Online Appendix for "Linearity-Generating Processes: A Modelling Tool Yielding Closed Forms for Asset Prices"

Xavier Gabaix

July 15, 2009

This online appendix discusses a few additional issues related to LG processes. Section 7 provides a simple way to estimate a LG stock process. Section 8 brings theoretical complements, including some additional examples. Section 9 expands the paper's discussion of the use of LG processes to analyze non-LG models.

7 A Simple Estimation of an LG Stock Process

A few papers develop full-fledged methods to estimate LG models. Binsbergen and Koijen (2009) study a LG model with time-varying risk premium and growth rate of stocks. Bekker and Bouwman (2009) estimate a three-factor LG term structure model. Carr, Gabaix and Wu (2009) show how calculate options in closed form in a LG term structure model, and perform an empirical analysis of bonds and fixed-income options.

The present section has the modest goal of presenting a transparent way of estimating a simple, one-factor LG process.

7.1 Method

Here we sketch a simple way to estimate LG processes for stocks. Consider the model of a stock with time-varying risk premium in Example 3. We take $x_t = -(\pi_t - \pi_*)/(1-\pi_*)$ to be the state variable. This makes the estimation easier, because it follows a LG process

$$x_{t+1} = \rho \frac{x_t}{1+x_t} + \varepsilon_{t+1}, \tag{57}$$

where $E_t[\varepsilon_{t+1}] = 0$. The LG moments are:

$$E_t\left[\frac{M_{t+1}D_{t+1}}{M_t D_t}\right] = \alpha(1+x_t), \qquad E_t\left[\frac{M_{t+1}D_{t+1}}{M_t D_t} x_{t+1}\right] = \rho\alpha x_t,$$

and by Theorem 2, the P/D ratio is

$$V_t = P_t/D_t = \frac{\alpha}{1-\alpha}\left(1 + \frac{x_t}{1-\alpha\rho}\right). \tag{58}$$

The goal is to estimate the two parameters α and ρ, and do to goodness of fit tests of the model. We note that we estimate a process, not a full economic model: the LG moments are silent about the nature of the pricing kernel M_t, e.g. which risk factors are priced. They simply present a modelling structure on the dynamics of the factor x_t and its link to observables – here the price-dividend ratio.

We propose two procedures, one via GMM and one via OLS. The GMM procedure is more compact, the OLS procedure is very transparent.

GMM Procedure The factor x_t is a latent variable. Still, it can be expressed $x_t = (1-\alpha\rho)\left(\frac{1-\alpha}{\alpha}V_t - 1\right)$. Hence, its equation of motion (57) can be rewritten $E_t[K_{t+1}] = 0$, where we define:

$$K_{t+1} = \frac{1-\alpha}{\alpha}V_{t+1} - 1 - \rho\frac{\frac{1-\alpha}{\alpha}V_t - 1}{1+\left(\frac{1-\alpha}{\alpha}V_t - 1\right)(1-\alpha\rho)}. \tag{59}$$

So, we obtain an estimate of α and ρ via GMM. For instance, moment conditions can take the form $E_t[K_{t+1}f(V_t)] = 0$ for some function f.

OLS procedure This procedure is a bit approximate, but may be useful as it is transparent. In the continuous-time limit, (58) writes

$$V_t = \frac{1}{R}\left(1 + \frac{x_t}{R+\phi}\right), \qquad R \equiv r + \pi_* - g_*, \tag{60}$$

with $\rho = 1 - \phi\Delta t$.

Step 1: Estimate $V_{t+1} = c + \rho_1 V_t$+noise, which gives a first estimate of ρ, valid when the nonlinearity is not very important (because (57) then implies $E_t[x_{t+1}] \simeq \rho x_t$, hence $E_t V_{t+1} \simeq c + \rho_1 V_t$).

Next, estimate $\widehat{R} = 1/\overline{V_t} \simeq (1-\alpha)/\alpha$, and, inspired by (60), form:

$$y_t \equiv \left(\widehat{R}V_t - 1\right)\left(\widehat{R} + (1-\rho_1)/\Delta t\right).$$

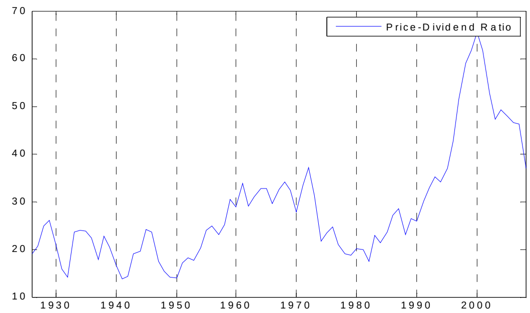

Figure 1: Price-Dividend ratio 1926-2008. Source: CRSP.

By construction, if the non-linear terms are small and the sample large, then $y_t \simeq x_t$.

Step 2: Estimate:

$$y_{t+1} = c + \rho_2 \frac{y_t}{1 + y_t} + \text{noise} \tag{61}$$

As a robustness check, it might be useful to run $y_{t+1} = c + \rho_2' y_t$+noise.

Step 3: Do a specification test. We define, using the coefficient of the regression (61), $L_{t+1} \equiv y_{t+1} - \widehat{\rho}_2 \frac{y_t}{1+y_t}$. With LG processes, we should have $E_t [L_{t+1}] = 0$. To operationalize this, we verify that running:

$$L_{t+1} = c + dy_t + ey_t^2 + \text{noise} \tag{62}$$

yields insignificant coefficients (via an F-test on (c, d, e)).

We next apply the procedure.

7.2 Empirical Application to the US Stock Market

Data We use the CRSP value-weighted index from 1926/12 to 2008/12, sampled at annual frequency. Using the CRSP ex- and cum-dividend returns yields the price-dividend ratio. Figure 1 shows the historical P/D ratio.

To remedy a too large heteroskedasticity of innovations to the P/D ratio, most papers use the log P/D ratio. We use the analogue with LG processes, which is to use GLS, with a weight on a regression being the P/D ratio V_t. For the GMM, we do the tests on $K_{t+1}' \equiv K_{t+1}/V_t$,

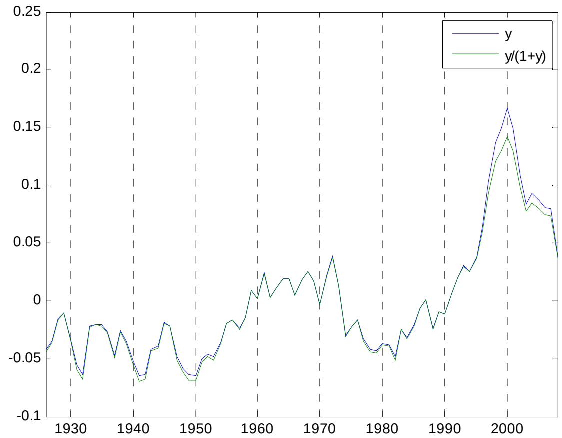

Figure 2: Estimated value y_t of the factor x_t in the process, and its LG transform $y_t/\left(1+y_t\right)$. As expected, the two are very close: their correlation is 0.998, and the correlation in their first time-difference is 0.994. In addition, the LG bound $y_t \geq -\phi$ is verified for the empirically found value of mean-revision, $\phi \simeq 10\%$.

which should also satisfy $E_t\left[K'_{t+1}\right] = 0$.

Results – GMM Procedure We use two moment conditions $E\left[K'_{t+1}\right] = 0$ and $E\left[K'_{t+1}\frac{V_t}{\overline{V}}\right] = 0$. This yields $\alpha = 0.974 \pm 0.0043$ ($\pm$ indicates one standard error), $\rho = 0.901 \pm 0.045$. For the overidentifying restriction, we use the third moment condition $E\left[K'_{t+1}\left(\frac{V_t}{\overline{V}}\right)^2\right] = 0$, which changes the estimates by a quantity less than 0.001. The $p-$value for the $J-$test is 0.14, which means that we do not reject the model.

Results – OLS Procedure Step 1 yields $\rho_1 = 0.908 \pm 0.041$, $R^2 = 0.85$. Also, $\widehat{R} = 1/\overline{V} = 3.52\%$ for $\overline{V} = 28.4$. The resulting estimate of the x_t factor is shown in Figure 2.

Step 2 yields $\rho_2 = 0.910 \pm 0.043$, $R^2 = 0.85$ for the LG regression (61). The non-LG version $y_{t+1} = \alpha + \rho'_2 y_t$ yields $\rho_2 = 0.908 \pm 0.041$, $R^2 = 0.85$. As expected, the LG and non-LG version yield essentially the same results.

Step 3 yields insignificant right-hand side coefficients (the $p-$value is 0.059). Hence, the specification test fails to reject the LG specification.

To complete the characterization of the LG process, we take $\alpha = \overline{V}/\left(1+\overline{V}\right) = 0.966$.

The s.e. on the estimate of $\overline{V}$ is the sample standard error times $\sqrt{(1+\rho)/(1-\rho)}$ (as V_t follows an approximate AR(1) process), so s.e.($\overline{V}$)$= 5.77$. By the delta method, s.e.(α) =s.e.$(\overline{V})\times(1-\alpha)^2 = 0.0067$. This is very close to the GMM estimate.

