappa}}}\overline{D}. \tag{8}$$

The linear approximation is

$$\begin{aligned}0\approx{}&\frac{\gamma\kappa\beta(\overline{P}+\overline{D})^{\kappa}}{\overline{D}}d_t\\&+\left[-\gamma\kappa\beta(\overline{P}+\overline{D})^{\kappa}\overline{D}+\beta\kappa\overline{D}\left(\frac{\overline{P}+\overline{D}}{\overline{P}}\right)^{\kappa-1}\right]E_td_{t+1}\\&-\kappa\overline{P}^{\kappa}p_t+\kappa\beta(\overline{P}+\overline{D})^{\kappa-1}\overline{P}E_tp_{t+1}.\end{aligned} \tag{9}$$

We let $\overline{D}=1$ to obtain the linearized equation

$$\begin{aligned}&p_t=\gamma d_t+\eta E_t(d_{t+1})+\beta^{\frac{1}{\kappa}}E_t(p_{t+1}),\\&\eta=\left[\left(1-\beta^{\frac{1}{\kappa}}\right)^{\kappa}\left(\beta^{\frac{1}{\kappa}}\right)^{1-\kappa}\right]\end{aligned} \tag{10}$$

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

Assume that the exogenous dividends process follows

$$d_t=\rho d_{t-1}+\varepsilon_t,\quad |\rho|<1,\ \varepsilon_t\sim i.i.d.(0,\sigma_\varepsilon^2),\ \sigma_\varepsilon^2<+\infty \tag{11}$$

with compact support $[-a,a]$, $a>0$. Then from this process for dividends we know that

$$E_t(d_{t+1})=\rho d_t \tag{12}$$

and so

$$\begin{aligned}&p_t=\beta^{\frac{1}{\kappa}}E_t(p_{t+1})+\theta d_t,\quad \theta\equiv\gamma+\eta\rho,\\&\eta=\left[\left(1-\beta^{\frac{1}{\kappa}}\right)^{\kappa}\left(\beta^{\frac{1}{\kappa}}\right)^{1-\kappa}\right],\quad \kappa=\frac{1-\alpha}{1-\gamma}\end{aligned} \tag{13}$$

is the fundamental expectational difference equation under investigation.

The REE value of ϕ is given by

$$\phi^{\text{REE}}=\frac{\theta\rho}{1-\beta^{\frac{1}{\kappa}}\rho} \tag{14}$$

which is unique and finite for all $\beta^{\frac{1}{\kappa}}\rho\neq1$ so we assume that condition holds.

For learning we conjecture that

$$p_t=\phi_{t-1}d_{t-1}+\xi_t,\quad \xi_t\sim i.i.d.(0,\sigma_\xi^2),\ \sigma_\xi^2<+\infty \tag{15}$$

implying

$$E_t(p_{t+1})=\phi_{t-1}d_t \tag{16}$$

[1] When $\gamma=\alpha$ we have the standard time-separable CRRA preference specification.

