Appendix: Investor Information, Long-Run Risk, and the Term Structure of Equity.

Mariano M. Croce*
UNC Chapel Hill

Martin Lettau†
UC Berkeley, CEPR and NBER

Sydney C. Ludvigson‡
NYU and NBER

First draft: August 15, 2005
This draft: June 6, 2014

*Department of Finance, Kenan-Flagler Business School, CB #3490, Chapel Hill, NC, 27599-3490; Email: crocem@kenan-flagler.unc.edu; Tel: (919)962-3179; http://public.kenan-flagler.unc.edu/faculty/crocem/

†Department of Finance, Haas School of Business, UC Berkeley, 545 Student Services Bldg., #1900, <Berkeley, CA, 94720-1900; Email: lettau@haas.berkeley.edu; Tel: (510) 643-6349; http://faculty.haas.berkeley.edu/lettau/

‡Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003; Email: sydney.ludvigson@nyu.edu; Tel: (212) 998-8927; Fax: (212) 995-4186; http://www.econ.nyu.edu/user/ludvigsons/

This material is based upon work supported by the National Science Foundation under Grant No. 0617858 to Lettau and Ludvigson. Ludvigson also acknowledges financial support from the Alfred P. Sloan Foundation and the CV Starr Center at NYU. The authors thank Dave Backus, John Y. Campbell, Timothy Cogley, Michael Gallmeyer, Lars Hansen, John Heaton, Thomas Sargent, Jay Shanken, Stijn Van Nieuwerburgh and seminar participants at the 2006 Society for Economic Dynamics conference, the summer 2006 NBER Asset Pricing meeting, Emory, NYU, Texas A&M, UCLA, and UNC Chapel Hill for helpful comments. Any errors or omissions are the responsibility of the authors.

Appendix to "Investor Information, Long-Run Risk, and the Term Structure of Equity"

This document is an on-line Appendix for the paper entitled "Investor Information, Long-Run Risk, and the Term Structure of Equity." The Appendix discusses (i) finite sample properties of VARMA versus ARMA representations for consumption and dividend growth, (ii) identification of parameters of the true data generating process, (iii) the innovations representation of the limited information problem, (iv) implications for the term structure of equity under full information when the true data generating process is a pair of ARMA(1,1) processes, (v) the numerical solution procedure (vi) estimates of VARMA and ARMA cash-flow processes using historical data.

In this document we refer to the true data generating process given in the text as

$$\Delta c_{t+1} = \mu_c + \underbrace{x_{c,t}}_{\text{LR risk}} + \sigma \underbrace{\varepsilon_{c,t+1}}_{\text{SR risk}} \tag{1}$$

$$\Delta d_{t+1} = \mu_d + \phi_x x_{c,t} + \phi_c \sigma \varepsilon_{c,t+1} + \sigma_d \sigma \varepsilon_{d,t+1} \tag{2}$$

$$x_{c,t} = \rho x_{c,t-1} + \sigma_{xc} \sigma \varepsilon_{xc,t} \tag{3}$$

$$\varepsilon_{c,t+1}, \varepsilon_{d,t+1}, \varepsilon_{xc,t} \sim N.i.i.d\,(0,1)\,. \tag{4}$$

1 VARMA versus ARMA Representations

The dynamic system has a multivariate Wold representation given as a first-order vector autoregressive-moving average representation ($VARMA\,(1,1)$) taking the form,

$$\begin{bmatrix} \Delta c_{t+1} \\ \Delta d_{t+1} \end{bmatrix} = \begin{bmatrix} \mu_c\,(1-\rho) \\ \mu_d\,(1-\rho) \end{bmatrix} + \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \Delta c_t \\ \Delta d_t \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v^V_{c,t+1} \\ v^V_{d,t+1} \end{bmatrix} - \underbrace{\begin{bmatrix} b_{cc} & b_{cd} \\ b_{dc} & b_{dd} \end{bmatrix}}_{\mathbf{b}} \begin{bmatrix} v^V_{c,t} \\ v^V_{d,t} \end{bmatrix}. \tag{5}$$

We assume that agents instead estimate the best fitting *univariate* Wold representations for consumption and dividend growth. Given the true data generating process, univariate Wold representations for Δc_t, and Δd_t can be written as a pair of first-order autoregressive-moving average ($ARMA(1,1)$) processes:

$$\Delta c_{t+1} = \mu_c\,(1-\rho) + \rho \Delta c_t + v^A_{c,t+1} - b_c v^A_{c,t} \tag{6}$$

$$\Delta d_{t+1} = \mu_d\,(1-\rho) + \rho \Delta d_t + v^A_{d,t+1} - b_d v^A_{d,t}. \tag{7}$$

To understand why we assume agents with limited information estimate the best fitting univariate processes rather than the multivariate counterpart, we need to discuss the tradeoff between forecast bias and forecast variance. Denote the predicted values of consumption and dividend growth formed using the univariate Wold representations as $\Delta\widehat{c}_{t+1}^{ARMA}$, and $\Delta\widehat{d}_{t+1}^{ARMA}$, respectively. Likewise, denote the predicted values of consumption and dividend growth formed using the multivariate Wold process $\Delta\widehat{c}_{t+1}^{VARMA}$, and $\Delta\widehat{d}_{t+1}^{VARMA}$, respectively. The multivariate Wold representation for the dynamic system (1)-(3) is a $VARMA\,(1,1)$, which has 11 free parameters. By contrast, each univariate process has four free parameters. Given that the true data generating process has a $VARMA$ representation, $\Delta\widehat{c}_{t+1}^{ARMA}$, and $\Delta\widehat{d}_{t+1}^{ARMA}$ will be biased predictors of consumption and dividend growth, while $\Delta\widehat{c}_{t+1}^{VARMA}$, and $\Delta\widehat{d}_{t+1}^{VARMA}$ will be unbiased. However, estimation error for the 7 additional parameters of the VARMA process will increase the variance of $\Delta\widehat{c}_{t+1}^{VARMA}$, and $\Delta\widehat{d}_{t+1}^{VARMA}$ over that of $\Delta\widehat{c}_{t+1}^{ARMA}$, and $\Delta\widehat{d}_{t+1}^{ARMA}$. As a consequence, agents face a choice between the unbiased but noisy predictors $\Delta\widehat{c}_{t+1}^{VARMA}$ and $\Delta\widehat{d}_{t+1}^{VARMA}$ and the biased but less-noisy predictors $\Delta\widehat{c}_{t+1}^{ARMA}$, and $\Delta\widehat{d}_{t+1}^{ARMA}$. If the sample size and off-diagonal of the $\mathbf{b}$ matrix in (5) are sufficiently small, the noise from estimating the additional $VARMA$ parameters can out-weigh the gains from eliminating the bias, implying that better forecasts are obtained using the more parsimonious univariate $ARMA$ specifications. This occurs because the off-diagonal elements of the b matrix, b_{cd} and b_{dc}, are close to zero, so the system behaves approximately as a pair of $ARMA\,(1,1)$ processes.

