# State-Space vs. VAR Models for Stock Returns

John H. Cochrane*

July 24, 2008

**Abstract**

State-space or latent-variable models for stock prices specify a process for expected returns and expected and unexpected dividend growth, and then derive dividend yields and returns from a present value relations. They are a useful structure for understanding and interpreting forecasting relations. In this note, I connect state-space representations with their observable counterparts, and VAR/ARMA representations recovered by forecasting regressions.

## 1 Introduction

This paper connects return forecasting regressions with "state-space" representations. In a "state-space" or latent variable representation, we start with a process for expected returns and expected dividend growth. We then derive the dividend yield as the present value of the dividend growth process. This is a nice system for interpreting forecasting regressions in terms of expected return variation and expected dividend growth variation.

As a simple and useful example of a "structural state-space representation", define $g_t = E_t(\Delta d_{t+1})$ and $\mu_t = E_t(r_{t+1})$, where both expectations are with respect to investors' information sets, and suppose these variables follow $AR(1)$ processes,

$$\begin{aligned}\mu_{t+1} &= \phi_\mu \mu_t + \varepsilon_{\mu,t+1} \\ g_{t+1} &= \phi_g g_t + \varepsilon_{g,t+1} \\ \Delta d_{t+1} &= g_t + \varepsilon_{d,t+1}.\end{aligned}$$

We can use the present value and return identities, derived below, to find dividend yields and returns,

$$dp_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} \left(r_{t+j} - \Delta d_{t+j}\right) = \frac{\mu_t}{1-\rho\phi_\mu} - \frac{g_t}{1-\rho\phi_g} = k_\mu \mu_t - k_g g_t$$

$$\begin{aligned}r_{t+1} &= -\rho dp_{t+1} + dp_t + \Delta d_{t+1} \\ &= \mu_t + \varepsilon_{d,t+1} - \rho(k_\mu \varepsilon_{\mu,t+1} - k_g \varepsilon_{g,t+1})\end{aligned}$$

*Graduate School of Business, University of Chicago, and NBER. 5807 S. Woodlawn, Chicago IL 60637, john.cochrane@chicagogsb.edu. http://faculty.chicagogsb.edu/john.cochrane/research/Papers/ I thank the CRSP for research support.

We can think of the first three equations as a specification of the environment and the last two as the result of asset pricing. The dividend yield and return shocks

$$\begin{aligned} \varepsilon_{dp,t} &= k_{\mu}\varepsilon_{\mu,t} - k_g\varepsilon_{g,t} \\ \varepsilon_{r,t+1} &= \varepsilon_{d,t+1} - \rho\varepsilon_{dp,t+1} \end{aligned}$$

also follow from the "structural" shocks by identities. Thus, we only need specify the covariance matrix of the "structural" shocks $cov\left(\left[\begin{array}{ccc} \varepsilon_g & \varepsilon_\mu & \varepsilon_d \end{array}\right]\right)$ to complete the specification of this model.

I first connect this "structural" representation with an "observable state-space representation." Agents observe the $\varepsilon$ shocks and the state variables $\mu, g$, but we do not. The observable representation is of the same form, but can be estimated using data on dividend yields, dividend growth and returns only. It is just a way to rewrite the Wold representation of dividend yields, dividend growth and returns in a form that looks a lot like the underlying "structure." It turns out that the observable representation has exactly the same form, but the shocks have one less dimension. In this example, the observable counterparts to the shocks $\varepsilon_g, \varepsilon_\mu, \varepsilon_d$ are driven by two regression errors, shocks to dividend yields and dividend growth which I denote $v_{dp}, v_d$. This observable representation allows us to assess what is and is not identified of the structure, and it is the simplest form in which to estimate the structure by maximum likelihood or other full informaiton methods.

Then, I connect the state-space representations with the MA($\infty$) ARMA and VAR representations of the data. The latter question lets us ask, what does a model like the above imply for forecasting regressions? How can we interpret the result of arbitrary forecasting regressions in the context of a state-space structure? What do forecasting regressions say about variation in expected dividend growth and expected returns? How can we form estimates of the state-space parameters from forecasting regressions?

Of course, one can specify much more general models of this sort, in which expected returns and dividend growth have more complex dynamics, feed back on each other, are driven by ex-post dividend growth, and evolve jointly with other variables that we can measure. In fact, any arbitrary VAR can be written in this form, and one task of these notes is to connect an arbitrary VAR to its state-space representation. The last section of this paper tackles this general case.

The VAR implications of the above simple system show some forecasting promise. For example, when $\phi_\mu = \phi_g$, we can write the implied return forecasting regression (Equation (37) below)

$$r_{t+1} \approx \frac{\alpha_{\mu,dp}}{k} dp_t - \frac{\alpha_r}{k} \sum_{j=0}^{\infty} \left(\phi - \frac{\alpha_r}{k}\right)^j r_{t-j} + v_{r,t+1} \tag{1}$$

where $\alpha$ and $k$ are parameters defined below. This regression says that in addition to the usual coefficient on the dividend yield, a long moving average of dividend growth should help to forecast returns. This specification would be easy for standard regressions to miss, as regression coefficients can be insignificant when taken one at a time, but important when taken together. This expression nicely sums the two sensible ways of learning about time varying expected returns.

The general analysis in the last section dampens one's enthusiasm however. I show how to translate any VAR in to a state-space model, and I show by example that a standard first order VAR has a state space representation very like this simple example, enhanced only with cross effects by which $\mu_t$ forecasts $g_{t+1}$ and vice versa. Thus, adoption of a state space model is unlikely to believably enhance forecasting ability. The only hope is to find some economic restriction for parameters in the state space formulation (rather than the essentially pedagogical restrictins embodied in my example) that are not clear when translated to the VAR formulation.

Cochrane (1991, 2004,2008) presents a one-state-variable version of this model, in which expected returns vary but expected dividend growth does not, in order to interpret regressions of returns and dividend growth on dividend yields. Van Binsbergen and Koijen (2008) present a two state variable nonlinear model of the above form, estimate it by maximum likelihood, and find it is able to increase the return forecast $\mathrm{R}^2$ over simple dividend yield regressions. Essentially, they find that the second term in (1) matters.

## 1.1 Identities

The Campbell-Shiller (1988) linearized identities linking prices, dividends and returns give a structure that helps to digest stock return predictability. Returns are, to a very good approximation

$$r_{t+1} = -\rho dp_{t+1} + dp_t + \Delta d_{t+1}. \tag{2}$$

As a result, in any forecasting regression of the form

$$\begin{aligned} r_{t+1} &= b_{r,dp} dp_t + b_{r,x} x_t + v_{r,t+1} \\ dp_{t+1} &= b_{dp,dp} dp_t + b_{dp,x} x_t + v_{dp,t+1} \\ \Delta d_{t+1} &= b_{d,dp} dp_t + b_{d,x} x_t + v_{d,t+1} \end{aligned}$$

($x$ can be a vector) the regression coefficients must be related by

$$b_{r,dp} = 1 - \rho b_{dp,dp} + b_{d,dp} \tag{3}$$

$$b_{r,x} = b_{d,x} - \rho b_{dp,x}, \tag{4}$$

and the errors must obey

$$v_{r,t+1} = -\rho v_{dp,t+1} + v_{x,t+1}.$$

These identities mean that that given dividend yields, we can drop either returns or dividend growth forecasts from the analysis. If we predict $\{dp, r\}$ or $\{dp, \Delta d\}$, we can recover the missing variable. Equivalently, they mean that return, dividend growth, and dividend yield forecasts are mechanically linked, so that any behavior of one has implications for the other. Since the return identity (2) holds for each data point, these identities hold for each sample as well as for the underlying population coefficients.

The identity (2) is accurate, but it is only approximate. Thus, it is important either to work out its consequences for empirical work, as the computer will not automatically detect redundant variables. Alternatively, one can work with data that imposes the approximate identity. When I do that, I use exact measures of returns and imply dividend growth. Fundamentally we are interested in returns, and making inferences about returns from data that is not exactly a return is a dangerous pastime.

Iterating the return identity forward, we obtain the Campbell-Shiller (1988) linearized present value identity,

$$dp_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j}. \tag{5}$$

Since this is an identity, $E_t$ can refer to any information set that includes $dp_t$, and the investor's information set in particular. The price-dividend ratio reveals a slice of investor's information about expected returns and expected dividend growth, which is why it is so central to forecasting.

This identity admits additional forecasting variables of course. First, other variables may help to predict returns if they also help to predict dividend growth. Second, other variables may help to predict one-period returns if they predict a different time-path of returns. A variable may help to predict a higher $r_{t+1}$ with no ability to predict dividends if it also predicts lower $r_{t+j}$, i.e. if its variation holds $E_t \sum_{j=1}^{\infty} \rho^j r_{t+j}$ constant. We will see these effects in understanding how dividend growth or returns, and lags of dividend yields, can help to forecast one-period returns.

The identity and logic can help us to assess the plausibility of additional variables. A variable may appear help to forecast one-year returns, but its implied dividend-growth or long-term return forecast may be implausible.

## 1.2 One state variable model reminder

A one-state variable model has been very useful in digesting dividend-yield forecasts. The two-state model is a straightforward generalization, so it's worth remembering the one-state-variable model first.

*Facts*

Consider dividend-yield forecasts of returns and dividend growth that use only current dividend yields,

$$
\begin{aligned}
r_{t+1} &= b_r dp_t + v_{r,t+1}; \quad b_r \approx 0.1; \quad \sigma(v_r) \approx 0.16 && (6)\\
\Delta d_{t+1} &= b_d dp_t + v_{d,t+1}; \quad b_d \approx 0; \quad \sigma(v_d) \approx 0.14 && (7)\\
dp_{t+1} &= b_{dp} dp_t + v_{dp,t+1}; \quad b_{dp} \approx 0.94; \quad \sigma(v_{dp}) \approx 0.08 && (8)\\
cov(v_d, v_{dp}) &\approx 0 && (9)
\end{aligned}
$$

By the identity

$$r_{t+1} = -\rho dp_{t+1} + dp_t + \Delta d_{t+1} \tag{10}$$

one of the return or dividend growth forecasts follows from the other; we have therefore

$$b_r = 1 - \rho b_{dp} + b_d$$

and the numbers quoted above reflect this identity

$$0.1 \approx 1 - 0.96 \times 0.94 + 0.$$

Taking innovations, we have

$$v_{r,t+1} = -\rho v_{dp_{t+1}} + v_{d,t+1}$$

and hence

$$\sigma^2(v_r) = \rho^2 \sigma^2(v_{dp}) + \sigma^2(v_d) - 2\rho cov(v_d, v_{dp}).$$

The numbers quoted above also reflect this identity,

$$0.16^2 \approx 0.96^2 \times 0.08^2 + 0.14^2$$

It's an interesting fact in the data that dividend yield innovations and dividend growth innovations are essentially uncorrelated. This fact implies that return innovations and dividend yield innovations are highly negatively correlated.

$$\begin{aligned} cov(v_{dp}, v_r) &= cov(v_{dp}, v_d - \rho v_{dp}) = -\rho var(v_{dp}) \\ cov(v_d, v_r) &= cov(v_d, v_d - \rho v_{dp}) = var(v_d) \end{aligned}$$

This correlation is easiest to digest in regression coefficients

$$\begin{aligned} \beta_{v_r, v_d} &= cov(v_r, v_d)/var(v_d) = 1 \\ \beta_{v_r, v_{dp}} &= cov(v_r, v_{dp})/var(v_{dp}) \approx -\rho \end{aligned}$$

*One state variable model*

These results are nicely digested with the following standard one-state-variable model. Let $\mu_t \equiv E_t(r_{t+1})$ denote expected return, and suppose it follows an AR(1),

$$\mu_t = \phi_\mu \mu_{t-1} + \varepsilon_{\mu,t} \tag{11}$$

Suppose expected dividend growth is a constant, so actual dividend growth follows

$$\Delta d_t = \varepsilon_{d,t} \tag{12}$$

All variables are demeaned. We can find the dividend yield by the Campbell-Shiller linearized present value relation

$$dp_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} (r_{t+j} - \Delta d_{t+j}) = \frac{\mu_t}{1 - \rho\phi_\mu} = k_\mu \mu_t \tag{13}$$

Thus, the dividend yield reveals the expected return. Actual returns follow from the identity (10)

$$\begin{aligned} r_{t+1} &= -\rho dp_{t+1} + dp_t + \Delta d_{t+1} \\ r_{t+1} &= \mu_t + (v_{d,t+1} - \rho v_{dp_{t+1}}) \\ r_{t+1} &= (1 - \rho\phi_\mu) dp_t + (v_{d,t+1} - \rho v_{dp_{t+1}}) \end{aligned}$$

In sum, if the world follows the "structural" model of dividend growth and expected returns given by (11)-(12), and prices are generated by (13), then we will *observe* a VAR representation

$$\begin{aligned} dp_{t+1} &= \phi_\mu dp_t + \varepsilon_{dp,t+1}; \quad \varepsilon_{dp,t+1} = k_\mu \mu_{t+1} \\ \Delta d_{t+1} &= \varepsilon_{d,t+1} \\ r_{t+1} &= (1 - \rho\phi_\mu) dp_t + \varepsilon_{r,t+1}; \quad \varepsilon_{r,t+1} = \varepsilon_{d,t+1} - \rho \varepsilon_{dp,t+1} = \varepsilon_{d,t+1} - \rho k_\mu \varepsilon_{\mu,t+1} \end{aligned}$$

Deliciously, the regression and "structural" model match perfectly. We can interpret the dividend yield forecast error as a pure "expected return shock." We can interpret the dividend growth realization and forecast error as a pure "cashflow shock." The empirical fact that these two shocks are essentially uncorrelated frees us from orthogonalization worries.