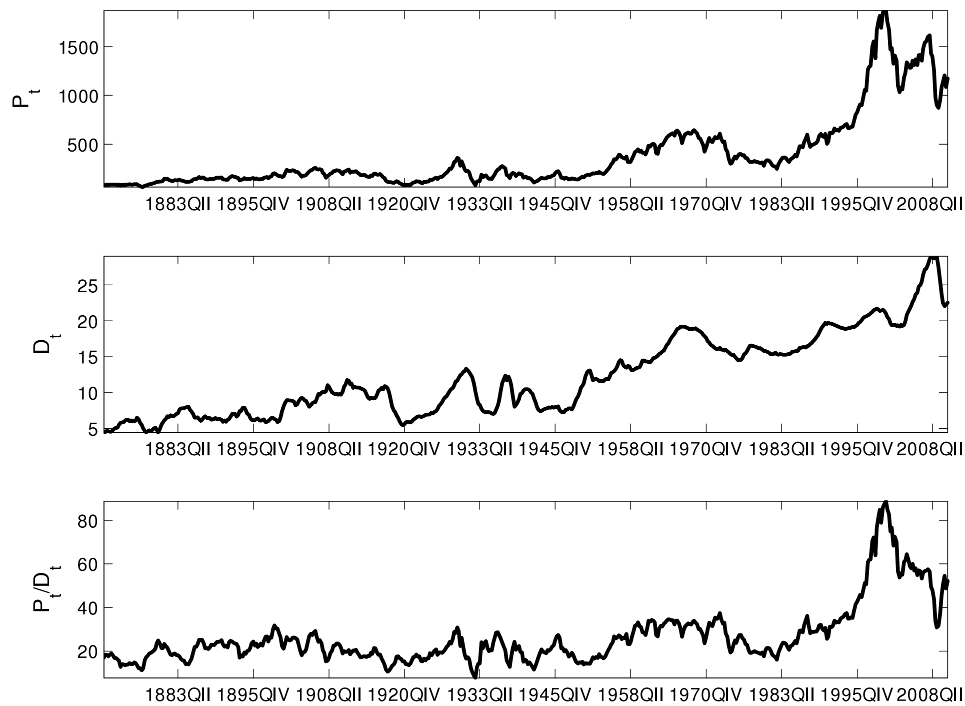

Fig. 1. S&P 500 data (1871QI–2010QIV).

which when inserted into (13) yields

$$p_t = \beta^{\frac{1}{\kappa}}\phi_{t-1}d_t + \theta d_t = \left(\beta^{\frac{1}{\kappa}}\phi_{t-1} + \theta\right) d_t \tag{17}$$

and hence the actual law of motion (ALM) is

$$p_t = \left(\beta^{\frac{1}{\kappa}}\phi_{t-1} + \theta\right)\rho d_{t-1} + \left(\beta^{\frac{1}{\kappa}}\phi_{t-1} + \theta\right)\varepsilon_t. \tag{18}$$

Now assume that ϕ_t evolves as per a CGSG algorithm

$$\phi_t = \phi_{t-1} + gd_{t-1}(p_t - \phi_{t-1}d_{t-1}), \quad g \in (0, 1) \tag{19}$$

and insert the ALM in place of p_t:

$$\phi_t = \lambda_t\phi_{t-1} + \psi_t, \tag{20}$$

$$\begin{aligned}\lambda_t &= 1 - g\left(1 - \rho\beta^{\frac{1}{\kappa}}\right)d_{t-1}^2 + g\beta^{\frac{1}{\kappa}}\varepsilon_t d_{t-1} \\ &= 1 - gd_{t-1}^2 + g\beta^{\frac{1}{\kappa}}d_t d_{t-1},\end{aligned} \tag{21}$$

$$\psi_t = g\theta d_t d_{t-1}. \tag{22}$$

The model in (20)–(22) follows the linear recursion with multiplicative and additive noise specification of Benhabib and Dave (2014).[2] The random variable λ_t that generates the multiplicative noise and is the source of large deviations and fat tails for the stationary distribution of ϕ_t needs to satisfy two characteristics. First, the distribution must have mean less than one or a stationary distribution for ϕ_t fails to exist. Second, for ϕ_t to have a fat tail even if the exogenous driving process is thin tailed, the distribution of λ_t needs to have some support above the unit circle: $P(|\lambda| > 1) > 0$. The first characteristic implies that the support of the *i.i.d.* noise $\varepsilon_t \in [-a, a]$ needs to be restricted to ensure that $E|\lambda_\infty| < 1$. Assuming that ε_t is uniformly distributed delivers the restriction

$$a < \sqrt{\frac{6\left(1 - \rho^2\right)}{g\left(1 - \beta^{\frac{1}{\kappa}}\rho\right)}}. \tag{23}$$

[2] Such formulations have yielded important insights in other domains as well. See Benhabib (2013) for an application in a monetary policy context and Benhabib et al. (2011) in a wealth distribution context.