Indeed, we find that this is the situation here, both in the model and in the data. Under the true data-generating process (1)-(3), the parameters b_{cd} and b_{dc} are small in absolute value in the benchmark calibration discussed below, equal to 0.004 and -0.08, respectively. These parameters are also small when estimated using historical data, where the restriction $b_{cd} = b_{dc} = 0$ cannot be statistically rejected at the five percent level in monthly, quarterly or annual data. This suggests that the $VARMA$ system is well approximated by two univariate $ARMA$ processes.

A more direct way to study the bias-variance tradeoff is to compare forecasts explicitly. Table A1 shows the results of short-sample simulations of the data generating process (1)-(4). We simulate 500 samples of the size currently available in historical data and use these samples to generate a set of recursive, out-of-sample forecasts of consumption and dividend growth. The data generating process is calibrated using the benchmark calibration discussed below. The table reports the difference, in percent, between the root mean squared

forecast error (RMSE) of the $ARMA$ estimation and the $VARMA$ estimation, averaged across the 500 samples. Thus, a negative number indicates that the $ARMA$ estimation has lower forecast error. As Table A1 shows, the estimated $ARMA$ processes routinely produce forecasts that are superior to those of the estimated $VARMA$ process. This result is robust across a range of initial estimation periods and forecast horizons. Thus, even when the true data-generating process is given by the dynamic system (1)-(4), the best-fitting univariate processes produce more accurate forecasts of consumption and dividend growth than does the true multivariate Wold representation.

Table A2 shows that a similar result holds qualitatively in historical data. We obtain annual, quarterly and monthly data on consumption and dividends and estimate $ARMA(1,1)$ and $VARMA(1,1)$ processes. As in Table A1, the table reports the difference in $RMSE$ between the $ARMA$ forecast and the $VARMA$ forecast. Over a range of different data frequencies, forecast horizons, and initial estimation samples, the $ARMA$ forecasts are more often than not superior to the $VARMA$ forecasts, sometimes sizeably so. For example, in quarterly data, the $ARMA$ consumption growth forecast error is almost 8 percent smaller than the counterpart based on the $VARMA$ specification. Unlike the $ARMA$ models, there is no case in which the $VARMA$ model produces a superior forecast for both consumption and dividend growth simultaneously.

These results show that, in samples of the size currently available, the additional parameters of the $VARMA$ specification cannot be estimated accurately enough to improve forecasts over a best-fitting univariate model. Indeed, the opposite holds: attempting to estimate the larger system worsens forecasts. Thus, even if the dynamic system (1)-(4) generates the data, it would be rational for agents faced with incomplete information to use the more parsimonious univariate specifications for forecasting consumption and dividend growth, when faced with the signal extraction problem of uncovering long-run consumption risk from observable data.

2 Identification

We assume that the general data generating process that agents with limited information would like to estimate takes the form

$$\Delta c_{t+1} = \mu_c + \underbrace{x_{c,t}}_{\text{LR risk}} + \sigma \underbrace{\varepsilon_{c,t+1}}_{\text{SR risk}} \tag{8}$$

$$\Delta d_{t+1} = \mu_d + x_{d,t} + \phi_c \sigma \varepsilon_{c,t+1} + \sigma_d \sigma \varepsilon_{d,t+1} \tag{9}$$

$$x_{c,t} = \rho x_{c,t-1} + \sigma_{xc} \sigma \varepsilon_{xc,t} \tag{10}$$

$$x_{d,t} = \rho_d x_{d,t-1} + \sigma_{xd} \sigma \varepsilon_{xd,t} \tag{11}$$

$$(\varepsilon_{c,t+1}, \varepsilon_{d,t+1}, \varepsilon_{xc,t}, \varepsilon_{xd,t}) \sim N.i.i.d\,(\mathbf{0}, \mathbf{\Omega}) \tag{12}$$

Market participants could in principal obtain a consistent estimate of these parameters simultaneously with estimates of $x_{c,t}$ and $x_{d,t}$, by writing the dynamic system above in state space form and applying maximum likelihood to the history of consumption and dividend data. Agents could use the Kalman filter to form an estimate of the unobservable conditional means $x_{c,t}$ and $x_{d,t}$, by sequentially updating a linear projection on the basis of consumption and dividend data observed through date t.

This system is not identified, however. The system (8)-(12) has 14 unknown parameters (including ten unknown parameters in $\mathbf{\Omega}$). Estimation of (5) identifies 11 parameters, three short of what's needed for exact identification. That is, given a sufficiently long sample of data on consumption and dividend growth, the parameters of the dynamic system (8)-(12) can be observed in certain combinations as the estimates ρ, b_{cc},..., b_{dd} and the variance-covariance matrix of $v_{c,t+1}$ and $v_{d,t+1}$, but this information is not enough to separately identify the parameters of (8)-(12). The true data generating process (1)-(3) is a special case of this system that imposes the restrictions $x_{d,t} = \phi_x x_{c,t}$, requiring $\rho_d = \rho$, $\sigma_{xd} = \phi_x \sigma_{xc}$, $\varepsilon_{xd,t} = \varepsilon_{xc,t}$, $x_0 = x_{d0} = 0$, and the shocks to (8)-(10) to be uncorrelated.