Dividend yields reveal expected returns. Actual return shocks are a combination of current dividend growth shocks $\varepsilon_d$ and dividend yield or expected return shocks. Returns are high if there is good cash flow news (expected to last forever) or if there is "good" (lower) expected return news. A change in prices with no contemporaneous change in dividends isolates a completely transitory price movement. All variation in dividend yields comes from varying expected returns. Since

$$\begin{aligned} var(\varepsilon_r) &= var(\varepsilon_d) + \rho^2 var(\varepsilon_{dp}) \\ 0.16^2 &= 0.14^2 + 0.96^2 \times 0.08^2 \end{aligned}$$

roughly $0.14^2/0.16^2 = 0.77\%$ of the variance of returns comes from current cashflow news, none from future cashflow news, and the rest from expected return news. Though many authors have confused the calculation, the statement "all variation in dividend yields comes from discount rates" and "discount rate and cashflow news each account for about half of the variation of returns" are in fact completely consistent. Returns are affected by current cashflows and dividend yields are only affected by expected future cashflows.

*Additional variables*

This model makes one dramatically wrong prediction. If we interpret expectations to include *all* available information, it says that no *other* variable may predict returns in addition to the dividend yield, and no other variable may predict dividend growth at all. This prediction is false, and integrating this fact into our understanding is the point of this article.

The one-state variable model is not *wrong*. If we accept the estimates in which lagged returns and lagged dividend growth do not help to predict returns or dividend growth, we can interpret both the VAR and the structural model to reflect expectations given an information set consisting of past returns, dividends, and dividend yields, $E_t(\cdot) = E(\cdot|I_t)$, $I_t = \{dp_t, r_t, \Delta d_t, dp_{t-1}, r_{t-1}\Delta d_{t-1}, ...\}$. The fact that *agents* may see more variables, and those additional variables may forecast returns, does not invalidate the VAR as a statement of the Wold representation of dividend yields, dividend growth and returns, nor does it invalidate our "model" as a representation of expectations *based on that information set.* Nor does it invalidate our decomposition of variables into "expected returns" and "expected cashflow" effects, so long as we understand "expected" to be respect to $I_t$ and not investor information sets.

Still, investors *do* see more, and other variables *do* forecast returns and dividend growth. My objective is to extend this lovely structure to accommodate that fact. We still will benchmark to the basic facts presented above – one constraint on our thinking about other variables is that if we condition down to dividend yields only, we must obtain the regressions above. But we will also be able to incorporate additional facts.

## 1.3 An ARMA(1,1) reminder

Our first task will be to take the system outlined in the introduction and find its observable implications – what does it imply for the time series behavior of dividend yields, dividend growth, and returns. The corresponding representation questions for a single variable are worth remembering, as the concepts are the same for the larger system but the algebra is much worse. Start with the "structural" system

$$\begin{aligned} \mu_t &= \phi \mu_{t-1} + \varepsilon_{\mu,t} \\ r_{t+1} &= \mu_t + \varepsilon_{r,t+1} \end{aligned} \tag{14}$$

We only observe the history of $r_t$, so our information set is $I_t = \{r_t, r_{t-1}, ...\}$. Thus, our task is connect an *observable* counterpart with this "structural" representation. Since the model is linear, the observable counterpart is just the univariate Wold representation of returns.

The easiest way to write the univariate representation is in state-space form. I'll call this the "observable state-space representation." Define

$$\hat{\mu}_t = E(r_{t+1}|I_t)$$

Then, we can write

$$\begin{aligned} \hat{\mu}_{t+1} &= \phi\hat{\mu}_t + \alpha v_{r,t+1} \\ r_{t+1} &= \hat{\mu}_t + v_{r,t+1}. \end{aligned} \tag{15}$$

where

$$v_{r,t+1} \equiv r_{t+1} - E(r_{t+1}|I_t)$$

$\alpha$ is a messy function of the structural parameters, including the covariance matrix of $\varepsilon_\mu, \varepsilon_g$, which we can find by matching second moments. However, to every structural model there *is* an alpha and for every alpha there are many structural models. Thus, for all observable questions we can simply leave $\alpha$ as a parameter to be estimated.

The difference between the "structural" model and this observable model is only in the shocks. Since in the end we are describing the univariate Wold representation of $\{r_t\}$, there can only be a single shock $v_{t+1}$, and therefore we must write an observable state-space model in this form. The shocks in (15)must be perfectly correlated. Equivalently, the correlation between shocks in (14) is not identified.

It's often convenient to express the model in AR($\infty$), MA($\infty$) or finite-order ARMA forms. In particular, this exercise will let us estimate the model from simple regressions, and it will let us connect restrictions on the "structural" form to their predictions for regressions. To derive these representations, solve (15)

$$\hat{\mu}_t = \frac{1}{1-\phi L}\alpha v_{r,t}$$

and substitute

$$r_{t+1} = \frac{1}{1-\phi L}\alpha v_{r,t} + v_{r,t+1} = \left(1 + \frac{\alpha L}{1-\phi L}\right) v_{r,t+1} = \left(\frac{1+(\alpha-\phi)L}{1-\phi L}\right) v_{r,t+1}.$$

This is the Wold moving average representation.

This model has a finite-order ARMA(1,1) representation,

$$(1-\phi L)r_{t+1} = [1 + (\alpha - \phi)L]\, v_{r,t+1}.$$

The AR($\infty$) representation lets us connect to forecasting regressions,

$$\begin{aligned} \frac{(1-\phi L)}{1+(\alpha-\phi)L} r_{t+1} &= v_{r,t+1} \\ \left(1 - \frac{\alpha L}{1-(\phi-\alpha)L}\right) r_{t+1} &= v_{r,t+1} \end{aligned}$$

$$r_{t+1} = \alpha \sum_{j=0}^{\infty} (\phi - \alpha)^j r_{t-j} + v_{r,t+1}$$

This latter formula is quite sensible. To recover a slow-moving time-varying conditional mean from this time series, filter it by applying a long moving average. If one had a good idea of $\phi$ (and we will in the multivariate setting), then a regression of $r_{t+1}$ on a weighted moving average could give a first estimate of $\alpha$, which of course governs the variation in the conditional mean.

The state-space representation (15) is exactly equivalent to this more familiar ARMA(1,1) representation. For example, we could use either representation to produce a series of innovations $\{v_t\}$ given a guess of the parameters, and then estimate parameters by maximum likelihood. They each represent the data with one additional state variable, either $\mu_t$ or the additional lag of the shock.

State-space or ARMA models are conventionally estimated by maximum likelihood. This is convenient and efficient but not necessary. The ARMA(1,1) is identified by matching the variance and first autocovariance. Thus one can find the observable parameters from those two moments, or equivalently, from a first order VAR. Yes, the first order VAR is "misspecified" in that its forecasts are not efficient. But it encodes the first two moments, and one can invert those moments to find the ARMA(1,1) parameters.

To prove that the "reduced form" (15) does in fact correspond to the structure (14), and to express the mapping between parameters, we can match second moments. (These are the "messy formulas" for $\alpha$.) From (14), we have

$$\begin{aligned}
var(r_t) &= var(\mu_t) + var(\varepsilon_r) = \frac{var(\varepsilon_\mu)}{1-\phi^2} + var(\varepsilon_r) \\
cov(r_t, r_{t-1}) &= cov(\phi\mu_{t-1} + \varepsilon_{\mu,t}, \mu_{t-1} + \varepsilon_{r,t}) = \frac{\phi var(\varepsilon_\mu)}{1-\phi^2} + cov(\varepsilon_r, \varepsilon_\mu) \\
cov(r_t, r_{t-2}) &= cov(\phi^2\mu_{t-3} + \phi\varepsilon_{\mu,t-2} + \varepsilon_{\mu,t-1} + \varepsilon_{r,t}, \mu_{t-3} + \varepsilon_{r,t-2}) = \phi\left(\frac{\phi var(\varepsilon_\mu)}{1-\phi^2} + cov(\varepsilon_r, \varepsilon_\mu)\right)
\end{aligned}$$

We see the familiar autocovariance function of the ARMA(1,1). The state-space observable model (15) has the same form, so it produces

$$\begin{aligned}
var(r_t) &= \left(\frac{\alpha^2}{1-\phi^2} + 1\right) var(v_r) \\
cov(r_t, r_{t-1}) &= \left(\frac{\phi\alpha^2}{1-\phi^2} + \alpha\right) var(v_r) \\
cov(r_t, r_{t-2}) &= \phi\left(\frac{\phi\alpha^2}{1-\phi^2} + \alpha\right) var(v_r)
\end{aligned}$$

The higher order autocorrelations die at the rate $\phi$ in both models, so the $\phi$ parameter is identified and it is the same. (Knowing this in advance I avoided the use of distinct symbols.) The remaining parameters then must satisfy

$$\begin{aligned}
\left(\frac{\alpha^2}{1-\phi^2} + 1\right) var(v_r) &= \frac{var(\varepsilon_\mu)}{1-\phi^2} + var(\varepsilon_r) \\
\left(\frac{\phi\alpha^2}{1-\phi^2} + \alpha\right) var(v_r) &= \frac{\phi var(\varepsilon_\mu)}{1-\phi^2} + cov(\varepsilon_r, \varepsilon_\mu)
\end{aligned}$$

These equations identify $\alpha$ and $var\,(v_r)$.

Solving from structure to observable, we have

$$\frac{\frac{\alpha^2}{1-\phi^2}+1}{\frac{\phi\alpha^2}{1-\phi^2}+\alpha} = \frac{\frac{1}{1-\phi^2}\frac{var(\varepsilon_\mu)}{var(\varepsilon_r)}+1}{\frac{\phi}{1-\phi^2}\frac{var(\varepsilon_\mu)}{var(\varepsilon_r)}+\frac{cov(\varepsilon_r,\varepsilon_\mu)}{var(\varepsilon_r)}}$$

$$\frac{\frac{1}{1-\phi^2}\alpha^2+1}{\frac{\phi}{1-\phi^2}\alpha^2+\alpha} = \frac{\frac{1}{1-\phi^2}\left[\beta_{\mu,r}^2+\frac{var(\varepsilon)}{var(\varepsilon_r)}\right]+1}{\frac{\phi}{1-\phi^2}\left[\beta_{\mu,r}^2+\frac{var(\varepsilon)}{var(\varepsilon_r)}\right]+\beta_{\mu,r}}$$

$$\frac{\frac{1}{1-\phi^2}\alpha^2+1}{\frac{1}{1-\phi^2}\alpha^2+\frac{\alpha}{\phi}} = \frac{\frac{1}{1-\phi^2}\left[\beta_{\mu,r}^2+\frac{var(\varepsilon)}{var(\varepsilon_r)}\right]+1}{\frac{1}{1-\phi^2}\left[\beta_{\mu,r}^2+\frac{var(\varepsilon)}{var(\varepsilon_r)}\right]+\frac{\beta_{\mu,r}}{\phi}}$$

This equation identifies $\alpha$, and then

$$\frac{var(v_r)}{var(\varepsilon_r)} = \frac{\frac{\beta_{\mu,r}^2+\frac{var(\varepsilon)}{var(\varepsilon_r)}}{1-\phi^2}+1}{\frac{\alpha^2}{1-\phi^2}+1}$$

determines $var\,(v_r)$.

Thus, for any $\alpha$, there *is* a structure with $\beta_{r\mu}=\alpha$ and $var(\varepsilon)=0$. I.e., we could be looking directly at the structure. However, many structures are compatible with a given $\alpha$. As $var(\varepsilon)$ increases – as the "structural" shocks become less correlated, and as agents gain a larger information advantage, $\alpha$ rises relative to $\beta_{r\mu}$. The general formula is a messy quadratic that is easy to solve numerically, but gives no particular intuition.

The "structural" model has two ways to create a variance larger than the geometrically decaying covariances would suggest: a larger $\beta_{r\mu}$ and a larger $var(\varepsilon)$. The observable model only has one such mechanism, a large $\alpha$. Thus, for example, a structural model with uncorrelated shocks $\beta_{r\mu}=0$ must still produce an observable model with perfectly correlated shocks and thus a large $\alpha$. *Correlation of the observable shocks does not imply correlation of the structural shocks.* This lesson is important in evaluating the multivariate observable models that follow.

# 2 A simple two-state-variable model

So, it is useful to interpret return forecasting regressions involving the dividend yield in terms of a simple "structural" model in which expected returns and expected and actual dividend growth vary over time; prices are generated as present values of dividends and returns are generated from the price and dividend process. We have to generalize the model to allow for variation in expected dividend growth, and to allow for other variables.