Concluding Remarks As expected, the values y_t and $y_t/(1+y_t)$ are very close: their correlation is 0.998, and the correlation in their first time-difference is 0.994.

This empirical example turns out to verify easily the working assumptions of the paper. As shown in Figure 2, the LG bound $y_t \geq -\phi$ is verified for the empirically found value of mean-revision, $\phi = 1-\rho_1 = 10\%$. Hence, the LG specification passes the specification test for the P/D ratio of the aggregate stock market index, and satisfies the lower bound restrictions of LG processes.

The reader may wish to read section 9.2, which presents a numerical implementation of the model, including a parametrization of the noise that respects the LG bounds.

8 Some Theoretical Complements

8.1 Some Other Examples

Example 10 *Stock price with stochastic equity premium (continuous time).*

Continuous time Consider two independent Brownian motions B_t, W_t, and a discount factor process and dividend process:

$$\frac{dM_t}{M_t} = -rdt - \frac{\pi_t}{\sigma}dB_t, \qquad \frac{dD_t}{D_t} = g_* dt + \sigma dB_t,$$

with $\pi_t = \pi_* + \widehat{\pi}_t$, where π_* is the constant part of the equity premium, and $\widehat{\pi}_t$ the transitory one, which follows what is the formal limit of (32), when $\rho_\pi = 1-\phi\Delta t$ and $\Delta t \to 0$:

$$d\widehat{\pi}_t = \left(-\phi\widehat{\pi}_t + \widehat{\pi}_t^2\right) dt + \sigma\left(\widehat{\pi}_t\right) dW_t,$$

where volatility $\sigma(\widehat{\pi}_t)$ ensures that a.s. $\widehat{\pi}_t < \phi$ (e.g. σ is sufficiently regular and $\sigma(\phi) = 0$, see Cheridito and Gabaix 2008). Defining $R \equiv r + \pi_* - g_*$, we show that $M_t D_t(1, \widehat{\pi}_t)$ is a

LG process with generator $\begin{pmatrix} R & -1 \\ 0 & R+\phi \end{pmatrix}$, by calculating the LG moments (21)-(22):

$$\begin{aligned} E_t\left[\frac{d(M_tD_t)}{M_tD_t}\right] &= (-r+g_*-\pi_t)\,dt = (-R-\widehat{\pi}_t)\,dt, \\ E_t\left[\frac{d(M_tD_t\widehat{\pi}_t)}{M_tD_t}\right] &= E_t\left[\frac{dM_tD_t}{M_tD_t}\right]\widehat{\pi}_t + E_t\left[d\widehat{\pi}_t\right] \\ &= (-R-\widehat{\pi}_t)\,\widehat{\pi}_t dt + \left(-\phi\widehat{\pi}_t + \widehat{\pi}_t^2\right)dt = -(R+\phi)\,\widehat{\pi}_t dt. \end{aligned}$$

Hence, Theorem 4 yields the price-dividend ratio:

$$P_t/D_t = \frac{1}{R}\left(1 - \frac{\pi_t - \pi_*}{R+\phi}\right).$$

Example 11 *Markov chains, and some economies with learning.*

There are n states. In state i the factor-augmented dividend grows at a rate G_i: $M_{t+1}D_{t+1}/(M_tD_t) = G_i$. Let $X_{it} \in \{0,1\}$ be equal to 1 if the state is i, 0 otherwise. The probability of going from state j to state i is called p_{ij}. Then, $M_tD_t(1, X_1, \ldots, X_n)$ is a LG process. Hence, a Markov chain belongs to the LG class.[16] As many processes are (arbitrarily) well-approximated by discrete Markov chains, they are (arbitrarily) well-approximated by LG processes.

Markov chains induced by learning naturally lead to LG processes. For a complete example, the reader is encouraged to read Veronesi (2000, 2005). He finds that if X_{it} is the agents' probability estimate that the economy is in state i, under canonical models with Gaussian filtration of information, the vector X_t follows an autoregressive process. He works out the prices of stocks and bonds in the economy, and finds that they are linear functions of X_t. Hence, some canonical learning models naturally give rise to LG processes.

Derivation of Example 11 $E_t\left[\frac{M_{t+1}D_{t+1}}{M_tD_t}\right] = \sum_i G_iX_{it}$, and

$$E_t\left[\frac{M_{t+1}D_{t+1}}{M_tD_t}X_{i,t+1}\right] = E_t\left[\frac{M_{t+1}D_{t+1}}{M_tD_t}\right]E_t\left[X_{i,t+1}\right] = \left(\sum_k G_kX_{kt}\right)\left(\sum_j p_{ij}X_{jt}\right) = \sum_j p_{ij}G_jX_{jt},$$

[16] Veronesi (2000), Veronesi and Yared (2000) and David and Veronesi (2006) have already seen that this type of Markov chain yielded prices that are linear in the factors.

as $X_{kt}X_{jt} = 0$ if $j \neq k$, and otherwise is equal to $X_{kt}X_{jt} = X_{jt}$, as exactly one of the X_{jt} is $\neq 0$.

Example 12 *Flexible LG parametrization of state variables and stochastic discount factor.*

A fairly general recipe to construct LG processes is the following. Take an n-dimensional process X_t, such that:

$$\begin{aligned} \frac{M_{t+1}D_{t+1}}{M_tD_t} &= \alpha + \beta' X_t + \varepsilon_{t+1} \\ X_{t+1} &= \frac{\gamma + \Gamma X_t}{a + \beta' X_t} + \eta_{t+1} - \frac{E_t\left[\varepsilon_{t+1}\eta_{t+1}\right]}{a + \beta' X_t}, \end{aligned} \tag{63}$$

with $E_t\left[\varepsilon_{t+1}\right] = 0$, $E_t\left[\eta_{t+1}\right] = 0$. Then, Eq. 8-9 are satisfied. Section 5.1 provides conditions to ensure $M_tD_t > 0$ for all times.

To interpret (63), consider the case $\gamma = E_t\left[\varepsilon_{t+1}\eta_{t+1}\right] = 0$. Eq. 63 then expresses that, when X_t is small, $E_t\left[X_{t+1}\right] = \frac{\Gamma X_t}{\alpha+\beta' X_t} \sim \frac{\Gamma}{\alpha} X_t$, which means that X_t follows approximately an AR(1). The corrective $1 + \beta'/\alpha \cdot X_t$ in the denominator is often small in practice, but ensures that the process is LG.

Example 13 *A LG process where the stock price is convex (not linear) in the growth rate of dividends.*

This "academic" example shows how one can obtain asset prices that are increasing in their variance, which is important in some applications (Johnson 2002, Pastor and Veronesi 2003). Consider an economy with constant discount rate r (i.e. $M_t = e^{-rt}$), and a stock with dividend $D_t = D_0 \exp\left(\int_0^t g_s ds\right)$, where[17] $dg_t = -\left(g_t^2/2 + \phi g_t\right)dt + \sqrt{k\left(G^2 - g_t^2\right)}dW_t$. Calculation shows that $e^{-rT}D_t\left(1, g_t, g_t^2\right)$ is a LG process with generator

$$\omega = \begin{pmatrix} r & -1 & 0 \\ 0 & r+\phi & -1/2 \\ -kG^2 & 0 & 2\phi + k + 1 \end{pmatrix}.$$

Hence by Theorem 4, the price-dividend ratio is:

$$P_t/D_t = \frac{2(\phi + r)(2\phi + k + r) + 2(2\phi + k + r)g_t + g_t^2}{2r\left(\phi + r\right)\left(2\phi + k + r\right) - kG^2}, \tag{64}$$

[17] We assume $0 < G < 2\left(\phi - k\right)$, and that the support of g_t is $[-G, G]$, with end points natural boundaries.

which is increasing in the parameter G of the volatility. In this example, the state vector is (g_t, g_t^2), which makes the price quadratic and convex in g_t.

8.2 LG Processes are the Only Ones that Yield Linearity

The paper shows that, in a certain sense, if bond prices are linear in the factors, then they come from an LG process. To see that, let us first consider the 1-factor case. Suppose that for $T = 1, 2$, $Z_t(T) = \alpha_T + \beta_T X_t$, for some numbers α_1, $\beta_1 \neq 0$, α_2, β_2. With $T = 1$, we get $E_t[M_{t+1}/M_t] = \alpha_1 + \beta_1 X_t$, so that condition (8) holds. Also,

$$\begin{aligned} \alpha_2 + \beta_2 X_t &= E_t\left[\frac{M_{t+2}}{M_t}\right] = E_t\left[\frac{M_{t+1}}{M_t} E_{t+1}\left[\frac{M_{t+2}}{M_{t+1}}\right]\right] = E_t\left[\frac{M_{t+1}}{M_t}(\alpha_1 + \beta_1 X_{t+1})\right] \\ &= \alpha_1(\alpha_1 + \beta_1 X_t) + \beta_1 E_t\left[\frac{M_{t+1}}{M_t} X_{t+1}\right] \end{aligned}$$

$$\Rightarrow E_t\left[\frac{M_{t+1}}{M_t} X_{t+1}\right] = \frac{1}{\beta_1}(\alpha_2 + \beta_2 X_t - \alpha_1(\alpha_1 + \beta_1 X_t)) = a'' + b'' X_t$$

hence (9) holds. We conclude that if both the 1 and 2-period maturity bonds are affine in X_t, then $M_t(1, X_t)$ is a LG process.