3 Innovations Representation

The $ARMA(1,1)$ processes may be recast in terms of the following pair of innovations representations:

$$\Delta c_{t+1} = \mu_c + \widehat{x}_{c,t} + v^A_{c,t+1} \tag{13}$$

$$\widehat{x}_{c,t+1} = \rho \widehat{x}_{c,t} + K v^A_{c,t+1} \tag{14}$$

$$\Delta d_{t+1} = \mu_d + \widehat{x}_{d,t} + v^A_{d,t+1} \tag{15}$$

$$\widehat{x}_{d,t+1} = \rho \widehat{x}_{d,t} + K^d v^A_{d,t+1}, \tag{16}$$

where $K \equiv \rho - b_c$ and $K^d \equiv \rho - b_d$. Here, $\widehat{x}_{c,t}$ and $\widehat{x}_{d,t}$ denote optimal linear forecasts based on the history of consumption and dividend data separately, i.e., $\widehat{x}_{c,t} \equiv \widehat{E}\left(x_{c,t}|\mathbf{z}_c^t\right)$, and $\widehat{x}_{d,t} \equiv \widehat{E}\left(x_{d,t}|\mathbf{z}_d^t\right)$, where $\mathbf{z}_c^t \equiv (\Delta c_t, \Delta c_{t-1}, ..., \Delta c_1)'$ and $\mathbf{z}_d^t \equiv (\Delta d_t, \Delta d_{t-1}, ..., \Delta d_1)'$.

The optimal forecasts are functions of the observable ARMA parameters and innovations:

$$\begin{aligned} \widehat{x}_{c,t} &= -\rho\mu_c + \rho\Delta c_t - b_c v_{c,t}^A \\ \widehat{x}_{d,t} &= -\rho\mu_d + \rho\Delta d_t - b_d v_{d,t}^A. \end{aligned}$$

4 ARMA(1,1) and Insurance–Full Information

Suppose the decision maker has full information and the true data generating process follows a pair of ARMA(1,1)s:

$$\Delta c_{t+1} = \mu(1-\rho) + \rho\Delta c_t + v_{c,t+1}^A - b_c v_{c,t}^A, \tag{17}$$

$$\Delta d_{t+1} = \mu(1-\rho) + \rho\Delta d_t + v_{d,t+1}^A - b_d v_{d,t}^A. \tag{18}$$

The two ARMA(1,1) processes above can be rewritten in the following equivalent way:

$$\Delta c_{t+1} = \mu + x_{c,t}^A + v_{c,t+1}^A \tag{19}$$

$$\Delta d_{t+1} = \mu + x_{d,t}^A + v_{d,t+1}^A \tag{20}$$

$$x_{c,t+1}^A = \rho x_{c,t}^A + \underbrace{(\rho - b_c)}_{K_c^A} v_{c,t+1}^A \tag{21}$$

$$x_{d,t+1}^A = \rho x_{d,t}^A + \underbrace{(\rho - b_d)}_{K_d^A} v_{d,t+1}^A \tag{22}$$

$$\left(v_{c,t+1}^A, v_{d,t+1}^A\right) \sim N.i.i.d\left(\mathbf{0}, \mathbf{\Omega_A}\right), \tag{23}$$

where

$$\Omega_A = \begin{bmatrix} \sigma_{v_c,A}^2 & \sigma_{v_c,A}\sigma_{v_d,A}\rho_{v_c v_d,A} \\ \sigma_{v_c,A}\sigma_{v_d,A}\rho_{v_c v_d,A} & \sigma_{v_d,A}^2 \end{bmatrix}.$$

In this setting, unlike in the model of the text, the shocks to the short-run (unforecastable) and long-run (forecastable) components are perfectly correlated. The SDF is driven just by one risk factor, i.e., $v_{c,t+1}^A$:

$$m_{t+1} - E_t[m_{t+1}] = -\underbrace{\left[\gamma + \kappa_c \frac{\gamma - 1/\psi}{1 - \rho\kappa_c} K_c^A\right]}_{>0} v_{c,t+1}^A. \tag{24}$$

With a single shock in each univariate process, how do we separate out components of the market risk premium attributable to long- versus short-run movements in consumption/dividend growth? This can be accomplished using the approximate return decomposition of (Campbell (1991)), which shows that the unexpected log stock return is associated with changes in expectations of future dividends or expectations of future real returns:

$$r_{t+1} - E_t\left[r_{t+1}\right] = \left(E_{t+1} - E_t\right)\sum_{j=0}^{\infty}\varphi^j \Delta d_{t+1+j} - \left(E_{t+1} - E_t\right)\sum_{j=1}^{\infty}\varphi^j \Delta r_{t+1+j},$$

where $\varphi \equiv 1/\left(1+\exp\left(\overline{d-p}\right)\right)$, and $\overline{d-p}$ is the mean log dividend-price ratio. The first sum begins at $j=0$ while the second begins at $j=1$, implying that return surprises may be written as the sum of three terms:

$$r_{t+1} - E_t\left[r_{t+1}\right] = \underbrace{\Delta d_{t+1} - E_t \Delta d_{t+1}}_{\text{cash-flow surprise}} + \underbrace{\left(E_{t+1} - E_t\right)\sum_{j=1}^{\infty}\varphi^j \Delta d_{t+1+j}}_{\text{news about future cash-flows}} - \underbrace{\left(E_{t+1} - E_t\right)\sum_{j=1}^{\infty}\varphi^j \Delta r_{t+1+j}}_{\text{news about future returns}}.$$