## 2.1 "Structural" state-space representation

Again, as a simple and useful example, define $g_t = E_t(\Delta d_{t+1})$ and $\mu_t = E_t(r_{t+1})$, where both expectations are with respect to investors' information sets, and suppose these variables follow

$AR(1)$ processes,

$$\mu_{t+1} = \phi_\mu \mu_t + \varepsilon_{\mu,t+1} \quad (16)$$
$$g_{t+1} = \phi_g g_t + \varepsilon_{g,t+1} \quad (17)$$
$$\Delta d_{t+1} = g_t + \varepsilon_{d,t+1}. \quad (18)$$

We can use the present value and return identities to find dividend yields and returns,

$$dp_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} (r_{t+j} - \Delta d_{t+j}) = \frac{\mu_t}{1-\rho\phi_\mu} - \frac{g_t}{1-\rho\phi_g} = k_\mu \mu_t - k_g g_t \quad (19)$$

$$\begin{aligned} r_{t+1} &= -\rho dp_{t+1} + dp_t + \Delta d_{t+1} \quad (20) \\ &= \mu_t + \varepsilon_{d,t+1} - \rho(k_\mu \varepsilon_{\mu,t+1} - k_g \varepsilon_{g,t+1}) \end{aligned}$$

We can think of the first three equations as a specification of the environment and the last two as the result of asset pricing. From (19) and (20), the dividend yield and return shocks

$$\varepsilon_{dp,t} = k_\mu \varepsilon_{\mu,t} - k_g \varepsilon_{g,t} \quad (21)$$
$$\varepsilon_{r,t+1} = \varepsilon_{d,t+1} - \rho \varepsilon_{dp,t+1} \quad (22)$$

also follow from the "structural" shocks by identities. Thus, we only need specify the covariance matrix of the "structural" shocks $cov\left(\left[ \begin{array}{ccc} \varepsilon_g & \varepsilon_\mu & \varepsilon_d \end{array} \right]\right)$ to complete the specification of this model.

Remember that this is *only* an example, as it embodies many restrictions that have no economic rationale. In (16)-(17), there is no reason that expected returns and expected dividend growth should evolve independently, or in AR(1) fashion. In reality, other variables coevolve with expected returns and expected dividend growth, and help us to recover their values. And, while tempting, there is no economic reason to impose a particular correlation structure on the shocks.

## 2.2 Observable state-space representation

*Agents* see the state variables $\mu$ and $g$, but we do not. Hence, we have to ask what are the observable implications of these "structures," and thus conversely how much of the structures can be identified from data and how much needs extra assumptions. The logic of this reduction from structural to observable is exactly analogous to what we just did in the ARMA(1,1) case.

Our information set consists of returns, dividend growth, dividend yields, and additional $z$ variables. $I_t = \{dp_t, r_t, \Delta d_t, z_t, \quad dp_{t-1}, r_{t-1}, \Delta d_{t-1}, z_{t-1}, ... \}$ From the return identity (10), one of $r$ or $\Delta d$ is superfluous. For example, we might summarize what we can learn about the data by a VAR involving $\{dp_t, \Delta d_t, z_t\}$.

We can write the observable implications of the "structural" model is simply to project everything on the information set $I_t$. Our best guesses of the unobserved state variables $\mu$ and $g$ appear in their place.

$$\begin{aligned} \hat{\mu}_t &\equiv E(r_{t+1}|I_t) = E(\mu_t|I_t) \\ \hat{g}_t &\equiv E(\Delta d_{t+1}|I_t) = E(g_t|I_t). \end{aligned}$$

Observable regression residuals $v$, appear in the place of unobservable "structural" shocks $\varepsilon$,

$$v_{x,t+1} = x_{t+1} - E(x_{t+1}|I_t).$$

Thus, the observable counterpart to the simple example (16)-(20) is

$$\hat{\mu}_{t+1} = \phi_\mu \hat{\mu}_t + v_{\mu,t+1} \tag{23a}$$
$$\hat{g}_{t+1} = \phi_g \hat{g}_t + v_{g,t+1} \tag{23b}$$
$$\Delta d_{t+1} = \hat{g}_t + v_{d,t+1}. \tag{23c}$$

and, either by projection or by applying the same present value and return identities,

$$dp_t = k_\mu \hat{\mu}_t - k_g \hat{g}_t \tag{24}$$
$$v_{dp,t+1} = k_\mu v_{\mu,t+1} - k_g v_{g,t+1} \tag{25}$$

$$r_{t+1} = \hat{\mu}_t + v_{r,t+1} \tag{26a}$$
$$v_{r,t+1} = v_{d,t+1} - \rho v_{dp,t+1}. \tag{26b}$$

In the simple example structure (16)-(21) there are three potentially independent shocks $\{\varepsilon_\mu, \varepsilon_g, \varepsilon_d\}$, since agents can see all the variables. In the end, this observable version is equivalent to the ARMA Wold representation for $\{dp_t, \Delta d_t\}$ (or $\{dp_t, r_t\}$), so the shocks $\{v_g, v_\mu, v_d\}$ must be linear functions of the two observable shocks $\{v_{dp}, v_d\}$ only. In addition, the $v$ shocks must obey (25). Hence, we must be able to write the observable shocks $v_\mu$ and $v_g$ in the form

$$\begin{aligned} v_{\mu,t+1} &= \alpha_{\mu,dp} v_{dp,t+1} + \alpha_{\mu,d} v_{d,t+1} \\ v_{g,t+1} &= \alpha_{g,dp} v_{dp,t+1} + \alpha_{g,d} v_{d,t+1} \\ 1 &= k_\mu \alpha_{\mu,dp} - k_g \alpha_{g,dp}; \quad \alpha_{\mu,d} - \alpha_{g,d} = 0 \end{aligned}$$

Imposing the constraints, we can parameterize the $\alpha$ by

$$v_{\mu,t+1} = \frac{\alpha_{\mu,dp}}{k_\mu} v_{dp,t+1} + \frac{\alpha_d}{k_\mu} v_{d,t+1} \tag{27}$$
$$v_{g,t+1} = \frac{\alpha_{\mu,dp} - 1}{k_g} v_{dp,t+1} + \frac{\alpha_d}{k_g} v_{d,t+1}. \tag{28}$$

(We could just as easily parameterize the second equation with $\alpha_{g,dp}$ and then $\alpha_{\mu,dp} = \alpha_{g,dp} - 1$.) These expressions make clear how the $v_\mu$ and $v_g$ shocks are observable – they are combinations of the observable regression errors $v_{dp}$ and $v_d$. Then $\hat{\mu}$ and $\hat{g}$ are moving averages of past dividend yield and dividend growth shocks, and thus moving averages of past dividend yield and dividend growth, which shows how they are observed. Given (27)-(28), we can also regard (23a) and (23b) as filtering formulas.

Equations (27) and (28) seem to mix "structure" ($v_\mu$ and $v_g$ ) and "result" ($v_{dp}$). If one wants to specify a "structure" before knowing the "result" $v_{dp}$, how does one proceed? In fact, however, we can read equations (27) and (28) as a recipe for constructing a valid "structural" covariance matrix before knowing the result. The form of the model places no restrictions on the covariance matrix of $\{v_{dp}, v_d\}$ – any covariance is possible. Hence, if we want to specify a valid model, we can start with two shocks, $\varepsilon_x, \varepsilon_y$, with arbitrary covariance matrix $\Sigma = cov(\varepsilon_x, \varepsilon_y)$. We can then pick arbitrary numbers $\alpha_{dp}$ and $\alpha_d$, and we can construct the model shocks $\{v_g, v_\mu, v_d\}$ by

$$\begin{bmatrix} v_{\mu,t} \\ v_{g,t} \\ v_{d,t} \end{bmatrix} = \begin{bmatrix} \frac{\alpha_{\mu dp}}{k_\mu} & \frac{\alpha_d}{k_\mu} \\ \frac{\alpha_{\mu,dp}-1}{k_g} & \frac{\alpha_d}{k_g} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{bmatrix},$$

meaning just that we specify that the shocks have covariance matrix

$$cov\left(\begin{bmatrix} v_{\mu,t} \\ v_{g,t} \\ v_{d,t} \end{bmatrix}\right) = \begin{bmatrix} \frac{\alpha_{\mu,dp}}{k_\mu} & \frac{\alpha_d}{k_\mu} \\ \frac{\alpha_{\mu,dp}-1}{k_g} & \frac{\alpha_d}{k_g} \\ 0 & 1 \end{bmatrix} \Sigma \begin{bmatrix} \frac{\alpha_{\mu,dp}}{k_\mu} & \frac{\alpha_{\mu,dp}-1}{k_g} & 0 \\ \frac{\alpha_d}{k_\mu} & \frac{\alpha_d}{k_g} & 1 \end{bmatrix}.$$

This is just a recipe for constructing a shock covariance matrix that obeys a certain set of restrictions; it is just a way of parameterizing a valid (for observability) $3 \times 3$ singular shock covariance matrix via $\alpha_{dp}, \alpha_d$ and $\Sigma$. If we now *construct* $dp_t$ by the present value identity, we will find that

$$\begin{aligned} v_{dp,t} &= k_\mu v_{\mu,t} - k_g v_{g,t} = \varepsilon_{x,t} \\ v_{d,t} &= \varepsilon_{y,t} \end{aligned}$$

In other words, dividend yields *reveal* the shock $\varepsilon_{x,t}$, just as dividend yields revealed the expected return shock $\varepsilon_{\mu,t}$ in the one-state-variable system (13). Thus, formulas (27)-(28) are just a parametric way of constructing a covariance matrix of shocks $\{v_g, v_\mu, v_d\}$ that is restricted in a particular way so that the "structural" shocks are observable from the information set $I_t$.

The project seems hard: find the Wold representation of $\{dp_t, r_t, \Delta d_t\}$ that follows from the structure (40)-(43). The answer turns out to be almost trivially simple: The state-space representation of the observable model is of the same form as that of the "structural" model, but with linear functions of the two observable shocks $v_d, v_{dp}$ in place of the three structural shocks $\varepsilon_g, \varepsilon_d, \varepsilon_{dp}$, and our best guesses of the state variables $\hat{g}$ and $\hat{\mu}$ in place of their unobservable actual values.

We may want to express the model in reference to a $\{dp, r\}$ VAR rather than a $\{dp, \Delta d\}$ VAR. In place of (27)-(28) we can write

$$\begin{aligned} v_{\mu,t+1} &= \frac{\alpha_{\mu,dp}}{k_\mu} v_{dp,t+1} + \frac{\alpha_r}{k_\mu} v_{r,t+1} \\ v_{g,t+1} &= \frac{\alpha_{\mu,dp}-1}{k_g} v_{dp,t+1} + \frac{\alpha_r}{k_g} v_{r,t+1}. \end{aligned}$$

The only difference is $v_r$ in place of $v_d$. Since $v_r = -\rho v_{dp} + v_d$, the parameterization doesn't add or subtract anything, but it form will be prettier when referencing a $\{dp, r\}$ VAR.)

## 2.3 From structure to observable state-space representations and back

The $\alpha$ coefficients (or other parametrizations of the covariance structure of $\{v_\mu, v_g, v_d, v_z\}$) are messy functions of the parameters of the structural models (16)-(18) or (64)-(66), which we can find by matching second moments. If one wishes to derive $\alpha$ coefficients from a particular view of the structure, one must make that connection. The formulas are straightforward, but do not give any useful intuition.

The *form* of the answer is given by (23a)-(26b) with a shock covariance matrix that we can summarize by the covariances of $\{v_{dp}, v_d\}$ and the $\alpha$ coefficients of (27) in every case. Therefore, for forecasting purposes one may simply treat the $\alpha$ as the free parameters to be estimated, and I show below how to infer them from a VAR. For everything we can observe, the $\alpha$ parameters are enough.

One may wish to go backward, to learn about the "structural" representation (16)-(18) or (64)-(66) from the observable representations. This step requires identification assumptions. The observable representation is in the same class, it just happens to have a singular covariance matrix. To learn about the "structural" representation, one must add identifying assumptions to deny the possibility of this singular covariance matrix, and then work backward again through the messy formulas connecting $\alpha$ and the observable shock covariance matrix to the structural $\varepsilon$ covariance matrix, or estimate the "structural" model with its identifying restrictions directly, e.g. by maximum likelihood. For example, in the observable representation of the simple AR(1) model, $v_\mu, v_g$, and $v_d$ are spanned by two shocks. One might specify that expected returns, expected dividend growth and actual dividend growth are all uncorrelated with each other. There is always an observationally equivalent structure in which the covariance matrix is singular, but one may want to make calculations, such as variance decompositions that impose another view.

However, I find no such assumptions plausible for empirical specifications (i.e. disconnected from specific economic models). It's pedagogically interesting to specify uncorrelated $\mu$, $g$ and actual dividend $\Delta d$ shocks, but there is no economic reason to impose any particular value for their correlation. Campbell and Cochrane (1998) is only one obvious example of an economic model in which there is only *one* underlying shock. In that model a decline in actual consumption ($\Delta d$) simultaneously triggers a change in expected returns. I can't think of any general economic statement that would limit or specify the correlation between expected returns, expected dividend growth, unexpected returns, unexpected dividend growth, and other state variables $z_t$. In that state of affairs, the observable representation is as well as we can do.

This model can be estimated straightforwardly by maximum likelihood. For each choice of parameters, we can simulate the model forward to create a time-series of $v$ shocks, and then we can evaluate the likelihood of those shocks. It can also be estimated by matching second moments or by inverting forecasting regressions via the formulas derived below.