Proposition 6 shows that the property holds with n factors.[18]

8.3 Plug-and-Verify Derivation for LG stocks

There is a elementary heuristic proof for the expression of stock prices. We seek a solution of the type $P_t/D_t \equiv V_t = c - 1 + h' X_t$, which we know exists, by summation of (12). The no-arbitrage equation is:

$$\frac{P_t}{D_t} = \frac{1}{D_t} E_t\left[\frac{M_{t+1}}{M_t}(D_{t+1} + P_{t+1})\right] = E_t\left[\frac{M_{t+1}}{M_t}\frac{D_{t+1}}{D_t}\left(1 + \frac{P_{t+1}}{D_{t+1}}\right)\right]$$

[18] The property that $\{(\alpha_T, \beta_T), T = 1, 2..\}$ spans $\mathbb{R}^{n+1}$ means that $E_t[M_{t+T}/M_t] = \alpha_T + \beta_T' X_t$ is the most compact representation of the process. More precisely, if it did not span $\mathbb{R}^{n+1}$, one could find a stricly lower dimensional process $x_t \in \mathbb{R}^m$, $m < n$, and constants A_T, B_T, such that $E_t[M_{t+T}/M_t] = A_T + B_T' x_t$. Indeed, call $\gamma_T = (\alpha_T, \beta_T')'$, and $V = Span\{\gamma_T, T \geq 0\}$. If V is a strict subset of $\mathbb{R}^{n+1}$, decompose $\mathbb{R}^{n+1} = V \oplus V^\perp$, call $B : V \to \mathbb{R}^{n+1}$ the natural injection, and $\langle \cdot, \cdot \rangle$ the restriction of the Euclidean product on V. Then, $\gamma'(T) Y_t = (B\gamma(T))' Y_t = \gamma(T)' B' Y_t$, so we have $Z_t(T) = \gamma(T)'(B' Y_t)$. Vector $B' Y_t$ has dimension $\dim V < n + 1$.

i.e.

$$\begin{aligned} c-1+h'X_t &= V_t = E_t\left[\frac{M_{t+1}}{M_t}\frac{D_{t+1}}{D_t}\left(1+V_{t+1}\right)\right] = E_t\left[\frac{M_{t+1}}{M_t}\frac{D_{t+1}}{D_t}\left(c+h'X_t\right)\right] \\ &= c\left(\alpha+\delta'X_t\right)+h'\left(\gamma+\Gamma X_t\right) \\ &= \left(c\alpha+h'\gamma\right)+\left(c\delta'+h'\Gamma\right)X_t \end{aligned}$$

i.e. (i) $c-1=c\alpha+h'\gamma$ and (ii) $h'=c\delta'+h'\Gamma$. (ii) gives $h'=c\delta'\left(I_n-\Gamma\right)^{-1}$, and plugging in (i) yields $c\left[1-\alpha-\delta'\left(I_n-\Gamma\right)^{-1}\gamma\right]=1$, hence c and (14).

8.4 Processes with Time-Dependent Coefficients

It is simple to extend the process to time-dependent deterministic coefficients, i.e. in Definition 2, to have $\alpha, \delta, \gamma, \Gamma$ functions of time. With $Y_t=\left(M_t, M_tX_t\right)^\top$, this is $E_t\left[Y_{t+1}\right]=\Omega_tY_t$, where $\Omega_t=\begin{pmatrix}\alpha_t & \delta_t' \\ \gamma_t & \Gamma_t\end{pmatrix}$. That implies $E_0\left[Y_T\right]=\prod_{t=0}^{T-1}\Omega_tY_0$. Hence, in the zero-coupon expressions, it is enough to replace Ω^T by $\prod_{t=0}^{T-1}\Omega_t$.

8.5 Closedness Under Addition and Multiplication

The product of two uncorrelated LG processes is LG. The product of two uncorrelated LG processes with respective dimensions d_1, d_2 (i.e., with d_1-1 and d_2-1 factors respectively) is LG, with dimension d_1d_2 (i.e., with d_1d_2-1 factors). The idea is simple, though it requires somewhat heavy notations.

We start in discrete time. Take two LG processes characterized by M_t^k, Y_t^k, Ω^k $(k=1,2)$ (we drop the D notation for simplicity) and consider a process with stochastic discount factor $M_t=M_t^1M_t^2$. Assume that, for any index i, j of the components, $cov_t\left(Y_{i,t+1}^1, Y_{j,t+1}^2\right)=0$. The innovations between processes are uncorrelated, but, importantly, not necessarily independent. Then, it is easy to verify that for any vector ψ^i, $E_t\left[\left(\psi^1Y_T^1\right)\left(\psi^2Y_T^2\right)\right]=E_t\left[\psi^2Y_T^2\right]E_t\left[\psi^2Y_T^2\right]$. In particular, $E_t\left[M_T^1M_T^2\right]=E_t\left[M_T^1\right]E_t\left[M_T^2\right]$.

Then, $M_t=M_t^1M_t^2$ is also the stochastic discount factor of a LG process. The underlying autoregressive process is $Y_t^1\otimes Y_t^2$, i.e. the vector made of the d_1d_2 components $Y_{it}^1Y_{jt}^2$, $i=1,\ldots,d_1, j=1,\ldots,d_2$. Because $E_t\left[Y_{i,t+1}^1Y_{j,t+1}^2\right]=\left(\sum_k\Omega_{ik}^1Y_k^1\right)\left(\sum_l\Omega_{jl}^2Y_l^2\right)=\sum_{k,l}\Omega_{ik}^1\Omega_{jl}^2Y_{kt}^1Y_{lt}^2$, the corresponding generator Ω is $\Omega_{ij,kl}=\Omega_{ik}^1\Omega_{jl}^2$, i.e., $\Omega=\Omega^1\otimes\Omega^2$.

The same reasoning holds in continuous time. Starting with processes M_t^k, Y_t^k, ω^k $(k = 1, 2)$, and assuming $\left\langle dY_{it}^1, dY_{jt}^2 \right\rangle = 0$, then $M_t^1 M_t^2$ is also a pricing kernel that comes from a LG process. The underlying autoregressive process is $Y_t^1 \otimes Y_t^2$. Because $E_t\left[d\left(Y_{it}^1 Y_{jt}^2\right)\right] = E_t\left[dY_{it}^1\right] Y_{jt}^2 + Y_{it}^1 E_t\left[dY_{jt}^2\right]$, the generator of the process is

$$\omega_{ij,kl} = \omega_{ik}^1 \delta_{jl} + \delta_{ik} \omega_{jl}^2 \tag{65}$$

where δ_{jl} is the Kronecker delta, i.e. $\omega = \omega^1 \otimes I_{d_2} + I_{d_1} \otimes \omega^2$. To make the above concrete, consider the following example.

Example 14 *Stock with decoupled LG processes for the growth rate and the risk premium.*

Consider processes with $dM_t/M_t = -rt - \lambda_t dB_t$, $dD_t/D_t = g_t dt + \sigma_t dB_t$, where $g_t = g_* + \widehat{g}_t$, which follows: $d\widehat{g}_t = -\phi_g \widehat{g}_t dt - \widehat{g}_t^2 dt + dN_t^g$, and the risk premium, $\pi_t = \lambda_t \sigma_t$ is decomposed $\pi_t = \pi_* + \widehat{\pi}_t$, which follows: $d\widehat{\pi}_t = -\phi_\pi \widehat{\pi}_t dt + \widehat{\pi}_t^2 dt + dN_t^\pi$, where N_t^g, N_t^π are martingales. Assume that the processes dN_t^g, dN_t^π and dB_t are uncorrelated. Then, the price of a stock, $P_t = E_0\left[\int_0^\infty M_{t+s} D_{t+s} ds\right] / M_t$, is $P_t/D_t = E_t\left[\int_{s=t}^\infty \exp\left(-\int_{u=t}^s (r + \pi_u - g_u)\, du\right) ds\right]$. In virtue of the properties of LG processes under multiplication that we just saw,

$$E_t\left[\exp\left(\int_t^s -\pi_u + g_u du\right)\right] = E_t\left[\exp\left(\int_t^s -\pi_u du\right)\right] E_t\left[\exp\left(\int_t^s g_u du\right)\right].$$

For general processes, the above equation would in general require the two processes to be independent – for instance, with stochastic volatility, the respective variance processes should be independent. For LG processes, the property required is the weaker: $\langle d\pi_t, dg_t \rangle = 0$ for all t's.