The first term above is today's dividend shock (short-run cash flow surprise); it is non-zero even when dividend growth is i.i.d. The second term above represents revisions in expectations, or "news," of future dividend growth rates. The third term above represents news about future returns. The last two terms are non-zero only if dividend growth and returns are *not* i.i.d., that is only if there is persistence in these variables. In the models of this paper, the last two terms are non-zero only if consumption growth has some persistence: persistence in consumption growth directly affects cash-flow news in the middle term on the right-hand-side, and indirectly affects return news through the risk-free rate when the EIS is non-zero. Thus we can decompose the market risk premium into components that would be present even if consumption growth were i.i.d. (movements attributable to the current cash-flow surprise) and to components present only if there is persistence in consumption growth (movements attributable to news about future growth rates and returns). For the ARMA(1,1) case considered in this subsection, the market return innovation (surprise) may be decomposed into components attributable to the cash-flow surprise and news variation:

$$r^{ex}_{m,t+1} - E_t[r^{ex}_{m,t+1}] = \underbrace{\upsilon^A_{d,t+1}}_{\text{cash-flow surprise}} \tag{25}$$

$$+ \underbrace{\kappa_d \frac{1}{1-\rho\kappa_d} K^A_c \left(\frac{K^A_d}{K^A_c}\upsilon^A_{d,t+1} - \frac{1}{\psi}\upsilon^A_{c,t+1}\right)}_{\text{news variation}}. \tag{26}$$

We can use this decomposition to compute the component of the market risk premium that is attributable to the current cash-flow surprise and news components, by computing the covariance of each component (25) and (26) with the SDF. The covariance of m_{t+1} with (25) gives component of the risk premium driven by exposure to purely short-run shocks, while the covariance of m_{t+1} with (26) gives the component of the risk premium driven by exposure to long-run consumption risk that drives the news components.

The market risk premium is $E[r^{ex}_{m,t+1}] = E[E_t[r^{ex}_{m,t+1}]] = E[-cov_t(r^{ex}_{m,t+1}, m_{t+1})]$. Since $E[r^{ex}_{m,t+1}] \approx -cov_t(r^{ex}_{m,t+1}, m_{t+1})$, we can write this covariance as

$$E[r^{ex}_{m,t+1}] \approx \underbrace{\left[\frac{\gamma - 1/\Psi}{1-\rho\kappa_c}\kappa_c K^A_c + \gamma\right]\rho_{v_c,v_d,A}\sigma_{v_c,A}\sigma_{v_{d,A}}}_{E[r^{ex}_{1,t+1}]=-cov_t(v^A_{d,t+1},m_{t+1})} \tag{27}$$

$$+ \underbrace{\frac{\kappa_d}{1-\rho\kappa_d}(\rho_{v_c,v_d,A}\frac{K^A_d\sigma_{v_d,A}}{K^A_c\sigma_{v_c,A}} - \frac{1}{\psi})\left(\frac{\gamma-1/\Psi}{1-\rho\kappa_c}\kappa_c + \frac{\gamma}{K^A_c}\right)(K^A_c\sigma_{v_c,A})^2}_{\kappa_d\frac{1-\rho}{1-\rho\kappa_d}S=-\frac{\kappa_d}{1-\rho\kappa_d}cov_t(\frac{K^A_d}{K^A_c}v^A_{d,t+1}-\frac{1}{\psi}v^A_{c,t+1},m_{t+1})}. \tag{28}$$

The term (27) is the component of the risk premium attributable to the current cash-flow surprise; note that the correlation $\rho_{v_c,v_d,A}$ must be non-zero in order for the dividend shock to be priced. The term (28) is the component attributable to movements in expected consumption growth; that is, attributable to persistence in consumption growth.

The term (28) varies only if expected consumption growth varies. This can be seen by focusing on the multiplicative term $(\rho_{v_c,v_d,A}\frac{K^A_d\sigma_{v_d,A}}{K^A_c\sigma_{v_c,A}} - \frac{1}{\psi})$, which, if zero, implies the entire term is zero. The first part of this term, $\rho_{v_c,v_d,A}\frac{K^A_d\sigma_{v_d,A}}{K^A_c\sigma_{v_c,A}} \equiv \phi^A$, is a measure of dividends' exposure to movements in expected consumption growth. To see this, consider the projection of the persistent component of dividend growth $x^A_{d,t}$ onto the persistent component of consumption growth $x^A_{c,t}$:

$$x^A_{d,t} = \beta_{x^A_d|x^A_c} \cdot x^A_{c,t} + resid_t,$$

where $resid_t$ is a residual and $\beta_{x^A_d|x^A_c}$ is the coefficient in this projection. For the full information model (17)-(18), $\beta_{x^A_d|x^A_c} = \frac{COV(x^A_d,x^A_c)}{V(x^A_c)}$. But since x^A_d and x^A_c are two AR(1) processes with the same persistence, this collapses to:

$$\beta_{x^A_d|x^A_c} = \frac{\mathrm{Cov}(K^A_d v^A_d, K^A_c v^A_c)}{V(K^A_c v^A_c)} = \rho_{v_c,v_d,A}\frac{K^A_d std(v^A_d)}{K^A_c std(v^A_c)} = \phi^A_x.$$

Recalling that the correlation $\rho_{v_c,v_d,A}$ must be non-zero in order for the persistent component of consumption growth to be priced into the dividend claim, it is clear that the exposure

of dividend growth to expected consumption growth is governed by $\rho_{v_c,v_d,A}\frac{K_d^A\sigma_{v_d,A}}{K_c^A\sigma_{v_c,A}}$. The exposure of dividend growth to expected returns is governed by the term $1/\psi$: the lower is ψ, the more the expected risk-free rate increases in response to an increase in expected consumption growth. Both terms are present only when expected consumption growth varies.

The decomposition above shows that short-run shocks are always a source of risk in this case. According to (27), as long as dividends are positively exposed to consumption risk (i.e., under the parameter restriction $\rho_{v_c,v_d,A} > 0$), dividends are always risky with respect to the short-run consumption shock $v_{c,t+1}^A$. This component of the risk premium is always positive.