The second characteristic can be seen from the fact that rare realizations of ε_t may well cause $g\beta^{\frac{1}{\kappa}}\varepsilon_t d_{t-1}$ in the λ_t term to "overwhelm" the term $1 - g\left(1 - \rho\beta^{\frac{1}{\kappa}}\right)d_{t-1}^2$ and generate support for λ_t above one. From (21), the support of λ_t above 1 unambiguously increases if β increases for all $\kappa \neq 0$. Increasing ρ can have an ambiguous effect: while the term $(1 - \rho\beta^{\frac{1}{\kappa}})$ declines and tends to raise λ_t for realizations of d_{t-1} and ε_t, the support of the stationary distribution of d_t gets bigger with higher ρ. While this can increase $g\left(1 - \rho\beta^{\frac{1}{\kappa}}\right)d_{t-1}^2$ and reduce the support of λ that is above 1 for large realizations of d_{t-1}^2, in our simulations the former effect seems to dominate. Finally decreasing g shrinks the support of λ_t that is above 1 so that as the gain parameter decreases, the tails of the stationary distribution of ϕ_t get thinner.

Next, the tail of the stationary distribution of ϕ_t depends on

$$\Lambda(\delta) = \lim_{n\to\infty}\sup\frac{1}{n}\log E\prod_{t=1}^{n}|\lambda_t|^{\delta} \quad \forall\delta \in \mathbb{R}. \tag{24}$$

In particular there is a unique positive $\xi < +\infty$ that solves $\Lambda(\delta) = 0$ such that

$$\begin{aligned}K_1(d_0) &= \lim_{\tau\to\infty}\tau^{\xi}P(\phi > \tau|d_0) \quad \text{and} \\ K_{-1}(d_0) &= \lim_{\tau\to\infty}\tau^{\xi}P(\phi < -\tau|d_0)\end{aligned} \tag{25}$$

and $K_1(d_0)$ and $K_{-1}(d_0)$ are not both zero. The parameter ξ, termed the tail index, governs the thickness of the tails of the stationary distribution of ϕ_t and can be estimated for any time series of real or simulated data using the methods of Clauset et al. (2009).

3. An empirical application

We focus first on the S&P 500 dataset of Shiller (1999, 2005) since its coverage includes major events in the US asset pricing history from post civil war volatility, the Great Depression to the Great Recession, plotted in Fig. 1. Prices track dividends; however the PDR shows a fair degree of volatility that motivates alternatives such as learning and recursive preferences.