## 2.4 AR/ARMA representations and forecasting regressions

The observable state-space model is

$$\begin{aligned}
\hat{\mu}_{t+1} &= \phi_\mu \hat{\mu}_t + v_{\mu,t+1} \\
\hat{g}_{t+1} &= \phi_g \hat{g}_t + v_{g,t+1} \\
\Delta d_{t+1} &= \hat{g}_t + v_{d,t+1}.
\end{aligned}$$

$$\begin{aligned}
dp_t &= k_\mu \hat{\mu}_t - k_g \hat{g}_t \\
v_{dp,t+1} &= k_\mu v_{\mu,t+1} - k_g v_{g,t+1}
\end{aligned}$$

$$\begin{aligned}
r_{t+1} &= \hat{\mu}_t + v_{r,t+1} \\
v_{r,t+1} &= v_{d,t+1} - \rho v_{dp,t+1}.
\end{aligned}$$

$$
\begin{aligned}
v_{\mu,t+1} &= \frac{\alpha_{\mu,dp}}{k_\mu} v_{dp,t+1} + \frac{\alpha_d}{k_\mu} v_{d,t+1} \\
v_{g,t+1} &= \frac{\alpha_{\mu,dp}-1}{k_g} v_{dp,t+1} + \frac{\alpha_d}{k_g} v_{d,t+1}.
\end{aligned}
$$

To derive the VAR, MA or ARMA representations, we substitute out the state variables $\hat{\mu}$ and $\hat{g}$

$$
\begin{aligned}
\hat{\mu}_t &= \frac{1}{1-\phi_\mu L} v_{\mu t} = \frac{1}{1-\phi_\mu L}\left[\frac{\alpha_{\mu,dp}}{k_\mu} v_{dp,t} + \frac{\alpha_d}{k_\mu} v_{d,t}\right] \\
\hat{g}_t &= \frac{1}{1-\phi_g L} v_{gt} = \frac{1}{1-\phi_g L}\left[\frac{\alpha_{\mu,dp}-1}{k_g} v_{dp,t} + \frac{\alpha_d}{k_g} v_{d,t}\right]
\end{aligned}
$$

$$
\begin{aligned}
dp_t &= \frac{1}{1-\phi_\mu L}\left(\alpha_{\mu,dp} v_{dp,t} + \alpha_d v_{d,t}\right) - \frac{1}{1-\phi_g L}\left[\left(\alpha_{\mu,dp}-1\right) v_{dp,t} + \alpha_d v_{d,t}\right] \\
dp_t &= \left(\frac{\alpha_{\mu,dp}}{1-\phi_\mu L} - \frac{\alpha_{\mu,dp}-1}{1-\phi_g L}\right) v_{dp,t} + \left(\frac{\alpha_d}{1-\phi_\mu L} - \frac{\alpha_d}{1-\phi_g L}\right) v_{d,t}
\end{aligned}
$$

### 2.4.1 MA representation

Hence we have the *MA representation*

$$
\begin{aligned}
dp_t &= \left(\frac{\alpha_{\mu,dp}}{1-\phi_\mu L} - \frac{\alpha_{\mu,dp}-1}{1-\phi_g L}\right) v_{dp,t} + \left(\frac{\alpha_d}{1-\phi_\mu L} - \frac{\alpha_d}{1-\phi_g L}\right) v_{d,t} \\
\Delta d_t &= \hat{g}_{t-1} + v_{d,t} = \frac{L}{1-\phi_g L}\frac{\alpha_{\mu,dp}-1}{k_g} v_{dp,t} + \left(1 + \frac{L}{1-\phi_g L}\frac{\alpha_d}{k_g}\right) v_{d,t}
\end{aligned}
$$

Collapsing the lag operator polynomials,

$$
\begin{aligned}
dp_t &= \left[\frac{1-\left[\phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu - \phi_g\right)\right] L}{\left(1-\phi_\mu L\right)\left(1-\phi_g L\right)}\right] v_{dp,t} + \alpha_d \frac{\left(\phi_\mu - \phi_g\right) L}{\left(1-\phi_\mu L\right)\left(1-\phi_g L\right)} v_{d,t} \\
\Delta d_t &= \frac{L}{1-\phi_g L}\frac{\alpha_{\mu,dp}-1}{k_g} v_{dp,t} + \frac{1-\left(\phi_g - \alpha_d/k_g\right) L}{1-\phi_g L} v_{d,t}
\end{aligned}
$$

or, finally, in a pretty vector format

$$
\begin{bmatrix} dp_t \\ \Delta d_t \end{bmatrix} = \begin{bmatrix} \frac{1-\left[\phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu - \phi_g\right)\right] L}{\left(1-\phi_\mu L\right)\left(1-\phi_g L\right)} & \frac{\alpha_d\left(\phi_\mu - \phi_g\right) L}{\left(1-\phi_\mu L\right)\left(1-\phi_g L\right)} \\ \frac{\alpha_{\mu,dp}-1}{k_g}\frac{L}{1-\phi_g L} & \frac{1-\left(\phi_g - \alpha_d/k_g\right) L}{1-\phi_g L} \end{bmatrix} \begin{bmatrix} v_{dp,t} \\ v_{d,t} \end{bmatrix}
$$

Returns follow from

$$
\begin{aligned}
r_{t+1} &= \hat{\mu}_t + v_{r,t+1} \\
&= \frac{1}{1-\phi_\mu L}\left[\frac{\alpha_{\mu,dp}}{k_\mu} v_{dp,t} + \frac{\alpha_d}{k_\mu} v_{d,t}\right] + v_{d,t+1} - \rho v_{dp,t+1}
\end{aligned}
$$

### 2.4.2 ARMA representation

Multiplying out the denominator lag polynomials, this model has a *ARMA(1,1) representation*

$$
\begin{aligned}
\left(1-\phi_\mu L\right)\left(1-\phi_g L\right) dp_t &= \left(1-\left[\phi_\mu-\alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)\right] L\right) v_{dp,t}+\alpha_d\left(\phi_\mu-\phi_g\right) L v_{d,t} \\
(1-\phi_g L)\Delta d_t &= \frac{\alpha_{\mu,dp}-1}{k_g} L v_{dp,t}+\left(1-\left(\phi_g-\frac{\alpha_d}{k_g}\right) L\right) v_{d,t}
\end{aligned}
$$

$$
\begin{bmatrix} \left(1-\phi_\mu L\right)\left(1-\phi_g L\right) & 0 \\ 0 & \left(1-\phi_g L\right) \end{bmatrix}\begin{bmatrix} dp_t \\ \Delta d_t \end{bmatrix}=\left(I-\begin{bmatrix} \phi_\mu-\alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right) & -\alpha_d\left(\phi_\mu-\phi_g\right) \\ \frac{1-\alpha_{\mu,dp}}{k_g} & \phi_g-\frac{\alpha_d}{k_g} \end{bmatrix} L\right)\begin{bmatrix} v_{dp,t} \\ v_{d,t} \end{bmatrix} \tag{32}
$$

Returns follow

$$
(1-\phi_\mu L) r_{t+1}=\left[\left(\frac{\alpha_{\mu,dp}}{k_\mu}+\rho\phi_\mu\right) v_{dp,t}+\left(\frac{\alpha_d}{k_\mu}-\phi_\mu\right) v_{d,t}\right]+\left(v_{d,t+1}-\rho v_{dp,t+1}\right)
$$

and thus also fit in the VARMA(1,1) representation.

### 2.4.3 AR representations and forecasting regressions

Finally, we can invert the MA lag polynomial to find the AR representation, which describes forecasting regressions. Alas, inverting or raising the right-hand matrix to powers does not lead to easily interpreted quantities, so I'll show the main features in a series of special cases.

*The case* $\phi_\mu = \phi_g$

Start with a special case, $\phi_\mu = \phi_g = \phi$ so $k_g = k_\mu = k = (1-\rho\phi)^{-1}$. In that case we have

$$
\begin{bmatrix} (1-\phi L)^2 & 0 \\ 0 & (1-\phi L) \end{bmatrix}\begin{bmatrix} dp_t \\ \Delta d_t \end{bmatrix}=\left(I-\begin{bmatrix} \phi & 0 \\ \frac{1-\alpha_{\mu,dp}}{k} & \phi-\frac{\alpha_d}{k} \end{bmatrix} L\right)\begin{bmatrix} v_{dp,t} \\ v_{d,t} \end{bmatrix}
$$

Canceling a root in the first equation, the autoregressive representation of the dividend yield is a simple AR(1),

$$
dp_t = \phi dp_{t-1} + v_{dp,t}.
$$

Thus, the structural parameter $\phi$ can be measured from the autoregression coefficient of dividend yields.

Dividend growth follows

$$
\begin{aligned}
(1-\phi L)\Delta d_{t+1} &= \frac{1-\alpha_{\mu,dp}}{k} v_{dp,t}-\left(\phi-\frac{\alpha_d}{k}\right) v_{d,t}+v_{d,t+1} \\
\Delta d_{t+1} &= \frac{1-\alpha_{\mu,dp}}{k} dp_t+\frac{1-\left(\phi-\frac{\alpha_d}{k}\right) L}{1-\phi L} v_{d,t+1}
\end{aligned}
$$

$$
\frac{1-\phi L}{1-\left(\phi-\frac{\alpha_d}{k}\right) L}\Delta d_{t+1}=\frac{1-\alpha_{\mu,dp}}{k}\left(\frac{1-\phi L}{1-\left(\phi-\frac{\alpha_d}{k}\right) L}\right) dp_t+v_{d,t+1} \tag{34}
$$

The last equation is the autoregressive representation for dividend growth, and describes the forecasting regression. More formally of course the autoregressive representation is

$$\left[\begin{array}{cc} 1-\phi L & 0 \\ -\frac{1-\alpha_{\mu,dp}}{k}\frac{1-\phi L}{1-\left(\phi-\frac{\alpha_d}{k}\right)L}L & \frac{1-\phi L}{1-\left(\phi-\frac{\alpha_d}{k}\right)L} \end{array}\right]\left[\begin{array}{c} dp_t \\ \Delta d_t \end{array}\right]=\left[\begin{array}{c} v_{dp,t} \\ v_{d,t} \end{array}\right]$$

We can also track the return forecasting regression in this system by using

$$\hat{\mu}_t=\frac{1}{1-\phi_\mu L}\left[\frac{\alpha_{\mu,dp}}{k_\mu}v_{dp,t}+\frac{\alpha_d}{k_\mu}v_{d,t}\right]$$

and thus

$$\begin{aligned} r_{t+1} &= \frac{1}{1-\phi L}\left[\begin{array}{cc} \frac{\alpha_{\mu,dp}}{k} & \frac{\alpha_d}{k} \end{array}\right]\left[\begin{array}{c} v_{dp,t} \\ v_{d,t} \end{array}\right]+v_{r,t+1} \\ &= \left[\begin{array}{cc} \frac{\alpha_{\mu,dp}}{k} & \frac{\alpha_d}{k} \end{array}\right]\left[\begin{array}{cc} 1 & 0 \\ -\frac{1-\alpha_{\mu,dp}}{k}\frac{1}{1-\left(\phi-\frac{\alpha_d}{k}\right)L}L & \frac{1}{1-\left(\phi-\frac{\alpha_d}{k}\right)L} \end{array}\right]\left[\begin{array}{c} dp_t \\ \Delta d_t \end{array}\right]+v_{r,t+1} \end{aligned}$$

$$r_{t+1}=\frac{\alpha_{\mu,dp}}{k}dp_t-\frac{\alpha_d}{k}\frac{1}{1-\left(\phi-\frac{\alpha_d}{k}\right)L}\left[\Delta d_t-\frac{(1-\alpha_{\mu,dp})}{k}dp_{t-1}\right]+v_{r,t+1} \quad (35)$$

To interpret (34) and (35), first, note that if $\alpha_{\mu,dp}=1$– if all dp shocks feed to returns, not dividend growth, and if $\alpha_d=0$ – dividend growth shocks as well as dp shocks don't affect expected dividend growth – then we return to the one-state variable model studied above. Dividend growth is just white noise,

$$\Delta d_{t+1}=v_{d,t+1}.$$

and return forecasts are just

$$r_{t+1}=\frac{1}{k}dp_t+v_{r,t+1}=(1-\rho\phi)\,dp_t+v_{r,t+1}$$

Second, allow arbitrary $\alpha_{dp}$ but start with $\alpha_d=0$ – we split dividend yield shocks between expected returns and dividend growth, but dividend growth shocks do not affect expected returns or expected dividend growth. Now we have a simple dividend yield and return forecast,

$$\begin{aligned} \Delta d_{t+1} &= \frac{1-\alpha_{\mu,dp}}{k}dp_t+v_{d,t+1}=\alpha_{g,dp}\,(1-\rho\phi)\,dp_t+v_{d,t+1} \\ r_{t+1} &= \frac{\alpha_{\mu,dp}}{k}dp_t+v_{r,t+1}=\alpha_{\mu,dp}(1-\rho\phi)dp_t+v_{r,t+1} \end{aligned} \quad (36)$$

$\alpha_{g,dp}$ and $\alpha_{\mu,dp}$ control whether dividend yields reveal expected returns or dividend growth, with the familiar $(1-\rho\phi)$ coefficient. With $\alpha_{\mu,dp}=1$, we recover the simple one-state variable model; with $\alpha_{g,dp}=1$ we recover the constant expected return benchmark.