By contrast, according to (28), dividends are risky with respect to persistent movements in consumption ($K_c^A v_{c,t+1}^A$) *if and only if* $\phi_x^A - \frac{1}{\psi} > 0$, i.e., if the revision of expected future dividends growth ($K_d^A v_{d,t+1}^A$) is big enough to overcome the insurance effect coming from the expected risk-free rate channel. In this case $S > 0$, however. When the long-run component of consumption growth is a source of risk–that is, adds to the risk premium rather than subtracts from it–the term structure slopes up. Thus this model, like the full information counterpart in the main text, implies that a downward sloping term structure is possible if and only if persistent consumption shocks are a source of *insurance* rather than risk.

Finally, note that equations (27)–(28) imply the following decomposition of the market equity premium under limited information:

$$E_t(r_{d,t+1}^{ex}) + .5V_t(r_{d,t+1}^{ex}) = E_t\left[r_{t+1}^{(1)ex} + .5V_t\left(r_{t+1}^{(1)ex}\right)\right] + \kappa_d\frac{1-\rho}{1-\kappa_d\rho}S. \tag{29}$$

This equation shows that, for any given excess return on the one-period strip, any parametrization of cash flows that delivers a downward sloping term structure $S < 0$ will make it more difficult to match evidence for a large equity premium.

5 Numerical Solution

We describe our numerical solution procedure for the full information specifications and the univariate signal extraction case. A description of the system signal extraction case is directly analogous and is omitted for brevity.

5.1 Model Solution in Full Information

Under Full Information, there is a single state variable, x_t. We discretize and bound its support by forming a grid of K points $\{x_1, x_2, \ldots x_K\}$ on the interval [-5V(x) +5V(x)]. We

choose K to be odd so that the unconditional mean of the state x is the middle point of our grid.

We discretize also the distribution of a standardized normal random variable by forming a grid of equidistant points $\{\epsilon_1, \epsilon_2, \ldots \epsilon_I\}$ over the interval [-5 +5], imposing:

$$p_i = \frac{e^{-\epsilon_i^2/2}}{\sum_1^I e^{-\epsilon_i^2/2}}, \; i = 1, 2, \ldots I$$

Again, we choose I to be odd so that $\epsilon_{(I-1)/2+1} = 0$.

Rewrite the Euler equations for the price-consumption ratio as:

$$\begin{aligned} W_c(x_k) &= \left(\sum_{i=1}^{I} \sum_{j=1}^{I} \delta^\theta e^{(1-\gamma)(\mu + x_k + \sigma\epsilon_i)} [1 + W_c(x'_{j|k})]^\theta p_i p_j \right)^{\frac{1}{\theta}} \\ x'_{j|k} &= \rho x_k + \sigma \varphi_x \epsilon_j \\ k &= 1, 2, \ldots, K, \end{aligned} \tag{30}$$

where $W_c(x_k)$ is the price-consumption ratio as a function of x in state k. The functional in (30) can be solved by noting that its right hand side is a contraction and treating $W_c(x)$ as the fixed point of $W_{c,n+1}(x) = T(W_{c,n}(x))$.

Approximate $W_{c,n}$ by a third order polynomial in x, and impose:

$$W_{c,n}(x'_{j|k}) = [1 \; x'_{j|k} \; (x'_{j|k})^2 \; (x'_{j|k})^3][\beta_{1,n} \; \beta_{2,n} \; \beta_{3,n} \; \beta_{4,n}]'$$

where the operator is initialized with an initial guess on the parameters β_0. Compute $W_{c,1}(x_k)$ for every $x_k \in \{x_1, x_2, \ldots x_K\}$, and stack the resulting values in the vector $\overrightarrow{W}_{c,1} \in R^K$. Using least squares the guesses are updated: $\beta_1 = (\Upsilon'\Upsilon)^{-1}\Upsilon'\overrightarrow{W}_{c,1}$, where:

$$\Upsilon = \begin{bmatrix} 1 & x_1 & (x_1)^2 & (x_1)^3 \\ 1 & x_2 & (x_2)^2 & (x_2)^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_k & (x_k)^2 & (x_k)^3 \\ & & & \end{bmatrix}$$

We repeat these steps until convergence (tolerance level = .1e-5).

Once $W_c(x) = [1 \; x \; x^2 \; x^3]\beta$ has been found, the stochastic discount factor has the following expression:

$$M_{k,i,j} = \delta^\theta e^{-\gamma(\mu + x_k + \sigma\epsilon_i)} \left(\frac{1 + W_c(\rho x_k + \sigma\varphi_x \epsilon_j)}{W_c(x_k)} \right)^{\theta - 1}$$

price-dividend ratios are found in a similar way by iterating until convergence the following recursion:

$$\begin{aligned} W_{d,n+1}(x_k) &= \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{l=1}^{I}\delta^{\theta}e^{-\gamma(\mu+x_k+\sigma\epsilon_{c,i})}\left(\frac{1+W_c(x'_{j|k})}{W_c(x_k)}\right)^{\theta-1}\times \\ &\quad \times[1+W_{d,n}(x'_{j|k})]e^{(\mu+\phi_x x_k+\phi_c\sigma\epsilon_{c,i}+\sigma\varphi_d\epsilon_{d,l})}p_ip_jp_l \end{aligned} \tag{31}$$

$$W_{d,n}(x'_{j|k}) = [1\ x'_{j|k}\ (x'_{j|k})^2\ (x'_{j|k})^3]\beta_{d,n}$$

The coefficients of the polynomial expansion for the price-dividends are updated by the following OLS formula: $\beta_{d,n+1} = (\Upsilon'\Upsilon)^{-1}\Upsilon'\overrightarrow{W}_{d,n+1}$.
For $n \to \infty$, $\beta_{d,n+1} \to \beta_d = (\Upsilon'\Upsilon)^{-1}\Upsilon'\overrightarrow{W}_d$.