Then, using (65) or direct calculations, $M_t D_t \left(1, \widehat{\pi}_t, \widehat{g}_t, \widehat{\pi}_t \widehat{g}_t\right)$ is LG with generator:

$$\omega = RI_4 + \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & \phi_\pi & 0 & -1 \\ 0 & 0 & \phi_g & 1 \\ 0 & 0 & 0 & \phi_\pi + \phi_g \end{pmatrix},$$

with $R = r + \pi_* - g_*$,[19] so that: by Theorem 4,

$$P_t/D_t = \frac{1}{R}\left[1 - \frac{\widehat{\pi}_t}{R+\phi_\pi} + \frac{\widehat{g}_t}{R+\phi_g} - \frac{\left(2R+\phi_\pi+\phi_g\right)\widehat{\pi}_t\widehat{g}_t}{\left(R+\phi_\pi\right)\left(R+\phi_g\right)\left(R+\phi_\pi+\phi_g\right)}\right]. \tag{66}$$

The central value is again the Gordon formula, $P_t/D_t = 1/R$. It is modified by the current level of the equity premium, and the growth rate of the stock.[20] A stock with a currently high growth rate g_t exhibits a higher price-dividend ratio, and this is amplified when the equity premium is low, as shown by the term $\widehat{\pi}_t\widehat{g}_t$.

The difference between formula (66) and formula (35) is that in (66), the processes for π_t and g_t are decoupled, whereas in (35), they were coupled, i.e. in their drift term there was a term $\widehat{g}_t$. The decoupling forces the presence of a cross term $\widehat{\pi}_t\widehat{g}_t$ in the expression of the price. In general, one obtains simpler expressions by having one multifactor LG processes, rather than the product of many different LG processes. With n coupled factors, the stock price has $n+1$ terms, while with n decoupled factors, the stock price has 2^n terms.

The sum of two LG processes is LG. This property is quite trivial, and mentioned for completeness. Consider two LG processes M_t^i, Y_t^i, Ω^i, with $M_t^i = \nu^i Y_t^i$, for $i = 1, 2$. Denote by d_i the dimension of Y_t^i. Then, $M_t = M_t^1 + M_t^2$ comes from a LG process of dimension $d_1 + d_2$. Indeed, define $Y_t = (Y_t^1, Y_t^2)$, $\nu = (\nu^1, \nu^2)$, and $\Omega = \begin{pmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{pmatrix}$. Then, $E_t\left[Y_{t+1}\right] = \Omega Y_t$, and $M_t = \nu' Y_t$.

8.6 Potential Escape Dynamics if the Conditions on Variance are not Respected

Section 5.1 provides sufficient conditions to keep the process well-defined, e.g. to keep prices positive and all quantities finite. For instance, consider the basic LG process in continuous

[19] Menzly, Santos and Veronesi (2004, Eq. 20) obtain a similar expression. This is natural because the LG class embeds their model, as Example 9 shows.

[20] This line of reasoning suggests the following non-LG variant. Suppose we have a process with $d\psi_t = \left(r_t\psi_t + \alpha r_t - \beta\right)dt + dN_t$,where dN_t is an adapted martingale, and is essentially arbitrary except for technical conditions. Then $V_t = \left(\psi_t + \alpha\right)/\beta$ is a solution of the perpetuity arbitrage equation: $1 - r_tV_t + E\left[dV_t\right]/dt = 0$. If the process well-defined for $t \geq 0$, then V_t is the price of a perpetuity, $V_t = E_t\left[\int_t^\infty e^{-\int_t^s r_u du} ds\right]$. For instance, with the process $d\left(1/r_t\right) = \phi\left(r_t - r^*\right)dt + dN_t$, the price of a perpetuity is: $V_t = \left(1/r_t + \phi/r_*\right)/\left(1+\phi\right)$.

time:

$$dg_t = -\left(\phi g_t + g_t^2\right) dt + \sigma\left(g_t\right) dB_t \tag{67}$$

For the process to be well-defined so that $g_t > -\phi$ a.s., the $\sigma(g)$ function has to go to 0 near $g = -\phi$.

However, it is instructive to examine the "escape dynamics" if $\sigma(g)$ is kept constant (against the prescription of this paper). Then, g_t can escape the region $(-\phi, \infty)$ at a random time τ. We examine the properties of the escape by considering $T^e(\phi, \sigma)$, the expected value of the time at which g_t would cross the boundary, $-\phi$, given it starts at 0. The following Lemma calculates its value.

Lemma 2 *If volatility is constant in process (67) (as opposed to going to zero at the boundary), the expected time till escape is:* $T^e(\phi, \sigma) = \frac{1}{\phi} F\left(\frac{\sigma^2}{2\phi^3}\right)$*, where we define:*

$$F(K) = \frac{1}{K} \int_0^1 \frac{\int_y^\infty \psi_K(x)\, dx}{\psi_K(y)} dy, \qquad \psi_K(y) = \exp\left(\frac{y^2}{2K} - \frac{y^3}{3K}\right). \tag{68}$$

Proof. First, we reduce the problem by dimensional analysis. Defining $y_t = g_t/\phi + 1$, we have $dy_t = -\phi(y_t - 1) y_t dt + \frac{\sigma}{\phi} dB_t$, where ϕ has the units of $[\text{time}]^{-1}$ and σ has the unit of $[\text{time}]^{-3/2}$. So, T^e can be written $T^e = \phi^{-1} F\left(\frac{\sigma^2}{2\phi^3}\right)$ for some function F to be determined (the prefactor 2 is simply for convenience), and where $K = \frac{\sigma^2}{2\phi^3}$ is a dimensionless parameter.

By homogeneity, it is enough to consider the case $\phi = 1$. Call $T(y)$ the expected time till reaching the boundary 0, given the process y_t starts at a value y. By the usual reasoning (whose heuristic form is $T(y_t) = E[T(y_{t+dt})] + dt$ for $y_t > 0$), we have:

$$-T'(y)\, y(y+1) + T''(y) K + 1 = 0.$$

This integrates to $T'(y) = \frac{\int_y^\infty \psi_K(x)dx}{K\psi_K(y)}$, as $\lim_{y\to\infty} T'(y) < +\infty$. Because $T(0) = 0$, we can integrate: $T(Y) = \int_0^Y \frac{\int_y^\infty \psi_K(x)dx}{K\psi_K(y)} dy$. Finally, $T^e = E[\tau \mid g_0 = 0] = E[\tau \mid y_0 = 1] = T(1)$. ■

To understand intuitively the behavior of the process, consider first approximating it by an Ornstein-Uhlenbeck process, $dg_t \simeq -\phi g_t dt + \sigma dB_t$. Then, the steady state distribution of g_t is Gaussian, with mean 0 and standard deviation $\sigma/\sqrt{2\phi}$. It is "safely" away from the boundary $-\phi$ if this boundary is, say, more than 3 standard deviation away from the mean. This corresponds to: $\phi > 3\sigma/\sqrt{2\phi}$, i.e. $\sigma^2/\phi^3 < 0.2$.

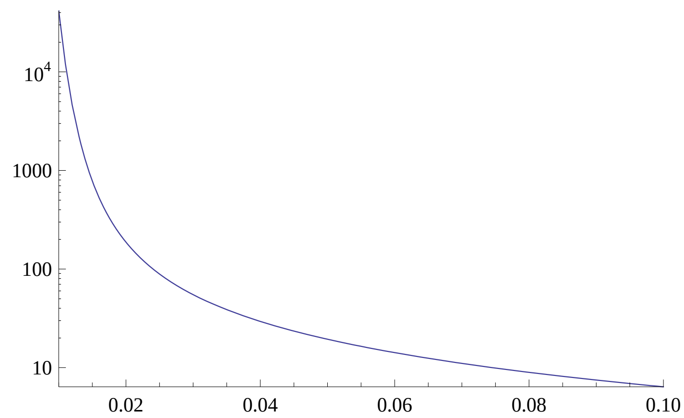

Figure 3: This Figure plots the expected time to escape for a LG process that would have a constant volatility. Note that this is against the prescription of this paper, which is that the volatility should go to 0 near the boundary, $-\phi$. The volatility σ is on the horizontal axis, the escape time T^e on the vertical axis (which is a log axis). Units are annual and $\phi = 0.13$.

We obtain a conclusion: we expect the time to escape function $F(K)$ (with $K = \sigma^2/\left(2\phi^3\right)$) to become very large for $K < 0.1$. More precisely, Lemma 2 gives:

Lemma 3 *If $\sigma^2/\phi^3 < 0.2$, then the expected time to escape is greater than $20/\phi$. If $\sigma^2/\phi^3 < 0.1$, it is greater than $109/\phi$.*

This is confirmed numerically: $F(K)$ is $5 \cdot 10^7, 36, 20, 2.8$ and 1.5 for $K = 0.01, 0.075, .1, .5$ and 1 respectively.

We consider the value $\phi = 0.13$ (the value calibrated in Gabaix (2009), see the explanations therein, and which is in the confidence interval of the estimation of section 7). Then, for $\sigma = 0.1\%$, 1%, 2%, 3%, 4%, and 10% respectively, T^e is equal to 10^{319}, $41{\cdot}10^3$, 187, 55, 29 and 6.4 year respectively.[21] Figure 3 plots this dependence. Note that the corresponding volatility of the P/D ratio is $\sigma/(R+\phi)$, which for a D/P ratio of 5% corresponds to a volatility coming purely from variation in the equity premium of 0.5%, 5%, 11%, 16%, 22% and 55% respectively. The corresponding values of σ^2/ϕ^3 are $4 \cdot 10^{-4}, 0.04, .18, .4, .72$, and 4.5.