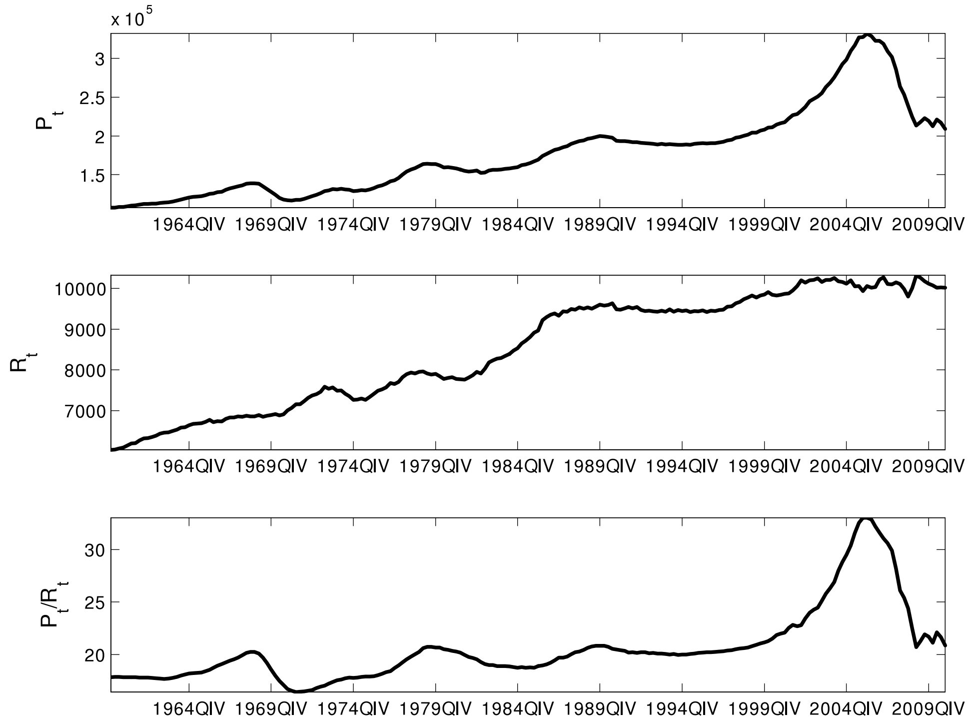

Fig. 2. Housing data (1960QI–2010QIV).

Table 1
Data characteristics, S&P 500, 1871QI–2010QIV.

Statistic	Value	Statistic	Value
$\widehat{\xi}$	3.5979	Std. Dev. (P_t/D_t)	13.7403
$s.e.(\widehat{\xi})$	0.1889	Corr. (P_t/D_t)	0.9845
$\widehat{\rho}$	0.9823	$r = \frac{\overline{D_t}}{\overline{P_t}}$	0.0322
$s.e.(\widehat{\rho})$	0.0078	$\beta = (1+r)^{-1}$	0.9688
Mean (P_t/D_t)	26.5784	σ_d	0.1837

Table 2
Data characteristics, housing, 1960QI–2010QIV.

Statistic	Value	Statistic	Value
$\widehat{\xi}$	6.6597	Std. Dev. (P_t/R_t)	3.4672
$s.e.(\widehat{\xi})$	0.7911	Corr. (P_t/R_t)	0.9381
$\widehat{\rho}$	0.9943	$r = \frac{\overline{R_t}}{\overline{P_t}}$	0.0478
$s.e.(\widehat{\rho})$	0.0125	$\beta = (1+r)^{-1}$	0.9544
Mean (P_t/R_t)	20.5717	σ_r	0.0409

Fig. 2 plots the data on housing.[3] Relative to Fig. 1, prices and rents are smooth and the series are less volatile.

Data characteristics including the estimates of ξ for the PDR are given in Table 1. We note that in estimating the autoregressive parameter of the dividends specification (ρ) we linearly detrend the dividends process to obtain the d_t series.

The estimate $\widehat{\xi}$ is low suggesting that the tail of the stationary distribution of ϕ_t has only its first few moments and is therefore thicker than that implied by a Normal distribution. The remaining statistics are in line with what is usually reported with post-WWII data although the standard deviation of linearly detrended dividends, σ_d, is higher than usual given that the data coverage includes several periods of volatility. For the PRR in the housing market, we make use of the quarterly, national data on rent and price.

For the housing market, with an estimated $\widehat{\xi}$ over 6, there is much less evidence that the stationary distribution of ϕ_t is different from a Normal distribution.

Next, we estimate the deep parameters $\vartheta = [g\ \gamma\ \beta\ \rho\ \alpha]$ of the model, as follows. First we feed the actual S&P 500 dividend series into our learning model (20)–(22) and estimate the parameters $\vartheta = [g\ \gamma\ \beta\ \rho]$ by minimizing the squared difference between the empirical ξ (reported in Table 1) and those generated by our model. That is, we implement a simulated minimum distance method to estimate ϑ as[4]

$$\min_{\vartheta} \left[\widehat{\xi} - \xi(\vartheta)\right]^2 . \tag{26}$$

The term $\xi(\vartheta)$ is generated as follows. For candidate parametrizations of ϑ we employ the linearly detrended S&P 500 dividends d_t to calculate ϕ_t as per (20)–(22). The ALM (18) then produces a corresponding p_t series which in turn delivers a price–dividend ratio P_t/D_t. We then estimate the ξ associated with the 'simulated' P_t/D_t, using the methods of Clauset et al. (2009) to produce a $\xi(\vartheta)$. The minimization procedure searches over the parameter space of ϑ to implement (26). Table 2 below reports the estimates and associated standard errors resulting from this estimation procedure. We also report associated ξ values obtained by simulating prices using the estimated parameters and the actual dividend data.[5]