To solve for zero coupon equity price-dividend Ratios note the following equivalence holds:

$$W_{d,t} = \sum_{n=1}^{\infty} W_{d,t}^n \tag{32}$$

where

$$\begin{aligned} W_{d,t}^0 &\equiv 1 \\ W_{d,t}^n &= E_t\left[e^{m_{t+1}+\Delta d_{t+1}}W_{d,t+1}^{n-1}\right],\ n=1,2,... \end{aligned}$$

Implement the following recursion across maturities:

$$\begin{aligned} W_d^n(x_k) &= \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{l=1}^{I}\delta^{\theta}e^{-\gamma(\mu+x_k+\sigma\epsilon_i)}\left(\frac{1+W_c(x'_{j|k})}{W_c(x_k)}\right)^{\theta-1}\times \\ &\quad \times[W_d^{n-1}(x'_{j|k})]e^{(\mu+\phi_x x_k+\phi_c\sigma\epsilon_i+\sigma\varphi_d\epsilon_l)}p_ip_jp_l \end{aligned} \tag{33}$$

where

$$\begin{aligned} k &= 1,2,...,K \\ W_d^{n-1}(x'_{j|k}) &= [1\ x'_{j|k}\ (x'_{j|k})^2\ (x'_{j|k})^3][\beta_1^{n-1}\ \beta_2^{n-1}\ \beta_3^{n-1}\ \beta_4^{n-1}]' \\ \beta^{n-1} &= (\Upsilon'\Upsilon)^{-1}\Upsilon'\overrightarrow{W}_d^{n-1}\ n=2,3,.... \\ \beta^0 &\equiv [1\ 0\ 0\ 0]' \end{aligned}$$

$$\lim_{n\to\infty}\sum_{j=1}^{n}\beta^{n-1} = \beta_d$$

This amounts to a sequence of quadrature problems that have to be solved recursively since the price of the asset with maturity n depends on the price of the asset with maturity $n-1$.

5.2 Model Solution in Limited Information

In Limited Information, the Price-Consumption Ratio and the stochastic discount factor depend just on one relevant state: $\widehat{x}$, here denoted $\widehat{\Delta c}$. We discretize and bound its support by forming a grid of K points $\{\widehat{\Delta c}_1, \widehat{\Delta c}_2, \ldots \widehat{\Delta c}_K\}$ on the interval [-5$V(\widehat{\Delta c})$ +5$V(\widehat{\Delta c})$]. We choose K to be odd so that the unconditional mean of the state $\widehat{\Delta c}$ is the middle point of our grid, $\widehat{\Delta c}_t \perp v_{c,t+1}$.
The Euler equation for the Price-Consumption ratio is:

$$\begin{aligned} W_c(\widehat{\Delta c}_k) &= \left(\sum_{j=1}^{I} \delta^\theta e^{(1-\gamma)(\mu+\widehat{\Delta c}_k+\sigma_{v_c}\epsilon_j)} [1 + W_c(\widehat{\Delta c}'_{j|k})]^\theta p_j \right)^{\frac{1}{\theta}} \\ &\text{where} \\ \widehat{\Delta c}'_{j|k} &= \rho \widehat{\Delta c}_k + (\rho - b_c)\sigma_{v_c}\epsilon_j \end{aligned} \tag{34}$$

solved by iterating until convergence the following recursion:

$$\begin{aligned} W_{c,n}(\widehat{\Delta c}_k) &= \left(\sum_{j=1}^{I} \delta^\theta e^{(1-\gamma)(\mu+\widehat{\Delta c}_k+\sigma_{v_c}\epsilon_j)} [1 + W_{c,n-1}(\widehat{\Delta c}'_{j|k})]^\theta p_j \right)^{\frac{1}{\theta}} \\ n &= 1, 2, \ldots \end{aligned}$$

where the function is interpolated by a third order polynomial in $\widehat{\Delta c}$ such that:

$$\begin{aligned} W_{c,n-1}(x'_{j|k}) &= [1 \ \widehat{\Delta c}'_{j|k} \ (\widehat{\Delta c}'_{j|k})^2 \ (\widehat{\Delta c}'_{j|k})^3][\beta_{1,n-1} \ \beta_{2,n-1} \ \beta_{3,n-1} \ \beta_{4,n-1}]' \\ \beta_n &= (\Phi'\Phi)^{-1}\Phi'\overrightarrow{W}_{c,n} \ \text{n} = 1, 2, 3, \ldots \\ &\text{where} \\ \Phi &= \begin{bmatrix} 1 & \widehat{\Delta c}_1 & (\widehat{\Delta c}_1)^2 & (\widehat{\Delta c}_1)^3 \\ 1 & \widehat{\Delta c}_2 & (\widehat{\Delta c}_2)^2 & (\widehat{\Delta c}_2)^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \widehat{\Delta c}_k & (\widehat{\Delta c}_k)^2 & (\widehat{\Delta c}_k)^3 \\ & & & \end{bmatrix} \\ \beta_0 &: \ \text{initial guess} \end{aligned}$$

The price-dividend ratio is a function of the state variable $\widehat{x}_d \equiv \widehat{\Delta d}$ and the shock v_d:

$$\begin{bmatrix} v_{c,t+1} \\ v_{d,t+1} \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{v_c} & \sigma_{v_c,c_d} \\ \sigma_{v_c,v_d} & \sigma_{v_d^2} \end{bmatrix}\right)$$

and

$$\begin{bmatrix} \widehat{\Delta c} \\ \widehat{\Delta d} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\widehat{\Delta c}} & \sigma_{\widehat{\Delta c},\widehat{\Delta d}} \\ \sigma_{\widehat{\Delta c},\widehat{\Delta d}} & \sigma^2_{\widehat{\Delta d}} \end{bmatrix}\right)$$

- A grid of combinations $(\widehat{\Delta d}_{g|k}, \widehat{\Delta c}_k)$ is stacked in a matrix S with dimension $(K \times G) \times 2$:

$$S = \begin{bmatrix} \widehat{\Delta c}_1 & \widehat{\Delta d}_{1|1} \\ \widehat{\Delta c}_1 & \widehat{\Delta d}_{2|1} \\ \vdots & \vdots \\ \widehat{\Delta c}_1 & \widehat{\Delta d}_{g|1} \\ \widehat{\Delta c}_2 & \widehat{\Delta d}_{1|2} \\ \vdots & \vdots \\ \widehat{\Delta c}_K & \widehat{\Delta d}_{g|K} \end{bmatrix}$$