We conclude that whenever σ^2/ϕ^3 is less than about 0.2, then the issues of escape dynamics are quite minor, or even trivial. Beyond that, it is important in numerical applications to

[21] The behavior of the time to escape when volatility is small is analyzed in Gardiner (2003, section 5.2).

consider the specifications to make the noise to go 0 near the boundary proposed in section 5.1 of this paper.

9 Approximating non-LG processes with LG processes

This section offers numerical complements to the scheme proposed in section 5.2 in the paper.

9.1 Variants on the Basic LG Projection

We study variants on the basic truncation, using the Ornstein-Uhlenbeck model $dx_t = -\phi x_t dt + \sigma dW_t$, $V(x) = E_{x_0=x}\left[\int_0^\infty e^{-RT+\int_0^T x_s ds} dT\right]$ as an example. We start by a simple, intuitive procedure, then propose a Hermite basis, then proceed to a numerical comparison.

9.1.1 Shifted Basic Truncation

The basic truncation is the one mentioned in the body of the paper. We review it here. The infinite-dimensional generator of the process $e^{\int_0^t x_s ds}(1, x_t, x_t^2, ...)$ associated with the OU has a generator given by (51):

$$\omega = \begin{pmatrix} R & -1 & 0 & 0 & \cdots \\ 0 & R+\phi & -1 & 0 & \cdots \\ -\sigma^2 & 0 & R+2\phi & -1 & \cdots \\ 0 & -3\sigma^2 & 0 & R+3\phi & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

So the first two terms of the LG projection are:

$$\begin{aligned} V^{[1]}(x) &= (1,0)\begin{pmatrix} R & -1 \\ 0 & R+\phi \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ x \end{pmatrix} \\ V^{[2]}(x) &= (1,0,0)\begin{pmatrix} R & -1 & 0 \\ 0 & R+\phi & -1 \\ -\sigma^2 & 0 & R+2\phi \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix} \end{aligned}$$

i.e.,

$$V^{[1]}(x) = \frac{1}{R}\left(1+\frac{x}{R+\phi}\right) \tag{69}$$

$$V^{[2]}(x) = \frac{1}{R-\frac{\sigma^2}{(R+\phi)(R+2\phi)}}\left(1+\frac{x}{R+\phi}+\frac{x^2}{(R+\phi)(R+2\phi)}\right). \tag{70}$$

A very simple method to obtain an alternative approximation is the following: Take the term (70), and drop the x^2 term: we get a first-order approximation to the stock price:

$$V^{[1]\prime}(x) = \frac{1}{R-\frac{\sigma^2}{(R+\phi)(R+2\phi)}}\left(1+\frac{x}{R+\phi}\right).$$

We will see that this expression is often more accurate than the basic expression $V^{[1]}(x)$. More general, the "shifted" procedure is to take the basic truncation $V^{[m+1]}$, which is a polynomial of degree $\leq m+1$ on x; drop the term of degree $m+1$; and call the resulting polynomial, of degree at most m, $V^{[m]\prime}(x)$. Under the assumption in the paper, it converges to the true price. For instance, from dropping the x^3 term in $V^{[3]}(x)$ we obtain:

$$V^{[2]\prime}(x) = \frac{1}{R-\frac{\sigma^2}{(R+\phi)(R+2\phi)-\frac{3\sigma^2}{R+3\phi}}}\left(1+\frac{x}{R+\phi-\frac{3\sigma^2}{(R+2\phi)(R+3\phi)}}+\frac{x^2}{(R+\phi)(R+2\phi)-\frac{3\sigma^2}{R+3\phi}}\right).$$

There are other procedures and bases that may be of interest. We describe them now.

9.1.2 Using Hermite Polynomials as a Basis

Instead of projecting in the basis $(1, x, x^2, \ldots)$, we can project in the basis $(1, H_1(x), H_2(x), \ldots)$, where H_k are Hermite polynomials. This seems like a natural thing to do, as when dealing with Gaussian variables, the Hermite polynomials usually have good properties.

Pick a $S > 0$, which has the units of an interest rate, and the Hermite polynomials $H_k(x) \equiv H_k^*(x/S)\, S^k$, where H_k^* are the standard Hermite polynomials, defined by $H_k^*(x) = (-1)^n e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}$. In particular, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - S^2$, $H_3(x) = x^3 - 3S^2 x$, $H_4(x) = x^4 - 6S^2x^2 + 3S^4$.

To pick the value of S, recall that if $X \sim N(0, S^2)$, then for all $k > 0$, $E[H_k(X)] = 0$.

As the OU process has a steady state distribution $N\left(0, \sigma^2/\left(2\phi\right)\right)$, we set

$$S = \frac{\sigma}{\sqrt{2\phi}}.$$

This way, $E\left[H_k\left(X\right)\right] = 0$ for $k > 0$.

Denote by Q the basis transformation matrix between basis $(1, x, x^2, \ldots)$ and basis $(1, H_1\left(x\right), H_2\left(x\right), \ldots)$. For instance, the third row of Q is $(-S^2, 0, 1, 0, 0, \ldots)$, the coefficients of $H_2\left(x\right)$.

Define $Y_t^H = QY_t$, the LG process expressed in the Hermite basis. We have $E\left[dY_t^H\right] = -\omega^H Y_t^H dt$, with $\omega^H = Q\omega Q^{-1}$.

We can define the LG truncation as in the power basis:

$$V^{m,H} = (1, 0, 0, \ldots)\left(\left(\omega^H\right)^{[m]}\right)^{-1}\left(1, H_1\left(x\right), \ldots, H_m\left(x\right)\right)'.$$

For instance, take the Ornstein-Uhlenbeck example of section 5.2. Calculations show that ω^H has a simple expression:

$$\omega^H = \begin{pmatrix} R & -1 & 0 & 0 & 0 & \cdots \\ -S^2 & R+\phi & -1 & 0 & 0 & \cdots \\ 0 & -2S^2 & R+2\phi & -1 & 0 & \cdots \\ 0 & 0 & -3S^2 & R+3\phi & -1 & \cdots \\ 0 & 0 & 0 & -4S^2 & R+4\phi & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

So the first two terms are:

$$\begin{aligned} V^{1,H}\left(x\right) &= (1, 0)\begin{pmatrix} R & -1 \\ -S^2 & R+\phi \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ x \end{pmatrix} \\ V^{2,H}\left(x\right) &= (1, 0, 0)\begin{pmatrix} R & -1 & 0 \\ -S^2 & R+\phi & -1 \\ 0 & -2S^2 & R+2\phi \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ x \\ x^2 - S^2 \end{pmatrix} \end{aligned}$$

i.e.,

$$V^{1,H}(x) = \frac{1+\frac{x}{R+\phi}}{R-\frac{\sigma^2}{(R+\phi)2\phi}}, \tag{71}$$

$$V^{2,H}(x) = \frac{1-v'+\frac{x}{R+\phi}+\frac{x^2}{(R+\phi)(R+2\phi)}}{R-(R+2\phi/3)\,v'}, \qquad v' = \frac{3\sigma^2}{2\phi(R+\phi)(R+2\phi)}. \tag{72}$$

We will see that in the OU case, the Hermite approximation converges very fast.

We note that other schemes might be useful, e.g. using the Galerkin method (Canuto et al. 2006), or perhaps the Taylor expansion method in Kristensen and Mele (2009).

9.1.3 An Alternative Intuitive Truncation

Theorem 5 analyzes a very stark truncation scheme. In practice, it is easy to do better. Consider the Ornstein-Uhlenbeck case with $D_t = \exp\left(\int_0^t x_s ds\right)$ and $dx_t = -\phi x_t dt + \sigma dW_t$ (x_t can be a growth rate, or minus a risk premium, expressed under the risk-neutral measure). We define $Y_t^1 = e^{-Rt}D_t$, and $Y_t^2 = e^{-Rt}D_t x_t$. We have: $E_t[dY_{1t}]/dt = (-R+x_t)\,Y_{1t} = -RY_{1t}+Y_{2t}$ and $dY_{2t}/dt = Y_{1t}\left(-(\phi+R)\,x_t + x_t^2\right)$.

To approximate x_t^2, we replace it by its steady state mean. To find it, we observe that $E_t[dx_t^2]/dt = -2\phi x_t^2+\sigma^2$, so that taking the expectation at time 0, we obtain $\lim_{t\to\infty} E_0[x_t^2] = \sigma^2/(2\phi)$. Hence we approximate $dY_{2t} \simeq Y_{1t}\left(-(\phi+R)\,x_t + \sigma^2/(2\phi)\right)$, and we approximate Y_t by Y_t^*, where Y_t^* follows: $E_t[dY_t^*] = -\omega^{(1)}Y_t^*$ with:

$$\omega^{(1)} = \begin{pmatrix} R & -1 \\ -\sigma^2/(2\phi) & R+\phi \end{pmatrix}. \tag{73}$$

Applying Theorem 4, we obtain:

$$V^{(1)}(x) = \frac{1+\frac{x}{R+\phi}}{R-\frac{\sigma^2}{(R+\phi)2\phi}}. \tag{74}$$

This is exactly the expression obtained with the Hermite truncation. We can proceed further.