Second, with $\alpha_{\mu,dp}=1$ again but now allow $\alpha_d\neq 0$, i.e. ex-post dividend growth shocks do affect expected dividend growth and expected returns, then dividend yields drop from the dividend forecast regression. To forecast dividends, we have only the standard dividend growth ARMA(1,1) that emerges from an AR(1) expected process, which features a long moving average of extra $\Delta d$ on the right hand side.

$$\frac{1-\phi L}{1-\left(\phi-\frac{\alpha_d}{k}\right)L}\Delta d_{t+1}=v_{d,t+1}$$

$$\left\{1-\frac{\frac{\alpha_d}{k_g}L}{1-\left(\phi-\frac{\alpha_d}{k_g}\right)L}\right\}\Delta d_{t+1} = v_{d,t+1}$$

$$\Delta d_{t+1} = \frac{\alpha_d}{k_g}\sum_{j=0}^{\infty}\left(\phi-\frac{\alpha_d}{k_g}\right)^j \Delta d_{t-j}+v_{d,t+1}$$

We can recover some information about the slow time-varying mean dividend growth by looking at long moving averages of actual dividend growth.

$$r_{t+1}=\frac{1}{k}dp_t+\frac{\alpha_d}{k}\sum_{j=0}^{\infty}\left(\phi-\frac{\alpha_d}{k_g}\right)^j \Delta d_{t-j}+v_{r,t+1}$$

We have the same dividend yield coefficient $1/k = 1-\rho\phi$. However, dividend growth shocks move expected returns and expected dividend growth in offsetting ways, so there is more information about expected returns available than in the dividend yield. *The same long moving average of dividend growth that helps to forecast dividend growth also helps to forecast returns. This observation summarizes most of the point of the AR(1) latent variable model.* Nicely, this representation simply adds two intuitive forecasters: the dividend yield and the moving average used with simple univariate ARMA(1,1) models. As with dividends, standard regressions might miss the fact that all dividends together help to forecast returns.

Finally, we can understand the general case with $\phi_\mu = \phi_d$, and both $\alpha_{\mu,dp} \neq 1$ and $\alpha_d \neq 0$. Now dividend forecasts reduce to we have

$$\Delta d_{t+1}=\frac{1-\alpha_{\mu,dp}}{k}dp_t+\frac{\alpha_d}{k}\sum_{j=0}^{\infty}\left(\phi-\frac{\alpha_d}{k}\right)^j\left(\Delta d_{t-j}-\frac{1-\alpha_{\mu,dp}}{k}dp_{t-j-1}\right)+v_{d,t+1}$$

We can think of this as the "basic" regression (36) with coefficient $\frac{1-\alpha_{\mu,dp}}{k}$, with the addition of a long moving average of "errors' from this regression as well to account for the $\alpha_d$ effect. Typically, the variation in ex-post dividend growth is much larger than its expected value, thus we may expect to a good approximation,

$$\Delta d_{t+1}\approx\frac{1-\alpha_{\mu,dp}}{k}dp_t+\frac{\alpha_d}{k}\sum_{j=0}^{\infty}\left(\phi-\frac{\alpha_d}{k}\right)^j\Delta d_{t-j}+v_{d,t+1}$$

Now, we have the usual coefficient plus a long moving average of past dividend growth.

Return forecasts (35) in this case reduce similarly to

$$r_{t+1}=\frac{\alpha_{\mu,dp}}{k}dp_t-\frac{\alpha_d}{k}\sum_{j=0}^{\infty}\left(\phi-\frac{\alpha_d}{k}\right)^j\left(\Delta d_{t-j}-\frac{1-\alpha_{\mu,dp}}{k}dp_{t-j-1}\right)+v_{r,t+1}$$

As always, what is added to the dividend growth forecast, given dividend yields, must be subtracted from the return forecasts. It too will be well approximated by

$$r_{t+1}\approx\frac{\alpha_{\mu,dp}}{k}dp_t-\frac{\alpha_d}{k}\sum_{j=0}^{\infty}\left(\phi-\frac{\alpha_d}{k}\right)^j\Delta d_{t-j}+v_{r,t+1}$$

The last two equations are a particularly nice representation, as they add together the two intuitive effects. The dividend yield helps to forecast dividend growth, by the usual present value logic. But the dividend yield also forecasts returns. Thus, a direct measure of expected dividend growth formed by a moving average helps as well.

These expressions show also how to obtain estimates of the structural model from regressions. $\phi$ follows from the dividend yield autocorrelation $\alpha_{\mu,dp}$ follows from the coefficient of dividend growth on dividend yields, and $\alpha_d$ follows from the regression coefficient of dividend yields on a long moving average of dividend growth.

Finally, this expression suggests the power of the state-space formulation. VAR research evaluates the individual significance of right hand variables, and could easily conclude that $\Delta d_{t-j}$ are individually insignificant, where the moving average of dividend growth might be both economically and statistically significant. Regression methods are very bad at detecting long-run phenomena of this sort.

*AR and ARMA representations for returns on the right hand side*

We may want to use $\{dp, r\}$ as the basis for the VAR rather than $\{dp, \Delta d\}$. In particular, we may want to forecast returns with dividend yields and lagged returns, rather than with dividend yields and lagged dividend growth. Going through the same steps again, we can find these representations simply by replacing $\alpha_d$ with $\alpha_r$, $v_d$ with $v_r$, and $\Delta d$ with $r$. Thus, the last general result becomes

$$r_{t+1} = \frac{\alpha_{\mu,dp}}{k} dp_t - \frac{\alpha_r}{k} \sum_{j=0}^{\infty} \left( \phi - \frac{\alpha_r}{k} \right)^j \left( r_{t-j} - \frac{\alpha_{\mu,dp}}{k} dp_{t-j-1} \right) + v_{r,t+1}$$

$$r_{t+1} \approx \frac{\alpha_{\mu,dp}}{k} dp_t - \frac{\alpha_r}{k} \sum_{j=0}^{\infty} \left( \phi - \frac{\alpha_r}{k} \right)^j r_{t-j} + v_{r,t+1} \tag{37}$$

Again, the regressions allow us to identify $\phi$ from dividend yield autocorrelation, $\alpha_{\mu,dp}$ from the dividend yield coefficient, and then $\alpha_r$ from the coefficient on a long moving average. Now a long moving average of past return "errors", which will be well approximated by past returns, should help to forecast returns. Again, this is the central message of the AR(1) latent variable model.

In both cases, we see the separate roles of $\alpha_{\mu,dp}$ and $\alpha_d$ or $\alpha_r$, which govern the underlying sensitivity of expected dividend growth to the dividend yield and to dividend growth or return shocks. The "cleanup" variables formed with a long moving averages only help if expected dividend growth (and hence return) variation is linked to the ex-post shocks $v_d$ and $v_r$, if $\alpha_d$ or $\alpha_r$ are not zero. If expected dividend growth or return is only affected by $dp$ shocks, $\alpha_{g,d} = 0$ or $\alpha_{g,r} = 0$, the original $dp$ regressions remain unchanged, and there is nothing more one can do to forecast than pure regressions on dividend yields, even when expected dividend growth does vary.

Regressions of this form can give us good preliminary estimates of $\alpha_{\mu,dp}$ and $\alpha_d$ or $\alpha_r$ to see how much the $\{dp, \Delta d\}$ or $\{dp, r\}$ systems really are affected by expected dividend growth variation. We can identify $\phi \approx 0.94$ from the dividend-yield autoregression, so proxies for the extra right hand variables are straightforward to construct.

*$\phi_\mu \neq \phi_g$ effects*

The case $\phi_\mu \neq \phi_g$ is interesting as well. There are two reasons why extra variables can help to forecast returns, given the dividend yield. Either they can forecast dividend growth in exactly the opposite way, or they can forecast different time-paths for dividend yields and dividend growth. $\phi_\mu \neq \phi_g$ captures the second sort of effect.

To see the effect without the complexity of the general case, consider the $\alpha_d = 0$ special case. Now the ARMA representation is

$$\begin{bmatrix} \left(1-\phi_\mu L\right)\left(1-\phi_g L\right) & 0 \\ 0 & \left(1-\phi_g L\right) \end{bmatrix} \begin{bmatrix} dp_t \\ \Delta d_t \end{bmatrix} = \left( I - \begin{bmatrix} \phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu - \phi_g\right) & 0 \\ \frac{1-\alpha_{\mu,dp}}{k_g} & \phi_g \end{bmatrix} L \right) \begin{bmatrix} v_{dp,t} \\ v_{d,t} \end{bmatrix}$$

The dividend yield follows an autonomous process, but more complex than an AR(1). It's an ARMA(2,1) with nearly -canceling roots, again the sort of process that is hard to capture with regressions.

$$\frac{\left(1-\phi_\mu L\right)\left(1-\phi_g L\right)}{1-\left[\phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)\right]L} dp_t = v_{dp,t}.$$

The dividend growth forecast is now

$$\begin{aligned}
\Delta d_{t+1} &= -\frac{1-\alpha_{\mu,dp}}{k_g}\frac{1}{1-\phi_g L} v_{dp_t} + v_{d,t+1} \\
&= -\frac{1-\alpha_{\mu,dp}}{k_g}\frac{1-\phi_\mu L}{1-\left[\phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)\right]L} dp_t + v_{d,t+1} \\
&= -\frac{1-\alpha_{\mu,dp}}{k_g}\left(1 - \frac{\alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)L}{1-\left[\phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)\right]L}\right) dp_t + v_{d,t+1} \\
&= -\frac{1-\alpha_{\mu,dp}}{k_g}\left(1 - \frac{\alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)L}{1-\left[\alpha_{g,dp}\phi_\mu - \alpha_{\mu,dp}\phi_\mu\right]L}\right) dp_t + v_{d,t+1}
\end{aligned}$$

$$\Delta d_{t+1} = -\frac{\alpha_{g,dp}}{k_g} dp_t + \frac{\alpha_{g,dp}\alpha_{\mu,dp}}{k_g}\left(\phi_\mu-\phi_g\right)\sum_{j=0}^{\infty}\left[\alpha_{g,dp}\phi_\mu - \alpha_{\mu,dp}\phi_\mu\right]^j dp_{t-j} + v_{d,t+1} \tag{38}$$

and the return forecast is correspondingly

$$\begin{aligned}
r_{t+1} &= \hat{\mu}_t + v_{r,t+1} \\
&= \frac{1}{1-\phi_\mu L}\left[\frac{\alpha_{\mu,dp}}{k_\mu} v_{dp,t} + \frac{0}{k_\mu} v_{d,t}\right] + v_{r,t+1} \\
&= \frac{1}{1-\phi_\mu L}\left[\frac{\alpha_{\mu,dp}}{k_\mu}\frac{\left(1-\phi_\mu L\right)\left(1-\phi_g L\right)}{1-\left[\phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)\right]L} dp_t\right] + v_{r,t+1} \\
&= \frac{\alpha_{\mu,dp}}{k_\mu}\frac{1-\phi_g L}{1-\left[\phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)\right]L} dp_t + v_{r,t+1} \\
&= \frac{\alpha_{\mu,dp}}{k_\mu}\left[1 - \frac{\left(1-\alpha_{\mu,dp}\right)\left(\phi_\mu-\phi_g\right)L}{1-\left[\phi_\mu - \alpha_{\mu,dp}\left(\phi_\mu-\phi_g\right)\right]L}\right] dp_t + v_{r,t+1}
\end{aligned}$$

$$r_{t+1} = \frac{\alpha_{\mu,dp}}{k_\mu} dp_t - \frac{\alpha_{\mu,dp}\alpha_{g,dp}}{k_\mu}\left(\phi_\mu-\phi_g\right)\sum_{j=0}^{\infty}\left[\alpha_{g,dp}\phi_\mu - \alpha_{\mu,dp}\phi_\mu\right]^j dp_{t-j} + v_{r,t+1} \tag{39}$$

As usual, we start with the dividend yield with its standard coefficient reflecting how a dividend yield shock is split between expected returns and expected dividend growth. Then, we get an extra term, reflecting different speeds of expected return and expected dividend growth adjustment. This time, a long moving average of dividend yields themselves provide the extra information. Since the dividend yield reflects influences with different horizons, the history of dividend yields can be used to tease out more information about expected returns than is present in the current dividend yield.

# 3 Dividend yield as a state variable, and an observable

## 3.1 Reexpressing the structural model

I have expressed the state-space model with expected returns and expected dividend growth $\mu$ and $g$ as state variables. Many other equivalent representations are possible. We can take any two of $\{\mu_t, g_t, dp_t\}$ as the state variables, and two of $\{\varepsilon_g, \varepsilon_\mu, \varepsilon_{dp}\}$ together with one of $\{\varepsilon_d, \varepsilon_r\}$ as the shocks.

In particular, when thinking about observation, and with the goal of generalizing the benchmark from regressions with only dp on the right hand side that expected returns vary but expected dividend growth does not, $var(g) = 0$, in mind, it is convenient to choose $dp$ as one state variable (since it is directly observable) and choose $g$ as the other. This expression quickly gives a useful representation.

Substituting the dividend yield for expected returns, we can rewrite the example system (16)-(22) as

$$g_{t+1} = \phi_g g_t + \varepsilon_{g,t+1} \tag{40}$$

$$dp_{t+1} = \phi_\mu dp_t + \left(\phi_\mu - \phi_g\right) k_g g_t + \varepsilon_{dp,t+1} \tag{41}$$

$$\Delta d_{t+1} = g_t + \varepsilon_{d,t+1} \tag{42}$$

Now we find $\mu_t$ and $r_t$ after the fact from identities

$$\mu_t = \frac{1}{k_\mu} dp_t + \frac{k_g}{k_\mu} g_t$$

$$r_{t+1} = \frac{1}{k_\mu} dp_t + \frac{k_g}{k_\mu} g_t + \left(\varepsilon_{d,t+1} - \rho \varepsilon_{dp,t+1}\right). \tag{43}$$

In the last equation we start to see how dividend-yield return-forecasting regressions may be modified.