The recursion used to find the price-dividend ratio is given by:

$$\begin{aligned} W_{d,n}(\widehat{\Delta c}_s, \widehat{\Delta d}_s) &= \sum_{j=1}^{I}\sum_{i=1}^{I} \delta^\theta e^{-\gamma(\mu+\widehat{\Delta c}_s+\sigma_{v_c}\epsilon_j)} \left(\frac{1+V_c(\widehat{\Delta c}'_{j|s})}{V_c(\widehat{\Delta c}_s)}\right)^{\theta-1} \times \\ &\quad \times [1+W_{d,n-1}(\widehat{\Delta c}'_{j|s}, \widehat{\Delta d}'_{i|s})] e^{\mu+\widehat{\Delta d}_s+\sigma_{v_d}\epsilon_i} p_{ij} \\ (\widehat{\Delta c}_s, \widehat{\Delta d}_s) &= [S_{s,1} S_{s,2}] \\ s &= 1, 2, ..., K \times G \end{aligned}$$

The price-dividend ratio is interpolated as above by a quadratic polynomial in the two

states:

$$
\begin{aligned}
W_{d,n-1}(\widehat{\Delta c}_s, \widehat{\Delta d}_s) &= [1\ \widehat{\Delta c}'_{j|k}\ \widehat{\Delta d}'_{i|k}\ (\widehat{\Delta c}'_{j|k})^2\ (\widehat{\Delta d}'_{i|k})^2\ \widehat{\Delta c}'_{j|k}\widehat{\Delta d}'_{i|k}] \times \\
&\quad \times [\beta^d_{1,n-1}\ \beta^d_{2,n-1}\ \beta^d_{3,n-1}\ \beta^d_{4,n-1}\ \beta^d_{5,n-1}\ \beta^d_{6,n-1}]' \\
\beta^d_n &= (\Phi^{d'}\Phi^d)^{-1}\Phi^{d'}\overrightarrow{W}_{d,n} \\
n &= 1, 2, 3, ... \\
&\text{where} \\
\Phi^d &= \begin{bmatrix} 1 & S_{1,1} & S_{1,2} & S^2_{1,1} & S^2_{1,2} & S_{1,1}S_{1,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & S_{G\times K,1} & S_{G\times K,2} & S^2_{G\times K,1} & S^2_{G\times K,2} & S_{G\times K,1}S_{G\times K,2} \end{bmatrix} \\
\beta_0 &: \text{initial guess}
\end{aligned}
$$

For zero coupon equity price-dividends, we implement the following recursion:

$$
\begin{aligned}
W^n_d(\widehat{\Delta c}_s, \widehat{\Delta d}_s) &= \sum_{j=1}^{I}\sum_{i=1}^{I} \delta^\theta e^{-\gamma(\mu+\widehat{\Delta c}_s+\sigma_{v_c}\epsilon_j)} \left(\frac{1+V_c(\widehat{\Delta c}'_{j|s})}{V_c(\widehat{\Delta c}_s)}\right)^{\theta-1} \times \qquad (35) \\
&\quad \times W^{n-1}_d(\widehat{\Delta c}'_{j|s}, \widehat{\Delta d}'_{i|s}) e^{\mu+\widehat{\Delta d}_s+\sigma_{v_d}\epsilon_i} p_{ij} \\
W^{n-1}_d(\widehat{\Delta c}'_{j|s}, \widehat{\Delta d}'_{i|s}) &= [1\ \widehat{\Delta c}'_{j|k}\ \widehat{\Delta d}'_{i|k}\ (\widehat{\Delta c}'_{j|k})^2\ (\widehat{\Delta d}'_{i|k})^2\ \widehat{\Delta c}'_{j|k}\widehat{\Delta d}'_{i|k}] \times \\
&\quad \times [\beta^{n-1}_1\ \beta^{n-1}_2\ \beta^{n-1}_3\ \beta^{n-1}_4\ \beta^{n-1}_5\ \beta^{n-1}_6]' \\
\beta^n_d &= (\Phi^{d'}\Phi^d)^{-1}\Phi^{d'}\overrightarrow{W}^n_d \\
n &= 1, 2, 3, ... \\
\beta^0_d &= [1\ 0\ 0\ 0\ 0\ 0].
\end{aligned}
$$

Table A1

ARMA vs VARMA: Short-sample Simulations

	Out of Sample RMSE					
	$I_s = 100\%$		$I_s = 75\%$		$I_s = 50\%$	
	Δc	Δd	Δc	Δd	Δc	Δd
$h = 1$	-.09	-.54	-.70	-.69	-1.1	-.54
$h = 6$	-.81	-.72	-.00	-.52	-1.2	-1.2
$h = 12$	-1.1	-1.0	-.12	-.40	-.60	-.79
$h = 24$	-1.0	-.00	-.31	-.00	-.30	-.51

Notes: The tables reports root-mean-squared errors (RMSE) of out-of-sample forecasts for univariate ARMA and multivariate VARMA models. We simulate 500 independent samples of monthly consumption and dividends growth using the calibration reported in figure 3. $N = 576$ is the total number of observations of monthly consumption and dividends growth. In each sample, we recursively estimate by maximum likelihood the parameters of the ARMA and the VARMA models and compute the h-period ahead forecast errors for consumption and dividends growth. I_s denotes the initial sample size (in percent of the total sample size N). The table reports the percentage relative difference—averaged across samples—in the RMSE of the ARMA and VARMA models for consumption and dividend growth (a negative number indicates that the ARMA model produces a superior forecast to the VARMA model).