Second-Order Terms To study second-order terms, we study the "pure" behavior of the process, without the effect of discounting, we look at the process when $R = 0$ to see which terms to drop and approximate. Then:

$$E_t\left[dY_{2t}\right]/dt = -\phi Y_{2t} + Y_{3t} \tag{75}$$

$$E_t\left[dY_{3t}\right]/dt = \sigma^2 Y_{1t} - 2\phi Y_{3t} + Y_{4t}, \tag{76}$$

Let us study the truncation to two terms. We replace Y_3 by the value that ensures the terms in (76) are equal to 0 (that is sometimes called the "adiabatic principle", see Gardiner (2003)), so $Y_3 \simeq \sigma^2 Y_{1t}/\left(2\phi\right)$, and injecting this in (75), with the discounting, we get: $E_t\left[dY_{2t}\right]/dt \simeq \frac{\sigma^2 Y_{1t}}{2\phi} - \left(\phi + R\right) Y_{2t}$. Hence, the LG approximation is the one in expression (73).

Let us do the same for the x^2 terms. For instance, with $R = 0$, $E_t\left[dY_{4t}\right]/dt = 3\sigma^2 Y_{2t} - 3\phi Y_{4t} + Y_{5t}$, so we replace $Y_{4t} \simeq Y_{2t}\sigma^2/\phi$. Injecting this in (76) yields the generator for the LG approximation:

$$\omega^{(2)} = \begin{pmatrix} R & -1 & 0 \\ 0 & R+\phi & -1 \\ -\sigma^2 & -\frac{\sigma^2}{\phi} & R+2\phi \end{pmatrix}.$$

The corresponding approximation is (77), by Theorem 4, $V_t^{(2)} = (1,0,0)\left[\omega^{(2)}\right]^{-1}(1, x_t, x_t^2)$, i.e.:

$$V^{(2)}(x) = \frac{1}{R} + \frac{x}{R\left(R+\phi\right)} + \frac{x^2 + \frac{\sigma^2}{R\phi}\left(x+\phi\right)}{R\left(R+\phi\right)\left(R + 2\phi - \frac{\sigma^2}{R\phi}\right)} \tag{77}$$

Approximation of arbitrary order Pick a truncation order m. We write $\omega^{[m+1]} = \begin{pmatrix} \omega^{[m]} & b \\ c' & D \end{pmatrix}$, and we use the approximation (coming heuristically from $Y_{m+2,t} \simeq -\frac{1}{D-R}c'\left(Y_{1t}, \ldots, Y_{m+1,t}\right)$, i.e., $\omega^{(m)} = \omega^{[m]} - \frac{1}{D-R}bc'$. This corresponds to the two operations that lead to $\omega^{(1)}$ and $\omega^{(2)}$ above. The LG price is simply

$$V_t^{(m)} = (1, 0, .., 0)\left[\omega^{(m)}\right]^{-1}(1, x, \ldots, x^m)'.$$

9.1.4 Numerical Evaluation of the four methods

To match the mean and volatility of the PD ratio, we calibrate the OU with $R = 3.5\%$, $\phi = 13\%, \sigma = 1.8\%$, all in annual units. That leads to a volatility of the log PD ratio of 10.3% per year.

Table 1: Mean Approximation Error with LG processes

		Mean Relative Error		
Number of factors in LG process	Shifted Basic	Basic	Hermite	Intuitive
1	$3.6 \cdot 10^{-2}$	$1.7 \cdot 10^{-1}$	$2.2 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$
2	$7.3 \cdot 10^{-3}$	$1.7 \cdot 10^{-2}$	$3.0 \cdot 10^{-3}$	$6.1 \cdot 10^{-3}$
3	$7.5 \cdot 10^{-4}$	$5.9 \cdot 10^{-3}$	$3.6 \cdot 10^{-4}$	$4.2 \cdot 10^{-4}$

Explanation: This Table shows the mean approximation error $E\left[\left|V^m(x) - V(x)\right|\right]/E\left[V(x)\right]$, averaging over the steady state distribution of the OU process, for various schemes proposed in this paper: the basic, $V^{[m]}$, the shifted basic $V^{[m]'}$, the truncation on a Hermite basis $V^{m,H}$, and the "intuitive" truncation $V^{(m)}$.

Table 1 shows the results for the mean approximation error $E\left[\left|V^m(x) - V(x)\right|\right]/E\left[V(x)\right]$, averaging over the steady state distribution of the OU process. The best performance is for the Hermite basis. The intuitive truncation procedure performs quite similarly. The simplest truncation's fit is a bit off with just 1 term. That's because it simply replaces $E\left[x_t^2\right]$ by 0, whereas the intuitive and Hermite truncation replace it by its mean, $\sigma^2/(2\phi)$, hence do better. However, the "shifted" simplest truncation has a better fit. Even for just the first order term, the average error is about 2% of the daily stock price movement. It would be already fantastic if finance could explain stock prices within 2%.

In addition, Figure 4 plots the LG approximation, and the exact expression.

We conclude that the first order approximation of the Ornstein-Uhlenbeck process by a LG process will be rather good, and useful for theoretical purposes.

The simplest, accurate procedure to recommend is to take the procedure mentioned in the body of the paper, but "shifted" as described above.

9.2 Numerical Comparison with and without LG terms

The "twist" term in LG processes may at first glance be strange. On the other hand, mathematically it is close to a simple AR(1). It may be useful to illustrate, in the same simulations, how an AR(1) and a LG process behave. This section compares twisted and non-twisted processes in a stock market model with a time-varying risk premium.

Time is discrete, with increments Δt. The stochastic discount factor and the dividend

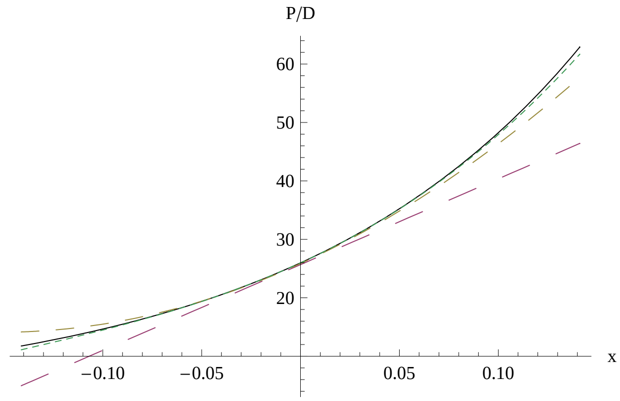

Figure 4: The Figure plots the true value of the P/D ratio of a stock with an Ornstein-Uhlenbeck process (solid line), and the approximation by a LG process with m =1, 2 and 3 factors (wide, medium and thin dashes, respectively), using the shifted basic truncation. Note that 95% of the time, the factor x is in the range $[-6.9\%, 6.9\%]$. The average relative error between the OU and the LG approximation is $E\left[\left|V^{[m]\prime}(x) - V(x)\right|\right]/E[V(x)]$, and is equal to 3.6%, 0.73% and 0.0075% for $m = 1, 2, 3$ respectively.

process follow

$$\frac{M_{t+\Delta t}}{M_t} = e^{-r\Delta t}\left(1 + \eta^M_{t+\Delta t}\right), \qquad \frac{D_{t+\Delta t}}{D_t} = e^{g\Delta t}\left(1 + \eta^D_{t+\Delta t}\right)$$

where $\eta^M_{t+\Delta t}$ and $\eta^D_{t+\Delta t}$ have zero expected value conditional at t.

We model $E_t\left[\frac{M_{t+\Delta t}}{M_t}\frac{D_{t+\Delta t}}{D_t}\right]$ as a function of a factor x_t, such that the risk premium is $\pi_* - x_t$ to a first order. We consider two processes x_t, an AR(1) and a LG process. The AR(1) formulation is: $E_t\left[\eta^M_{t+\Delta t}\eta^D_{t+\Delta t}\right] = e^{-\pi_*\Delta t}e^{x_t\Delta t}$, so that risk premium on a one-period ahead dividend claim is $\pi_* - x_t$ to the leading order. For the perturbation x_t, we can postulate the AR(1):

$$x_{t+\Delta t} = F(x) + \sigma(x_t)\sqrt{\Delta t}\varepsilon_{t+\Delta t}, \qquad F(x) = \overline{x} + (1 - \phi\Delta t)(x_t - \overline{x}),$$

where $\varepsilon_{t+\Delta t}$ is uniformly distributed, with mean 0 and variance 1. The constant $\overline{x}$ term is an convexity adjustment, that will be determined soon.