## 3.2 Adding an observable

One direction of generalization is easy to accommodate at the same time. Other variables may help to forecast returns, by helping us to form better estimates of the hidden state variables $\mu_t$ and $g_t$. It is easy to extend this simple example by adding vector of variables $z_t$ which are related to expected returns and expected dividend growth, but with potentially serially correlated errors

$$z_{t+1} = \beta_g g_t + \beta_\mu \mu_t + \phi_z z_t + \varepsilon_{z,t+1}. \tag{44}$$

We can also think of this specification as

$$z_t = \tilde{\beta}_g g_t + \tilde{\beta}_\mu \mu_t + \phi_z z_{t-1} + \tilde{\varepsilon}_{z,t},$$

where I use tildes because the coefficients are not the same. In this latter version, $z_t$ wanders in a stationary but serially correlated way around the an underlying linear combination of $\mu_t$ and $g_t$.

Since agents see $\mu_t$ and $g_t$, this specification does not alter our formulas for dividend yields or returns in the "structural" representations; we just add this equation to our specification of the model. The $z$ variables matter of course to us who must make imperfect estimates of $\mu$ and $g$.

In the context of the model reexpressed with $dp, g$ as state variables (40)-(43), we can use $dp_t = k_\mu \mu_t - k_g g_t$ to write

$$\begin{aligned} z_{t+1} &= \beta_g g_t + \beta_\mu \left( \frac{1}{k_\mu} dp_t + \frac{k_g}{k_\mu} g_t \right) + \phi_z z_t + \varepsilon_{z,t+1} \\ z_{t+1} &= \frac{\beta_\mu}{k_\mu} dp_t + \left( \beta_g + \beta_\mu \frac{k_g}{k_\mu} \right) g_t + \phi_z z_t + \varepsilon_{z,t+1} \\ z_{t+1} &= \gamma_{dp} dp_t + \gamma_g g_t + \phi_z z_t + \varepsilon_{z,t+1} \end{aligned} \tag{45}$$

This is still a restrictive and largely pedagogical example. In general, of course, we will want to allow $\mu$ and $g$ to respond to $z$, writing

$$\begin{bmatrix} \mu_t \\ g_t \\ z_t \end{bmatrix} = A(L) \begin{bmatrix} \varepsilon_{\mu,t} \\ \varepsilon_{g,t} \\ \varepsilon_{z,tt} \\ \varepsilon_{d,t} \end{bmatrix}$$

In this case dividend yield and return formulas are affected by the presence of the $z$ variable, since $z$ helps agents to forecast future expected returns and dividend growth. The form of $z$ in (44) is just a way to extend the simple example slightly, not to purchase complete generality. I consider this general model below.

### 3.3 Observable state-space representation

As before, we can write the observable counterpart to the structural model (40)-(43) with the same state-space representation, but driven by shocks of the observables relative to their history. This construction follows simply by projecting every part of the linear structural model onto the observed information set.

Denote the information set $I_t = \{dp_t, r_t, \Delta d_t, z_t, \quad dp_{t-1}, r_{t-1}, \Delta d_{t-1}, z_{t-1}, ... \}$ From the identity (10), of course, one of $r$ or $\Delta d$ is superfluous. For example, we might summarize what we can learn about the data by a (potentially infinite order) VAR involving $\{dp_t, \Delta d_t, z_t\}$.

The observable errors are

$$v_{x,t+1} = x_{t+1} - E(x_{t+1} | I_t)$$

with this new larger information set.

Define as before our best guesses of the hidden state variables as

$$\begin{aligned} \hat{\mu}_t &\equiv E(r_{t+1}|I_t) = E(\mu_t|I_t) \\ \hat{g}_t &\equiv E(\Delta d_{t+1}|I_t) = E(g_t|I_t). \end{aligned}$$

Then the observable model corresponding to (40)-(42) and (45) is

$$\hat{g}_{t+1} = \phi_g \hat{g}_t + (\alpha_{g,dp} v_{dp,t+1} + \alpha_{g,d} v_{d,t+1} + \alpha_{g,z} v_{z,t+1}) \tag{46}$$

$$dp_{t+1} = \phi_\mu dp_t + \left(\phi_\mu - \phi_g\right) k_g \hat{g}_t + v_{dp,t+1} \tag{47}$$

$$\Delta d_{t+1} = \hat{g}_t + v_{d,t+1} \tag{48}$$

$$z_{t+1} = \gamma_{dp} dp_t + \gamma_g \hat{g}_t + \phi_z z_t + v_{z,t+1} \tag{49}$$

Returns and expected returns $\hat{\mu}$ follow by identities, in particular.

$$\begin{aligned} r_{t+1} &= \frac{1}{k_\mu} dp_t + \frac{k_g}{k_\mu} \hat{g}_t + v_{r,t+1}; \\ v_{r,t+1} &= v_{d,t+1} - \rho v_{dp,t+1} \end{aligned} \tag{50}$$

Once again, the shock to our best guess of dividend growth $\hat{g}$ must be a combination of the observable shocks as shown in (46). Again, if one wants to derive $\alpha$ coefficients from a particular view of the structure, it is a straightforward if messy job of matching second moments. Absent theoretical restrictions on the covariance matrix of shocks (and they are still absent), the $v$ are again as well as we can do to describe the system dynamics.

$z_t$ is observable, but its forecasting variable $g_t$ is not. This difference in right hand variable ($g$ vs. $\hat{g}$) means that the $z$ regression error $v_{z,t+1}$ is different from the structural shock $\varepsilon_{z,t+1}$, and it means the AR representation of the system, presented below, leads to a more complex forecasting equation for $z$ than in the structural model.

In the special case $\alpha_{g,d} = \alpha_{g,dp} = \alpha_{g,z} = 0$, expected dividend growth (given our information set $I_t$) is constant, and (46)-(50) reduce to the familiar one-state-variable system

$$\begin{aligned} \Delta d_{t+1} &= v_{d,t+1} \\ dp_{t+1} &= \phi_\mu dp_t + v_{dp,t+1} \\ r_{t+1} &= \left(1 - \rho\phi_\mu\right) dp_t + (v_{d,t+1} - \rho v_{dp,t+1}) \end{aligned}$$

Though $z$ coevolves with expected returns, the dividend yield reveals all we need to know about expected returns in this case.

As expected dividend growth varies more, as the $\alpha$ increase, we begin to have a new state variable $\hat{g}_t$ that forecasts returns in addition to the dp ratio, seen in (50). In this representation, the coefficient of returns on the dividend yield is unchanged, but greater variation in expected dividend growth allows greater variation in expected returns. We can think of $\hat{g}_t$ here as cleaning up the dividend yield, removing some of the movement in dividend yields caused by dividend growth forecasts. We can write the return forecasting equation as

$$r_{t+1} = (1 - \rho\phi_\mu) \left[ dp_t + E\left( \sum_{j=1}^{\infty} \rho^{j-1} g_t | I_t \right) \right] + v_{r,t+1}.$$

The additional state variable $\hat{g}_t$ is visible in (46) as an accumulation of errors,

$$\hat{g}_t = \sum_{j=0}^{\infty} \phi_g^j \left( \alpha_{g,dp} v_{dp,t-j} + \alpha_{g,d} v_{d,t-j} + \alpha_{g,z} v_{z,t-j} \right),$$

or we can simply view (46) as a filtering formula. The AR(1) structure of latent variables means that the history of dividend growth and dividend yields is summarized by this one additional latent variable for the purposes of forecasting returns.

The special case $\phi_\mu = \phi_g$ is an interesting simplification. Then, (47) and (50) reduce to

$$dp_{t+1} = \phi_\mu dp_t + v_{dpt+1}$$

$$r_{t+1} = (1 - \rho\phi_\mu) dp_t + \hat{g}_t + v_{r,t+1}. \tag{51}$$

The complication in the dp transition equation (47) came from the fact that dp is the sum of two AR(1)s with different coefficients (expected returns and expected dividend growth). Now that it is the sum of two AR(1)'s with the same coefficient, it retains the simple AR(1) structure of the model with no variation in expected dividend growth. The return forecasting equation now adds just the current expected dividend growth. Equivalently, since $k_g = k_\mu$, the response of dividend yield to expected return is the same as its response to expected dividend growth.

### 3.3.1 Other choices for state variables and shocks.

As with the structural system, we can rewrite the observable system in many ways by exploiting the identities

$$dp_t = k_\mu \hat{\mu}_t - k_g \hat{g}_t \tag{52}$$

$$v_{dpt+1} = k_\mu v_{\mu t+1} - k_g v_{gt+1} \tag{53}$$

$$v_{rt+1} = -\rho v_{dpt+1} + v_{dt+1} \tag{54}$$

We can choose any two of the three state variables $\{\hat{g}_t, \hat{\mu}_t, dp_t\}$ and the other is implied by the identity (52). We can choose to represent the system in terms of *one* the three state variable shocks $\{v_{dp}, v_\mu, v_g\}$, and one of $\{v_r, v_d\}$, the other following from (54). The difference relative to the "structural" model is that we can now only observe one state variable shock rather than two. As in (46), one of the state variable shocks must be expressed in terms of the other shocks to the system. In the end, the observable implications come down to those of a $\{dp, \Delta d, z\}$ or $\{dp, r, z\}$ VAR.

## 3.4 VAR and ARMA representations, forecasting regressions

Reexpressing the state-space model in terms of $\hat{g}, dp$ as state variables rather than $\hat{g}, \hat{\mu}$ does not change the VAR and ARMA representations, which substitute out all of the unobserved state variables.

Adding $z$ variables does change the VAR and ARMA representations. Empirically, the history of dividend growth and returns don't add as much to return forecasts as additional variables do.

Thus, though algebraically more complex, this is the important case to consider. (The algebra is in the Appendix.)

$$\begin{aligned}
\Delta d_{t+1} &= \alpha_{g,dp} dp_t + \alpha_{g,z} z_t + \alpha_{g,d} V_{d,t} + \alpha_{g,z} V_{z,t} + v_{d,t+1} \\
z_{t+1} &= \left(\alpha_{g,dp}\gamma_g + \gamma_{dp}\right) dp_t + \left(\alpha_{g,z}\gamma_g + \phi_z\right) z_t + \alpha_{g,d}\gamma_g V_{d,t} + \alpha_{g,z}\gamma_g V_{z,t} + v_{z,t+1}
\end{aligned}$$

where

$$\begin{aligned}
V_{d,t} &\equiv \frac{1}{1-\tilde{\phi}L} \left(\Delta d_t - \alpha_{g,dp} dp_{t-1} - \alpha_{g,z} z_{t-1}\right) \\
V_{z,t} &\equiv \frac{1}{1-\tilde{\phi}L} \left[\phi z_{t-1} - \left(\alpha_{g,z}\gamma_g + \phi_z\right) z_{t-1} - \left(\alpha_{g,dp}\gamma_g + \gamma_{dp}\right) dp_{t-1}\right]
\end{aligned}$$

$$\tilde{\phi} \equiv \phi - \alpha_{g,d} - \alpha_{g,z}\gamma_g.$$

The return AR representation is most convenient with reference to the $\{dp, r\}$ VAR,

$$\begin{aligned}
r_{t+1} &= \left(1 - \rho\phi + \alpha_{g,dp}\right) dp_t + \alpha_{g,z} z_t + \alpha_{g,r} V_{r,t} + \alpha_{g,z} V_{z,t} + v_{r,t+1} \\
z_{t+1} &= \left(\gamma_{dp} + \gamma_g \alpha_{g,dp}\right) dp_t + \left(\phi_z + \gamma_g \alpha_{g,z}\right) z_t + \gamma_g \alpha_{g,r} V_{r,t} + \gamma_g \alpha_{g,z} V_{z,t} + v_{z,t+1}
\end{aligned}$$

where

$$\begin{aligned}
V_{r,t} &\equiv \frac{1}{1-\tilde{\phi}L} \left[r_t - \left((1-\rho\phi) + \alpha_{g,dp}\right) dp_{t-1} - \alpha_{g,z} z_{t-1}\right] \\
V_{z,t} &\equiv \frac{1}{1-\tilde{\phi}L} \left[\phi z_{t-1} - \left(\phi_z + \alpha_{g,z}\gamma_g\right) z_{t-1} - \left(\alpha_{g,dp}\gamma_g + \gamma_{dp}\right) dp_{t-1}\right] \\
\tilde{\phi} &\equiv \phi - \alpha_{g,r} - \alpha_{g,z}\gamma_g.
\end{aligned}$$

We have a natural extension of the intuition from previous formulas. Now dividend yields $dp_t$ as well as the state variable $z_t$ should help to forecast returns and dividend growth. Again "cleanup" variables formed by long moving averages also help. Interestingly, "cleanup" variables help for $z$ as well, since $z$ dynamics are influenced by the unobserved $\hat{g}$.

*Interesting special cases*

The formulas are also interesting to see what forces are at work, and to see which effects one might suppose to work do not in fact show up.