Table A2

ARMA vs. VARMA: US Data

			Out of Sample RMSE			
Freq.	N	h	$I_s = 75\%$		$I_s = 50\%$	
			Δc	Δd	Δc	Δd
Annual	77	1	-1.4	-2.2	$-.30$	3.3
		5	$-.12$	-1.5	$-.61$	$-.90$
Quarterly	235	1	-2.4	$-.43$	-7.7	.30
		5	-5.5	$-.82$	-3.4	$-.11$
Monthly	576	1	$-.30$	$-.12$	.61	$-.50$
		5	$-.61$	$-.15$	$-.52$	$-.61$

Notes: The table reports h-period out-of-sample root-mean-squared forecast errors (RMSE) for univariate ARMA models and multivariate VARMA models using consumption and dividend data measured at different frequencies. N denotes the number of observations and I_s is the initial sample in the recursive estimation. The table reports the percentage relative difference—averaged across samples—in the RMSE of the ARMA and VARMA models for consumption and dividend growth (a negative number indicates that the ARMA model produces a superior forecast to the VARMA model).

6 Estimation of ARMA and VARMA processes

This section presents estimates of the VARMA and ARMA cash-flow processes using historical data on consumption and dividends. It should be noted that, choosing the parameters of the incomplete information process by estimating them does not allow us to compare that specification to the full information version because the latter isn't identified from estimates of either the VARMA or ARMA (see above). Thus, it should be understood that these parameters correspond to a model where the true data generating process is the ARMA or VARMA, and there is no information problem.

The parameters are estimated for two samples, with and without the Great Recession period (2009-2012). We then plug the estimated parameters into formulas for the term

spread and equity premium implied by a model where the true cash flow process is either the VARMA or ARMA process. Table A4 below reports the estimates and the implied values for the term spread of equity and the equity premium. The equity term structure spread is negative and the equity premium is sizeable in the ARMA LI model. Specifically, in the whole sample 1931-2012, both the annual equity premium and the spread have an absolute value of 3% using estimated parameter values. From the derivations above, we know that these estimated parameters imply that the cash-flow model must be one of long-run insurance rather than long-run risk.

In the bottom portion of Table A4, we present the asset pricing implications of the VARMA model estimated with and without the Great Recession data. For each set of estimated parameters, we also report the asset pricing implications of our model when we zero-out the 'off-diagonal' Kalman-gains, $K_{cd} = K_{dc} = 0$. We note three facts. First, under all of our VARMA estimates, the equity term structure spread is negative, as in the ARMA model. Second, the off-diagonal Kalman-gains are close to zero and statistically insignificant. When all other parameters are kept constant and the condition $K_{cd} = K_{dc} = 0$ is imposed, the equity term structure tends to be more downward sloping. In the VARMA model, similar to full information, a more negative term spread comes at the cost of having a lower equity premium. Since in the VARMA estimation of the model there is a lower correlation between $\upsilon_{c,t}$ and $\upsilon_{d,t}$, the short-run risk in these specifications (even those with the off diagonal elements zeroed out) is lower and the overall equity premium is not as high as under the ARMA model.

Table A4

Sample	Ψ	γ	$\widehat{\rho}$	$\widehat{K}_{cc}$	$\widehat{K}_{dd}$	$\widehat{K}_{cd}$	$\widehat{K}_{dc}$	$\widehat{\sigma}_{v_c}$	$\widehat{\sigma}_{v_d}$	$\widehat{\rho}_{v_c,v_d}$	EP	S	RF
						ARMA(1,1)							
1931-08	1.8	10	0.999	0.037	0.007	–	–	1.8	7.9	0.55	2.49	-3.85	0.79
1931-12	1.8	10	0.999	0.040	0.009	–	–	2.0	8.6	0.52	3.00	-2.96	0.58
						VARMA(1,1)							
1931-08	1.8	10	0.998	0.043	0.011	0.000	0.003	1.7	6.1	0.49	1.85	-0.67	0.87
	1.8	10	0.998	0.043	0.011	0	0	1.7	6.1	0.49	1.52	-2.27	0.88
1931-12	1.8	10	0.999	0.026	0.004	0.000	0.005	1.7	6.1	0.45	1.23	-2.32	1.22
	1.8	10	0.999	0.026	0.004	0	0	1.7	6.1	0.45	0.75	-7.45	1.23

Notes - The top portion of this table reports the point estimates of the parameters of two ARMA(1,1) processes for consumption and dividends growth:

$$\begin{aligned}\Delta c_t &= \mu(1-\rho) + \rho\Delta c_{t-1} - (\rho - K_{cc})v_{c,t-1} + v_{c,t} \\ \Delta d_t &= \mu(1-\rho) + \rho\Delta d_{t-1} - (\rho - K_{dd})v_{d,t-1} + v_{d,t}.\end{aligned}$$

The bottom portion of this table reports the point estimates of the parameters of a joint VARMA(1,1) process for consumption and dividends growth:

$$\begin{aligned}\Delta c_t &= \mu(1-\rho) + \rho\Delta c_{t-1} - (\rho - K_{cc})v_{c,t-1} + K_{cd}v_{d,t} + v_{c,t} \\ \Delta d_t &= \mu(1-\rho) + \rho\Delta d_{t-1} - (\rho - K_{dd})v_{d,t-1} + K_{dc}v_{c,t} + v_{d,t}.\end{aligned}$$

The standard deviations of the consumption and the dividend innovations are denoted as σ_{v_c} and σ_{v_d}, respectively. The correlation of these innovations is $\rho_{c,d}$. Real consumption is obtained from the Bureau of Economic Analysis. Real dividends are from the R. Shiller data set, available online. For each set of estimates, we report the model-implied equity premium (EP), average equity term structure spread (S), and average risk-free rate (RF). For each VARMA(1,1) estimation, we also report the asset pricing implications of our model when $K_{cd} = K_{dc} = 0$. When computing the asset pricing statistics from our model, we impose $\mu = 0.02$ for both dividends and consumption across all cases. The preference parameters are calibrated to the values reported in the table.

References

Campbell, John Y., "A Variance Decomposition for Stock Returns," *Economic Journal*, 1991, *101*, 157–179.