The estimates are reasonable: the discount factor β is about 0.97 at an annual frequency, and the dividend process is quite persistent with a coefficient ρ of 0.96. Compared to other estimates in the literature, the risk aversion parameter α of 1.24 is rather small, and the inverse of IES parameter γ of 2.66 (though not precisely estimated) also suggests a modest preference for intertemporal substitution. Finally, the gain parameter g of 0.29 is not too different from that in Benhabib and Dave (2014). With learning we

[3] The data can be downloaded from the Lincoln Institute of Land Policy at http://www.lincolninst.edu/subcenters/land-values/rent-price-ratio.asp.

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

[5] Starting values for the minimization procedure were $\vartheta_0 = [0.25\ 2.5\ 0.96\ 0.96\ 3]$.

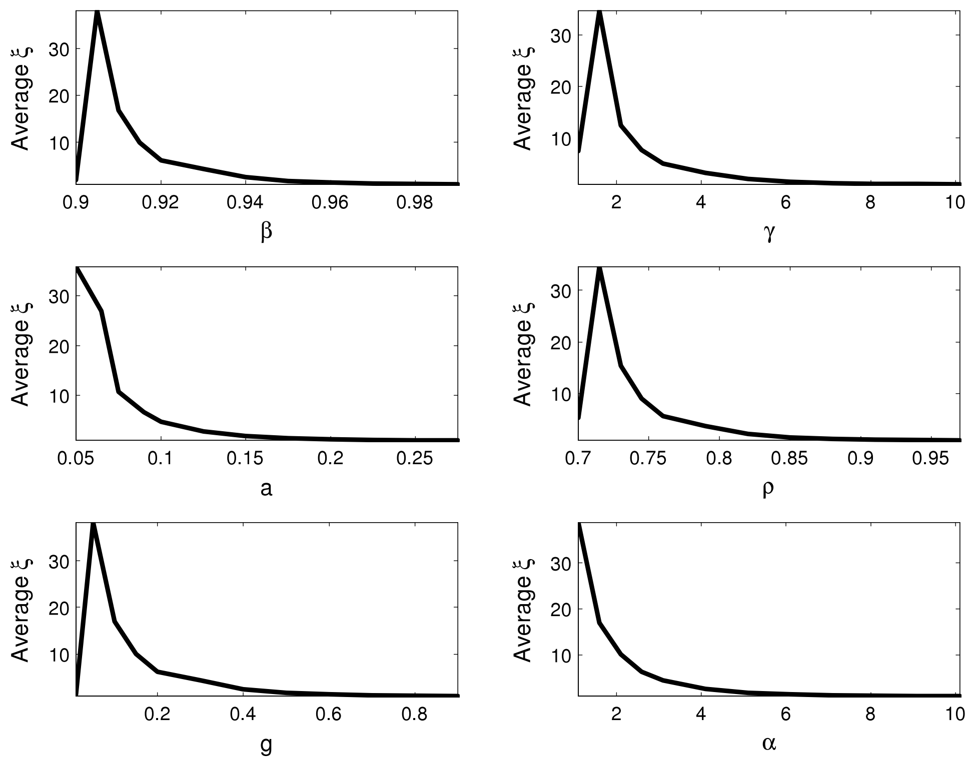

Fig. 3. Simulation results.

Table 3
Parameter estimates, dividends and prices from the S&P 500 (1871QI–2010QIV).