We also consider a LG process. The covariance is expressed as: $E_t\left[\eta^M_{t+\Delta t}\eta^D_{t+\Delta t}\right] =$

$e^{-\pi_* \Delta t}\left(1 + x_t^* \Delta t\right)$, where x_t^* follows:

$$x_{t+\Delta t}^* = F^*\left(x_t^*\right) + \sigma\left(x_t^*\right)\sqrt{\Delta t}\varepsilon_{t+\Delta t}, \qquad F^*\left(x^*\right) = \frac{\left(1 - \phi\Delta t\right) x^*}{1 + x^* \Delta t}. \tag{78}$$

So, the risk premium on a dividend claim is, to the leading order, $\pi_* - x_t^*$. The processes mean-revert with a speed ϕ. The starred variables relate to the LG process, while the non-starred variables relate to the AR(1).

To ensure that the process remains within $[X_{\min}, X_{\max}]$ with $X_{\min} < 0 < X_{\max}$, we define the baseline volatility as:

$$\overline{\sigma}\left(x\right) = \sqrt{2K}\left(1 - x/X_{\min}\right)\left(1 - x/X_{\max}\right), \tag{79}$$

with $K > 0$. It goes to 0 fast enough at $X_{\min} < 0$ and $X_{\max} > 0$, ensuring that x_t is within $[X_{\min}, X_{\max}]$. The average volatility of X is fairly well approximated by: $K^{1/2}\xi$, with $\xi \simeq 1.35$. Hence, we set $v_x = K\xi^2/\left(2\phi\right)$, and $\overline{x} = -K\xi^2/\left(2\phi^2\right)$.

Furthermore, to keep x_t in $[X_{\min}, X_{\max}]$, we need ε_t to be bounded, say that $\varepsilon_t \in [\varepsilon_{\min}, \varepsilon_{\max}]$. Then, we can take the truncated volatility to be:

$$\sigma\left(x\right) = \min\left\{\overline{\sigma}\left(x\right), \frac{F\left(x\right) - X_{\min}}{\varepsilon_{\max}\sqrt{\Delta t}}, \frac{X_{\max} - F\left(x\right)}{\varepsilon_{\max}\sqrt{\Delta t}}\right\}. \tag{80}$$

The last two terms ensure that $x_{t+1} \in [X_{\min}, X_{\max}]$. They matter only when x is very close to the boundaries. Away from the boundaries, as in the continuous time limit, $\sigma\left(x\right) = \overline{\sigma}\left(x\right)$.

We shall compare the movements of x_t and x_t^*, as well as the price-dividend ratios of the two processes. In the numerical implementation, we take ε_t uniform on $[\varepsilon_{\min}, \varepsilon_{\max}] = \left[-\sqrt{3}, \sqrt{3}\right]$, so that $var\left(\varepsilon_t\right) = 1$. We use annual units, and take $R = r + \pi_* - g = 0.035$, for a central P/D ratio of 28. For the simulations, each period lasts a month, $\Delta t = 1/12$. Also, $\left(X_{\min}, X_{\max}\right) = \left(-11\%, 80\%\right)$ and $\phi = 0.13$ (which means that x_t mean-reverts with a half-life of mean-reversion of 5 years). Finally, $K = 0.18$, which corresponds to a annual volatility of the log P/D ratio of 11%, based solely on the volatility of the stochastic discount factor. With a dividend volatility of 11%, as in Campbell and Cochrane (1999), the total stock return volatility is 15%. It is easy to increase this volatility, for instance by introducing a positive correlation between innovations to dividends, and innovations to x_t.

We simulate 10,000 years of data. Figures 5-6 illustrate a typical run over 100 years.

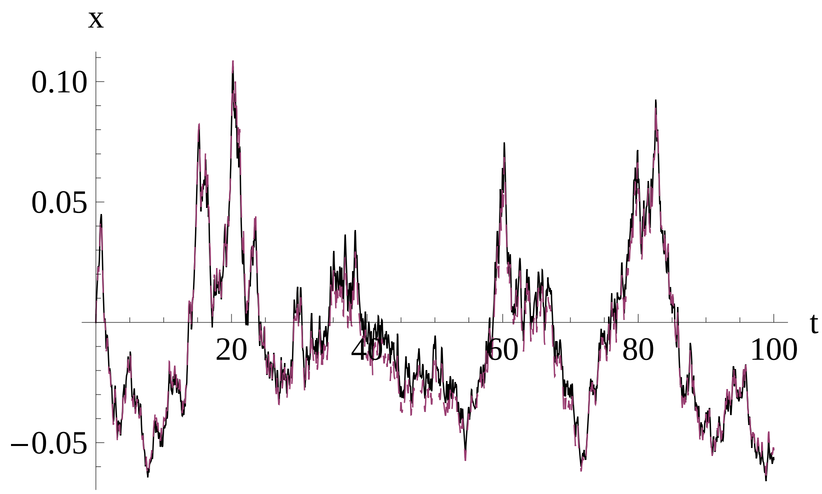

Figure 5: Processes for x_t (AR(1), dashed line) and x_t^* (LG, solid line), simulated over 100 years. The curves are quite close, which was expected as the two processes are identical up to second order terms.

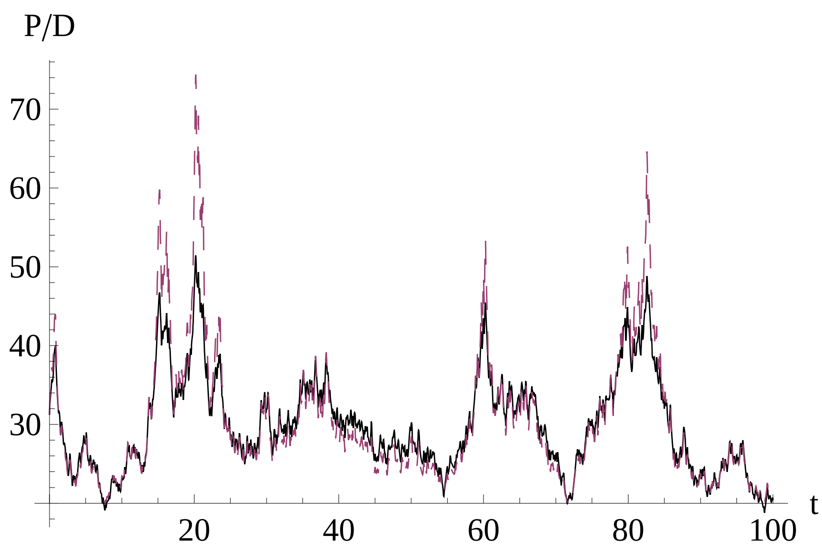

Figure 6: The price-dividend ratios, simulated over 100 years. The solid black line represents the P/D associated with the LG process, and dashed purple line the P/D ratio associated with the AR(1) process.

They show x_t and x_t^*, the corresponding price/dividend ratios P_t/D_t and P_t^*/D_t^*. (For the LG process, the PD ratio is in closed form; for the AR(1) process, a numerical solution is used). Given that the processes for x_t and x_t^* are similar up to second order terms, they are close, as confirmed by Figure 5.

The correlations are $corr\,(x_t, x_t^*) = 0.985$ and $corr\,(\ln P_t/D_t, \ln P_t^*/D_t^*) = 0.958$. The correlation in the returns generated by the two models is 0.97.

The conclusion is that the processes are indeed quite close. Of course, even if they had been quite different, this would not have been a problem for LG processes. We do not want to say that the true model is an AR(1), that a LG process approximates. It could as well be that the true model is a LG process, than an AR(1) model approximates. Or rather, as models are just approximation of a complex economic reality, the respective advantage of LG vs affine models depends on the specific task at hand. The modeler should be able to pick whichever approximation is most convenient. It is simply reassuring that the modelling choice does not make a large difference in terms of the economic processes.

9.3 Example of a Model Projected in the LG Space: Bansal-Yaron (2004)

A side benefit from LG processes is that they allow to linearize models, by projecting them in an "LG space". To illustrate this, we present here the analysis of the Bansal-Yaron (BY, 2004) model, solved (approximately, like BY) with the LG method. It yields, of course, the same expressions (to a first order) as the original BY framework, but with a different method. Depending on the reader's taste, this method may or may not be easier to use than the more traditional "linearize, plug, and verify" method. Hence, this subsection is mostly of illustrative rather than substantive interest.

The essentials of the BY model are as follows. The pricing kernel comes from an Epstein-Zin-Weil preference:

$$\frac{M_{t+1}}{M_t} = \delta^\theta G_{t+1}^{-\theta/\psi} R_{a,t+1}^{\theta-1}$$

where G_{t+1} is the growth rate of consumption, $R_{a,t+1}$ is the gross return on an asset that delivers aggregate consumption, δ the subjective discount factor, $\theta \equiv \frac{1-\gamma}{1-1/\psi}$, where γ is risk aversion and ψ the intertemporal elasticity of substitution. The processes are:

$$\begin{aligned}
\ln G_{t+1} &= \mu + x_t + \sigma_t \eta_{t+1} \\
x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1} \\
\sigma^2_{t+1} &= \sigma^2 + \nu_1 \left(\sigma^2_t - \sigma^2\right) + \sigma_w w_{t+1}
\end{aligned}$$

where $e_{t+1}, u_{t+1}, w_{t+1}$ i.i.d. standard Gaussian, so that x_t is the expected growth rate of consumption, and σ_t its stochastic volatility.