First, suppose $\alpha_{g,z} = 0$. Expected dividend growth still moves around, and $z$ still gives us independent information about the hidden state variable $\hat{g}_t$, but *shocks* to $g$ are unaffected by shocks to $z$. (The shocks $v$ are all arbitrarily correlated, so "unaffected" here means in a multiple regression sense; shocks $v_g$ are spanned by shocks $v_{dp}$ and $v_d$)

Now our formulas specialize to

$$\begin{aligned}
\Delta d_{t+1} &= \alpha_{g,dp} dp_t + \alpha_{g,d} V_{d,t} + v_{d,t+1} \\
z_{t+1} &= \left(\alpha_{g,dp}\gamma_g + \gamma_{dp}\right) dp_t + \phi_z z_t + \alpha_{g,d}\gamma_g V_{d,t} + v_{z,t+1}
\end{aligned}$$

where

$$V_{d,t} \equiv \frac{1}{1-\tilde{\phi}L}\left(\Delta d_t - \alpha_{g,dp} dp_{t-1}\right)$$

and

$$\begin{aligned} r_{t+1} &= \left(1-\rho\phi+\alpha_{g,dp}\right) dp_t + \alpha_{g,r} V_{r,t} + v_{r,t+1} \\ z_{t+1} &= \left(\gamma_{dp}+\gamma_g \alpha_{g,dp}\right) dp_t + \phi_z z_t + \gamma_g \alpha_{g,r} V_{r,t} + v_{z,t+1} \end{aligned}$$

where

$$V_{r,t} \equiv \frac{1}{1-\tilde{\phi}L}\left[r_t - \left(\left(1-\rho\phi\right)+\alpha_{g,dp}\right) dp_{t-1}\right].$$

If $\alpha_{g,z}=0$ – if expected dividend growth $\hat{g}_t$ is not affected by *shocks* to the information variable $z_t$, then $z$ disappears entirely from forecasting returns and dividend growth (and, as always dividend yields).

### 3.5 Calibration to dp-only forecasts

Naturally, we can estimate the multivariate return and dividend growth forecasts specified by the two-state variable models. However, we can also ask what are their implications for *single* regressions that only include dividend yields. Yes, these regressions are not as efficient as they could be, as they ignore information that could improve forecasts. But they are regressions, and a successful two-state model must mimic what happens if one runs the "wrong" regression as well as what happens if one runs the "right" regressions. The advantage of this preliminary step is that we know a good deal about simple dividend-yield regressions. As it turns out, the simple facts about dividend yield forecasts in (6)-(9) allow us to restrict the two-state-variable model considerably.

To generate $b_d = 0$ – the fact that dividend yields alone do not predict dividend growth – we must have either $var(g) = 0$ (of course) or

$$\beta_{k_\mu \mu, k_g g} \equiv \frac{cov(k_g g_t, k_\mu \mu_t)}{var(k_g g_t)} = 1. \tag{55}$$

(Derivations below.) A regression of the the "permanent component" of expected returns $k_\mu \mu_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$ on the permanent component of expected dividend growth must be one. Intuitively, a rise in expected dividend growth must come with a corresponding rise in expected returns on average, so that the rise in expected dividend growth does not raise the price-dividend ratio.

In terms of more fundamental parameters of the model, we have a similar constraint on the regression coefficient of the expected return shock on the expected dividend growth shock,

$$\beta_{\varepsilon_\mu, \varepsilon_g} \equiv \frac{cov(\varepsilon_\mu, \varepsilon_g)}{var(\varepsilon_g)} = \frac{1-\rho\phi_\mu}{1-\rho\phi_g} \frac{1-\phi_g \phi_\mu}{1-\phi_g^2} \tag{56}$$

The terms on the right hand side of (56) correct for the potentially different persistence of expected return and expected dividend growth shocks. In the useful baseline $\phi_\mu = \phi_g$, we have $\beta_{\varepsilon_\mu, \varepsilon_g} = 1$ as well, with the same intuition as for (55).

$b_r = 0.1$ and $b_{dp} = 0.94$ – the two items are the same with the identity $b_r = 1 + b_d - \rho b_{dp}$and $b_d = 0$, $\rho \approx 0.96$ – imply that

$$1 - \rho\phi_\mu = b_r \tag{57}$$

and

$$\phi_\mu = b_{dp} \approx 0.94 \tag{58}$$

This is the same formula that obtains when $var(g_t) = 0$. Thus, even in the presence of dividend growth forecastability, and even though the price-dividend ratio does not reveal expected returns directly, the fact that $b_d = 0$ implies that we can recover the *autocorrelation* coefficient of expected returns $\phi_\mu$ – the structural parameter in (17) – from the univariate autocorrelation of the dividend-price ratio $b_{dp}$.

Finally, the covariance matrix of the shocks in (6)-(9) restricts the covariance matrix of the shocks in the model (16)-(18). There are only two independent shocks in the regressions (6)-(9), which we might as well take as the uncorrelated dividend yield and dividend growth shocks. The regression shocks are related to the structural shocks by

$$v_{dp,t+1} = k_g \left(\phi_\mu - \phi_g\right) g_t + \varepsilon_{dp,t+1} \tag{59}$$

$$v_{d,t+1} = g_t + \varepsilon_{d,t+1} \tag{60}$$

In the $\phi_\mu = \phi_g$ special case, $g_t$ drops from the $dp$ transition equation (41), so the regression is the structural equation and the regression error measures the structural error. In general, as usual, regression and structural errors are not the same. (59)-(60) and $cov(v_d, v_{dp}) = 0$ then imply

$$\begin{aligned} var(\varepsilon_d) &= var\left(v_d\right) - var\left(g\right) && (61)\\ var(\varepsilon_{dp}) &= var(v_{dp}) - \left(\phi_\mu - \phi_g\right)^2 k_g^2 var(g) && (62)\\ cov(\varepsilon_{dp}, \varepsilon_d) &= -\left(\phi_\mu - \phi_g\right) k_g var(g) && (63) \end{aligned}$$

It's easiest to describe the parameters of the model in the form (40)-(42), but of course other representations and the variance and covariance of other shocks $\varepsilon_r, \varepsilon_\mu$follow.)

In sum, we may stay consistent with the evidence of simple dividend-yield regressions (6)-(9) with a wide range of $g$ processes. We may choose the persistence $\phi_g$ and the properties of the shock $\varepsilon_{gt}$ – its variance and correlation with the other shocks $\varepsilon_{dp}, \varepsilon_d$ (or $\varepsilon_\mu, \varepsilon_d$, or $\varepsilon_\mu, \varepsilon_r$). However, once that is done, the remaining parameters of the model are pinned down by the simple dividend-yield regressions. We identify the $\phi_\mu \approx 0.94$ structural parameter from dividend yield autocorrelation. As we increase $var(\varepsilon_g)$ and hence $var(g_t)$, we have to choose $var(\varepsilon_d)$, $var(\varepsilon_{dp})$ and $cov(\varepsilon_d, \varepsilon_{dp})$ as indicated by (61)-(63). The surprise is not that there is *some* information to add from other sources; the surprise is that *so much* of the model structure is pinned down already by dividend yield regressions.

### 3.5.1 Derivations

$b_d = 0$ means

$$\begin{aligned} 0 &= cov(\Delta d_{t+1}, dp_t)\\ &= cov(g_t + \varepsilon_{d,t+1}, k_\mu \mu_t - k_g g_t) = cov(g_t, k_\mu \mu_t - k_g g_t)\\ &= k_\mu cov(g_t,\, \mu_t) - k_g var(g_t)\\ &= cov(k_g g_t,\, k_\mu \mu_t) - var(k_g g_t), \end{aligned}$$

establishing (55). Translating to shocks,

$$1=\beta_{k_\mu\mu,k_g g}=\frac{k_\mu\frac{1}{1-\phi_g\phi_\mu}\sigma_{\varepsilon_g,\varepsilon_\mu}}{k_g^2\frac{1}{1-\phi_g^2}\sigma^2_{\varepsilon_g}}=\frac{1-\rho\phi_g}{1-\rho\phi_\mu}\frac{1-\phi_g^2}{1-\phi_g\phi_\mu}\frac{\sigma_{\varepsilon_g,\varepsilon_\mu}}{\sigma^2_{\varepsilon_g}}=\frac{1-\rho\phi_g}{1-\rho\phi_\mu}\frac{1-\phi_g^2}{1-\phi_g\phi_\mu}\beta_{\varepsilon_\mu,\varepsilon_g},$$

establishing (56).

To find the model prediction for $b_r$,

$$\begin{aligned}
b_r &= \frac{cov(r_{t+1},dp_t)}{var(dp_t)}=\frac{cov(\mu_t,k_\mu\mu_t-k_g g_t)}{var(k_\mu\mu_t-k_g g_t)}=\left(1-\rho\phi_\mu\right)\frac{cov(k_\mu\mu_t,k_\mu\mu_t-k_g g_t)}{var(k_\mu\mu_t-k_g g_t)}\\
&= \left(1-\rho\phi_\mu\right)\frac{cov(k_\mu\mu_t,k_\mu\mu_t-k_g g_t)}{cov(k_\mu\mu_t,k_\mu\mu_t-k_g g_t)-cov(k_g g_t,k_\mu\mu_t-k_g g_t)}
\end{aligned}$$

Using $b_d=0$ and hence $cov(g_t,k_\mu\mu_t-k_g g_t)=0$, we obtain

$$b_r=(1-\rho\phi_\mu)$$

and hence (57).

To derive (59) and (60), write the definition of regression errors,

$$\begin{aligned}
v_{dp,t+1} &= dp_{t+1}-b_{dp}dp_t=dp_{t+1}-\phi_\mu dp_t\\
&= k_\mu\phi_\mu\mu_t-k_g\phi_g g_t+\varepsilon_{dp,t+1}-\phi_\mu\left(k_\mu\mu_t-k_g g_t\right)\\
&= k_g\left(\phi_\mu-\phi_g\right)g_t+\varepsilon_{dp,t+1}
\end{aligned}$$

$$v_{d,t+1}=\Delta d_{t+1}-b_d dp_t=\Delta d_{t+1}=g_t+\varepsilon_{d,t+1}$$

### 3.5.2 Implications for long-run returns

What do regressions look like at longer horizons? Suppose you run

$$r_{t+2}=a_r+b_r^{(2)}dp_t+v_{t+2}$$

in data generated from the two-state model. How does $b_r$ vary with horizon? The one state variable model

$$b_r^{(2)}=b_r(1+\phi)$$

seems to work well in the data. Are there other patterns we should see?

## 4 A General Two-State Variable Model

In the end, we want to rewrite an arbitrary VAR into a state-space form, to find its implications for expected returns and dividend growth. We want to translate VAR specification assumptions to state-space specification assumptions and vice-versa. For that application, I need to generalize this system to allow for dynamics beyond a univariate AR(1), to allow for additional variables $z_t$ which coevolve with expected returns and dividend growth in arbitrary ways. In addition, we want to understand more generally the relationship between restrictions in the "structural" state-space model, such as the AR(1) of the simple example, and restrictions on VAR representations, such as finite lags, the presence or absence of variables, and relations between coefficients such as the long moving average form of regression coefficients in the simple example.

## 4.1 "Structural" representation

The general form of the state-space model is

$$\begin{bmatrix} \mu_t \\ g_t \\ z_t \end{bmatrix} = A(L) \begin{bmatrix} \varepsilon_{\mu,t} \\ \varepsilon_{g,t} \\ \varepsilon_{z,t} \\ \varepsilon_{d,t} \end{bmatrix} \tag{64}$$

$$\Delta d_{t+1} = g_t + \varepsilon_{d,t+1}. \tag{65}$$

(I write moving averages for technical convenience, but the equivalent autoregressive representations are often more intuitive as in (16)-(18).) Relative to the simple examples, this specification allows all three state variables $(\mu, g, z)$ to help forecast the others, and it allows them all to react to ex-post dividend growth shocks. Dividend yields follow[1] by

$$dp_t = E_t \sum_{j=0}^{\infty} \rho^j \left( \mu_{t+j} - g_{t+j} \right) = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \frac{LA(L) - \rho A(\rho)}{L - \rho} \begin{bmatrix} \varepsilon_{\mu,t} \\ \varepsilon_{g,t} \\ \varepsilon_{z,t} \\ \varepsilon_{d,t} \end{bmatrix} \tag{66}$$

and dividend yield and return shocks follow; evaluating (66) at $L = 0$,

$$\varepsilon_{dp,t+1} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} A(\rho) \begin{bmatrix} \varepsilon_{\mu,t} \\ \varepsilon_{g,t} \\ \varepsilon_{z,t} \\ \varepsilon_{d,t} \end{bmatrix} \tag{67}$$

$$\varepsilon_{r,t+1} = -\rho \varepsilon_{dp,t+1} + \varepsilon_{d,t+1} \tag{68}$$

## 4.2 Observable state-space representation

The state-space representation of the observable implications of the general model (64)-(68) consists of exactly the same model, projected on the information set $I_t$, and recognizing the loss of one shock. It is simply

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \\ z_t \end{bmatrix} = A(L) \begin{bmatrix} v_{\mu,t} \\ v_{g,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix} \tag{69}$$

$$\Delta d_{t+1} = \hat{g}_t + v_{d,t+1} \tag{70}$$

[1] Here I use the magic Hansen-Sargent (1980) formula, that if

$$x_t = a(L)\varepsilon_t$$

then

$$E_t \sum_{j=0}^{\infty} \rho^j x_{t+j} = \frac{La(L) - \rho a(\rho)}{L - \rho} \varepsilon_t$$

To prove it, just write out the moving average representations of both sides.