Parameter	Estimate	Std. err.
g	0.2930	0.0029
γ	2.6595	1.5504
β	0.9666	0.0001
ρ	0.9577	0.0997
α	1.2446	0.2813
Associated ξ	3.5979	

Table 4
Parameter estimates, rents and prices from the US housing market (1960QI–2010QIV).

Parameter	Estimate	Std. err.
g	0.3096	6.1908
γ	3.2965	4.8639
β	0.9542	0.1536
ρ	0.9595	0.5263
α	1.1104	0.3821
Associated ξ	6.6597	

no longer require a large CRRA coefficient to explain the data (see Table 4).

For the housing market, the parameter magnitudes are not too different from those for the stock market in Table 3, but they are much less precisely estimated. Only the discount factor β and risk aversion parameter α are significant, and we cannot reject the null hypothesis that each of the other parameters is zero. The results suggest that learning and intertemporal substitution are relatively unimportant for the housing market.

4. Simulations and comparative statics

Our final step is to examine how the parameter governing the tail index ξ varies with the deep parameters of the model. Given a value for the vector $[g\ \gamma\ \beta\ \rho\ \alpha]$ we simulate the learning algorithm that updates ϕ, and estimate ξ from the simulated data using the maximum likelihood procedure of Clauset et al. (2009). We simulate 1000 series, each of length 5000, for ϕ_t. We then feed the simulated series into the model to produce $\{P_t\}$ and $\{P_t/D_t\}$. We then estimate ξ for each simulated series and produce an average ξ over the 1000 simulations. We then vary each element of $(\rho, g, \beta, \gamma, a, \alpha)$ while keeping the others at their baseline (which are the estimates in Table 3). The results are plotted in Fig. 3.

The plots in Fig. 3 show one distinguishing feature. For small values of β, ρ and g the average ξ values are high. This is because $\lambda_t = 1 - g\left(1 - \rho\beta^{\frac{1}{\kappa}}\right) d_{t-1}^2 + g\beta^{\frac{1}{\kappa}} \varepsilon_t d_{t-1}$ includes terms raised to the power of $\frac{1}{\kappa}$ and since the baseline value of κ is 0.1474, it is not clear *a priori* how this affects the support of λ_t. Under recursive preferences the impact of the parameters is no longer monotonic as under the CRRA framework. The lack of monotonic behavior is around implausible values for β and ρ (given the data) prompting the parameter estimates reported in the previous section to choose to settle at values in the monotonic region.

5. Conclusion

We estimate a model of learning under recursive preferences using data for the stock and housing markets. We examine the extent to which asset pricing volatility can be accounted for when the model is written as a linear recursion with multiplicative and additive noise. The results suggest that learning *and* a recursive preference formulation play an important role in the stock market, but there is no such evidence for the illiquid housing market.

References

Benhabib, J., 2013. A note on regime switching, monetary policy and multiple equilibria. In: Sargent, Thomas, Vilmunen, Jouko (Eds.), Macroeconomics at the Service of Public Policy. Oxford Univ. Press, Festschrift for Seppo Honkapohja.
Benhabib, J., Bisin, A., Zhu, S., 2011. The distribution of wealth and fiscal policy in economies with finitely-lived agents. Econometrica 79, 123–157.

Benhabib, J., Dave, C., 2014. Learning, large deviations and rare events. Rev. Econ. Dynam. 17 (3), 367–382.
Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. SIAM Rev. 51 (4), 661–703.
Epstein, Larry G., Zin, Stanley E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica 57 (4), 937–969.
Epstein, Larry G., Zin, Stanley E., 1991. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: an empirical analysis. J. Polit. Econ. 99 (2), 263–286.
Lucas Jr., R.E., 1978. Asset prices in an exchange economy. Econometrica 46 (6), 1429–1445.
Shiller, R.J., 1999. Market Volatility, 6th printing. MIT Press, Cambridge.
Shiller, R.J., 2005. Irrational Exuberance, second ed. Broadway Books.