We first derive the basic "case I" of the BY model, with constant volatility $\sigma^2_t \equiv \sigma^2$. The market value of the consumption claim will be (in the linearization)

$$P_t = \frac{C_t}{R}\left(1 + b x_t\right), \tag{81}$$

for some constants b and R to be determined. The process x_t follows $E_t x_{t+\Delta t} = \left(1 - \Phi_x \Delta t\right) x_t$, where the BY notation is $\rho = 1 - \Phi_x \Delta t$. We will consider the limit of small time intervals, $E_t\left[dx_t\right] = -\Phi_x x_t dt$, and, given the stochastic structure postulated by BY, the expected return on the consumption claim is:

$$\begin{aligned}
E\left[r^a_{t+dt}\right]/dt &= E\left[\frac{d\left(P_t/C_t\right)}{P_t/C_t} + \frac{dC_t}{C_t} + \frac{C_t}{P_t}dt\right]/dt = \frac{bE\left[dx_t\right]/dt}{1 + bx_t} + \left(\mu + x_t\right) + \frac{R}{1 + bx_t} \\
&= R + \mu + \left(1 - \left(R + \Phi_x\right) b\right) x_t + O\left(x_t^2\right)
\end{aligned}$$

This linearization is the moral equivalent of the Campbell-Shiller linearization of the return used by BY.

Instead of using the Euler equation, we calculate the LG moments. To keep the derivation elementary, we proceed in discrete time, with small intervals dt, then take the limit $dt \to 0$:

$$E\left[\frac{M_{t+dt}C_{t+dt}}{M_t C_t}\right] = E\left[\delta^\theta G_{t+dt}^{1-\theta/\psi} e^{(\theta-1)r^a_{t+dt}}\right] = \exp\left(-R' dt + b' x_t dt\right)$$

for R' a constant, and

$$b' = \left(1 - \frac{\theta}{\psi}\right) + \left(\theta - 1\right)\left(1 - \left(R + \Phi_x\right) b\right) = \theta\left(1 - \frac{1}{\psi}\right) - \left(\theta - 1\right)\left(R + \Phi_x\right) b$$

We summarize the first LG moment, $E\left[\frac{d(M_tC_t)}{M_tC_t}\right]/dt = -R' + b'x_t$. Also,

$$E\left[\frac{d(M_tC_t)}{M_tC_t}x_t\right]/dt = (-R' + b'x_t)\,x_t - \Phi_x x_t \simeq -(R' + \Phi_x)\,x_t$$

So, $M_tC_t(1, x_t)$ is (approximately) a LG process, with generator $\omega = \begin{pmatrix} R' & -b' \\ 0 & R+\Phi_x \end{pmatrix}$. Hence, by Theorem 4, the price of a consumption claim, $P_t = E_t\left[\int_{s=t}^{\infty} M_sC_s/M_t ds\right]$, is $P_t = \frac{C_t}{R'}\left(1 + \frac{b'x_t}{R+\Phi_x}\right)$. Equating this with (81), we have: $R' = R$, and

$$b = \frac{b'}{R+\Phi_x} = \frac{\theta\left(1-\frac{1}{\psi}\right) - (\theta-1)(R+\Phi_x)\,b}{R+\Phi_x}$$

We deduct $b = \frac{1-\frac{1}{\psi}}{R+\Phi_x}$. Hence, the generator of $M_tC_t(1, x_t)$ is $\omega = \begin{pmatrix} R & -\left(1-\frac{1}{\psi}\right) \\ 0 & R+\Phi_x \end{pmatrix}$.

We conclude that the price of a consumption claim is

$$\frac{P_t}{C_t} = \frac{1}{R}\left(1 + \frac{1-\frac{1}{\psi}}{R+\Phi_x}x_t\right) \tag{82}$$

Note that, as BY, we have not solved for R, but this could easily be done.

We next verify that that the LG expression (82) matches the BY approximate expression,

$$\left(\frac{P_t}{C_t}\right)^{BY} \simeq e^{A_0 + A_1 x_t \Delta t} \simeq e^{A_0}\left(1 + A_1 x_t \Delta t\right)$$

Identifying $e^{A_0} = 1/R$. BY find

$$A_1\Delta t = \frac{1-\frac{1}{\psi}}{1-\rho_x\kappa}\Delta t = \frac{1-\frac{1}{\psi}}{1-(1-\Phi_x\Delta t)(1-R\Delta t)}\Delta t = \frac{1-\frac{1}{\psi}}{\Phi_x + R}$$

as $\Delta t \to 0$. So the BY expression is identical to (82), up to second order terms.

In conclusion, as expected the LG way of solving the BY problem, and the first order approximation of BY yield the same expressions, to the first order. The same finding would hold for the other expressions in their paper.

The LG method allows easily to calculate finite-maturity claims, something which is not done in the original BY paper, but is immediate to do after the LG moments have been calculated. For instance, the price $P_t(T)$ of a consumption dividend paying at $t+T$ is:

$$P_t(T) = C_t e^{-RT}\left(1+\frac{1-e^{-\Phi_x T}}{\Phi_x}\left(1-\frac{1}{\psi}\right)x_t\right). \tag{83}$$

This could be useful, given the interest for dividend strips in the value / growth literature (e.g., Lustig, van Nieuwerburgh and Verdelhan (2008)).

Derivation of bond values.

We can also derive the bond values in the BY model. Using the same linearization, we get: $E\left[\frac{dM_t}{M_t}\right]/dt = -r-\frac{1}{\psi}x_t$, for a constant r. Likewise, $E\left[\frac{d(M_t x_t)}{M_t}\right]/dt = -(r+\Phi_x)x_t$. So $M_t(1,x_t)$ is (approximately) a LG process, with generator $\omega = \begin{pmatrix} r & \frac{1}{\psi} \\ 0 & r+\Phi_x \end{pmatrix}$. Hence the price of a real bond, $Z_t(T) = E_t[M_{t+T}/M_t]$ is

$$Z_t(T) = e^{-rT}\left(1-\frac{1-e^{-\Phi_x T}}{\Phi_x}\frac{x_t}{\psi}\right). \tag{84}$$

and the linearized bond yield is:

$$y_t(T) = r+\frac{1-e^{-\Phi_x T}}{\Phi_x T}\frac{x_t}{\psi}. \tag{85}$$

If $\psi>1$, a high long run growth rate x_t increases both stock prices and bond yields. Bansal and Shaliastotich ("A Long-Run Risks Explanation of Predictability Puzzles in Bond and Currency Markets", 2009) present similar expressions.

Incorporating stochastic volatility

With the same method, one can show that $M_t C_t\left(1, x_t, \theta\left(\sigma_t^2-\sigma^2\right)\right)$ is a LG process with generator:

$$\omega = \begin{pmatrix} R & -\left(1-\frac{1}{\psi}\right) & -b_\sigma \\ 0 & R+\Phi_x & 0 \\ 0 & 0 & R+\Phi_\sigma \end{pmatrix}, b_\sigma = \frac{\theta}{2}\left(\left(1-\frac{1}{\psi}\right)^2+\left(\frac{1-\frac{1}{\psi}}{R+\Phi_x}\varphi_e\right)^2\right),$$

where with the BY notations the autocorrelation of σ_t^2 is $\nu_1 = 1 - \Phi_\sigma \Delta t$. This yields the price/dividend ratio:

$$\frac{P_t}{C_t} = \frac{1}{R}\left[1 + \frac{\left(1 - \frac{1}{\psi}\right)}{R + \Phi_x} x_t + b_\sigma \frac{\sigma_t^2 - \sigma^2}{R + \Phi_\sigma}\right],$$

which is the LG analogue of BY's expression. The price of a consumption strip is:

$$P_t(T) = C_t e^{-RT}\left(1 + \frac{1 - e^{-\Phi_x T}}{\Phi_x}\left(1 - \frac{1}{\psi}\right) x_t + \frac{1 - e^{-\Phi_\sigma T}}{\Phi_\sigma} b_\sigma \left(\sigma_t^2 - \sigma^2\right)\right). \tag{86}$$

10 References

Canuto, C., M. Y. Hussaini, A. Quarteroni and T.A. Zhang (2006): *Spectral Methods: Fundamentals in Single Domains.* Springer, New York.

Gardiner, C.W. (2003): *Handbook of Stochastic Methods.* Springer, New York.

Kristensen, D, and A. Mele (2009): "Adding and Substracting Black-Scholes: A New Approach to Approximating Derivative Prices in Continuous Time Models," Working Paper, Columbia U.

Johnson, T. C. (2002): "Rational Momentum Effects," *The Journal of Finance*, 57, 585-608.

Lustig, H., S. van Nieuwerburgh and A. Verdelhan (2008): "The Wealth-Consumption Ratio," Working Paper, UCLA.

Pastor, L., and P. Veronesi (2003): "Stock Valuation and Learning about Profitability," *Journal of Finance*, 58, 1749-1789.