dividend yields follow by

$$dp_t = E_t \sum_{j=0}^{\infty} \rho^j \left( \hat{\mu}_{t+j} - \hat{g}_{t+j} \right) = \left[ \begin{array}{ccc} 1 & -1 & 0 \end{array} \right] \frac{LA(L) - \rho A(\rho)}{L - \rho} \left[ \begin{array}{c} v_{\mu,t} \\ v_{g,t} \\ v_{z,t} \\ v_{d,t} \end{array} \right] \tag{71}$$

and the dividend yield and return shocks follow by

$$\begin{aligned} v_{dp,t+1} &= \left[ \begin{array}{ccc} 1 & -1 & 0 \end{array} \right] A(\rho) \left[ \begin{array}{c} v_{\mu,t+1} \\ v_{g,t+1} \\ v_{z,t+1} \\ v_{d,t+1} \end{array} \right] \\ v_{r,t+1} &= -\rho v_{dp,t+1} + v_{d,t+1} \end{aligned} \tag{72}$$

In the structure (66)-(68), there are four potentially independent shocks, $\varepsilon_\mu, \varepsilon_g, \varepsilon_z, \varepsilon_d$, and we complete the structural specification with an arbitrary $4 \times 4$ shock covariance matrix. Our observable representation is equivalent to a VAR (Wold representation) of the observable variables $\{dp, \Delta d, z\}$ or $\{dp, r, z\}$ so it is driven by only three shocks, which must be linear combinations of the VAR shocks $\{v_{dp}, v_d, v_z\}$ or $\{v_{dp}, v_r, v_z\}$. Choosing the $\{dp, \Delta d, z\}$ representation, we must be able to write shocks to the observable model in the form

$$\left[ \begin{array}{c} v_{\mu,t} \\ v_{g,t} \end{array} \right] = \left[ \begin{array}{ccc} \alpha_{\mu,dp} & \alpha_{\mu,z} & \alpha_{\mu,d} \\ \alpha_{g,dp} & \alpha_{g,z} & \alpha_{g,d} \end{array} \right] \left[ \begin{array}{c} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{array} \right]. \tag{73}$$

Only one row of the $\alpha$ matrix is free, as we must have the counterpart of (67),

$$v_{dp,t} = \left[ \begin{array}{ccc} 1 & -1 & 0 \end{array} \right] A(\rho) \left[ \begin{array}{c} v_{\mu,t} \\ v_{g,t} \\ v_{z,t} \\ v_{d,t} \end{array} \right] = \left[ \begin{array}{ccc} 1 & -1 & 0 \end{array} \right] A(\rho) \left[ \begin{array}{ccc} \alpha_{\mu,dp} & \alpha_{\mu,z} & \alpha_{\mu,d} \\ \alpha_{g,dp} & \alpha_{g,z} & \alpha_{g,d} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] \left[ \begin{array}{c} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{array} \right]$$

and hence

$$\left[ \begin{array}{ccc} 1 & 0 & 0 \end{array} \right] = \left[ \begin{array}{cc} 1 & -1 \end{array} \right] \left[ \begin{array}{cccc} A_{\mu\mu}(\rho) & A_{\mu g}(\rho) & A_{\mu z}(\rho) & A_{\mu d}(\rho) \\ A_{g\mu}(\rho) & A_{gg}(\rho) & A_{gz}(\rho) & A_{gd}(\rho) \end{array} \right] \left[ \begin{array}{ccc} \alpha_{\mu,dp} & \alpha_{\mu,z} & \alpha_{\mu,d} \\ \alpha_{g,dp} & \alpha_{g,z} & \alpha_{g,d} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] \tag{74}$$

For example, as in (27) we can express this restriction as a set of restrictions that let us solve for the $\alpha_\mu$ given the $\alpha_g$ or vice-versa

$$\begin{aligned} 1 &= \left[ A_{\mu\mu}(\rho) - A_{g\mu}(\rho) \right] \alpha_{\mu,dp} + \left[ A_{\mu g}(\rho) - A_{gg}(\rho) \right] \alpha_{g,dp} \\ 0 &= \left[ A_{\mu\mu}(\rho) - A_{g\mu}(\rho) \right] \alpha_{\mu,z} + \left[ A_{\mu g}(\rho) - A_{gg}(\rho) \right] \alpha_{g,z} + \left[ A_{\mu z}(\rho) - A_{gz}(\rho) \right] \\ 0 &= \left[ A_{\mu\mu}(\rho) - A_{g\mu}(\rho) \right] \alpha_{\mu,d} + \left[ A_{\mu g}(\rho) - A_{gg}(\rho) \right] \alpha_{g,d} + \left[ A_{\mu d}(\rho) - Ad(\rho) \right] \end{aligned}$$

This is the general version of the restriction in (27)-(28).

### 4.2.1 Model expressed in terms of observable shocks

With the substitution (73), we can equivalently rewrite the model in terms of the observable shocks directly

$$\begin{aligned} \begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \\ z_t \end{bmatrix} &= \tilde{A}(L) \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix} \\ \Delta d_{t+1} &= \hat{g}_t + v_{d,t+1} \end{aligned} \tag{75}$$

The only (but crucial) difference in the form of this representation is that we have the observable shocks directly on the right hand side rather than indirectly via $v_g$ and $v_\mu$. I use a tilde on $\tilde{A}$ to distinguish it from the original $A$ matrix.

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \\ z_t \end{bmatrix} = A(L) \begin{bmatrix} v_{\mu,t} \\ v_{g,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix} = A(L) \begin{bmatrix} \alpha_{\mu,dp} & \alpha_{\mu,z} & \alpha_{\mu,d} \\ \alpha_{g,dp} & \alpha_{g,z} & \alpha_{g,d} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix} = \tilde{A}(L) \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix}$$

Dividend yields now follow by

$$dp_t = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \frac{L\tilde{A}(L) - \rho\tilde{A}(\rho)}{L-\rho} \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix} \tag{76}$$

and the dividend yield and return shocks follow by

$$\begin{aligned} v_{dp,t+1} &= \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \tilde{A}(\rho) \begin{bmatrix} v_{dp,t+1} \\ v_{z,t+1} \\ v_{d,t+1} \end{bmatrix} \\ v_{r,t+1} &= -\rho v_{dp,t+1} + v_{d,t+1} \end{aligned} \tag{77}$$

By construction,

$$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \tilde{A}(\rho)$$

so we recover the same dividend yield shock we started with. Hence we can also write

$$dp_t = \frac{L\left[\tilde{A}_\mu(L) - \tilde{A}_g(L)\right] - \rho \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}{L-\rho} \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix}$$

where $\tilde{A}_\mu$ denotes the $\mu$ row of $\tilde{A}$.

### 4.2.2 From VAR to state-space

Above, we asked, given the restrictions of a "structural" representation, what restrictions on the VAR representation follow? In this case the inverse question is just as interesting: Given a VAR representation (78), how do we construct the state-space representation (69)? This inverse problem is the interesting problem for empirical work. When we run regressions, we want to interpret them through the eyes of a state-space model like (69). In this case the "structural" model is specified loosely enough that any VAR can be interpreted through its eyes.

This exercise is actually quite easy. Any set of return and dividend growth forecast regressions implies a model for expected returns and expected dividend growth, as the right hand side of the regressions. We just have to track the dynamics of the right hand sides of two regressions, and we have the dynamics for the state variables $\hat{\mu}$ and $\hat{g}$. Since the rest follows from identities, we can reverse the logic and know that starting with those dynamics produces the dividend yield and return as the "result."

Thus, suppose we start with a set of forecasting regressions of returns and dividend growth on $dp$, $\Delta d$ (or $r$) and $z$,

$$\begin{bmatrix} dp_{t+1} \\ z_{t+1} \\ \Delta d_{t+1} \\ r_{t+1} \end{bmatrix} = a(L) \begin{bmatrix} dp_t \\ z_t \\ \Delta d_t \end{bmatrix} + \begin{bmatrix} v_{dp,t+1} \\ v_{z,t+1} \\ v_{d,t+1} \\ v_{r,t+1} \end{bmatrix}. \qquad (78)$$

As usual, one equation is irrelevant; the return or dividend growth row can be inferred from the others, and we can use $r_t$ on the right hand side in place of $\Delta d_t$. Our objective is to translate between this VAR representation and the equivalent observable state-space representation of the form (64)-(68).

Using $\{dp, z, \Delta d\}$ as state variables, our regressions imply the moving average representation

$$\begin{bmatrix} dp_t \\ z_t \\ \Delta d_t \end{bmatrix} = [I - La^*(L)]^{-1} \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix}$$

where $a^*(L)$ denotes the $dp$, $z$ and $\Delta d$ (first three) rows of $a(L)$ defined in (78). Since $\hat{\mu}_t = E(r_{t+1}|I_t)$, $\hat{g}_t = E(\Delta d_{t+1}|I_t)$ we can then find the state variables

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{r,dp}(L) & a_{r,z}(L) & a_{r,d}(L) \\ a_{d,dp}(L) & a_{r,z}(L) & a_{r,d}(L) \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} dp_t \\ z_t \\ \Delta d_t \end{bmatrix}$$

and so write the moving average representation

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{r,dp}(L) & a_{r,z}(L) & a_{r,d}(L) \\ a_{d,dp}(L) & a_{r,z}(L) & a_{r,d}(L) \\ 0 & 1 & 0 \end{bmatrix} [I - La^*(L)]^{-1} \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix}. \qquad (79)$$

Of course we have directly

$$\Delta d_{t+1} = \hat{g}_t + v_{d,t+1}.$$

*That's it.* We have a model of the form (75). We can say "suppose expected returns and dividend growth are generated by this process, suppose we find dividend yields by the present value formula (76) and returns from the identity linking returns to dividend yields and dividend growth." We will recover the return regressions that we started with. (It's a lot of fun to go through the construction of $dp_t, r_t$ and verify you get back to where you started. )

Alas, *this equivalence is bad news for the state-space project* if its goal is to improve return forecasts. We can construct a state-space model to correspond to *any* VAR. For example, we can derive a state-space representation of a Campbell-Shiller (1988) two-lag VAR. State-space models do not *have* to produce regressions with long moving averages on the right hand side. That finding is a result of one particular state-space model, and that example was chosen for its pedagogically simple structure. We have no economic reason to impose the particular structure of the simple example rather than the time-series structure of expected returns and dividend growth that results from an arbitrary VAR. The only hope is that one might make a calculation of this sort and find that a sensible VAR implies an implausible state-space specification, for example with sawtooth impulse response functions. Otherwise, state-space models will be a useful interpretation step rather than any particular help in forecasting.

### 4.3 Expression in terms of expected return and expected dividend growth shocks

As in the state-space model, it's interesting to express the shocks as a "shock to expected returns" and another "shock to expected dividend growth," rather than in terms of the more nebulous "shock to the dividend yield," i.e. in the form of (69) rather than (75).

Evaluating (79) at $L = 0$, we have

$$\begin{bmatrix} v_{\mu,t} \\ v_{g,t} \end{bmatrix} = \begin{bmatrix} a_{r,dp}(0) & a_{r,z}(0) & a_{r,d}(0) \\ a_{d,dp}(0) & a_{d,z}(0) & a_{d,d}(0) \end{bmatrix} \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix} \tag{80}$$

Thus, the first regression coefficients $a(0)$ are the alpha parameters of (73), which shows us how to identify and estimate the $\alpha$ parameters.

We want to rewrite (79),

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{r,dp}(L) & a_{r,z}(L) & a_{r,d}(L) \\ a_{d,dp}(L) & a_{r,z}(L) & a_{r,d}(L) \\ 0 & 1 & 0 \end{bmatrix} [I - La^*(L)]^{-1} \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix}$$

In terms of $\mu$ and $g$ shocks so defined. This means finding a matrix $Q$ that inverts the transformation in (80), i.e. such that

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{r,dp}(L) & a_{r,z}(L) & a_{r,d}(L) \\ a_{d,dp}(L) & a_{r,z}(L) & a_{r,d}(L) \\ 0 & 1 & 0 \end{bmatrix} [I - La^*(L)]^{-1} Q \begin{bmatrix} v_{\mu,t} \\ v_{g,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix} \tag{81}$$

To be such an inverse, we must have

$$I_3 = \begin{bmatrix} Q_{dp,\mu} & Q_{dp,g} & Q_{dp,z} & Q_{dp,d} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{r,dp}(0) & a_{r,z}(0) & a_{r,d}(0) \\ a_{d,dp}(0) & a_{d,z}(0) & a_{d,d}(0) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

This condition for inverse puts three restrictions on the four free elements of $Q$.

$$\begin{bmatrix} 1 = Q_{dp,g}a_{d,dp}(0) + Q_{dp,\mu}a_{r,dp}(0) \\ 0 = Q_{dp,g}a_{d,z}(0) + Q_{dp,\mu}a_{r,z}(0) + Q_{dp,z} \\ 0 = Q_{dp,g}a_{d,d}(0) + Q_{dp,\mu}a_{r,d}(0) + Q_{dp,d} \end{bmatrix}$$

We may choose either $Q_{dp,g}$ or $Q_{dp,\mu}$; the other follows from the first condition, and $Q_{dp,z}$ and $Q_{dp,d}$ follow from those two.

To understand this arbitrariness consider the case that the coefficients $a_{.,z}, a_{.,d} = 0$, i.e. of contemporaneous variables only $dp_t$ forecasts returns. Then, we must find shocks to expected returns and expected dividend growth as

$$\begin{bmatrix} v_{\mu,t} \\ v_{g,t} \end{bmatrix} = \begin{bmatrix} a_{r,dp}(0) \\ a_{d,dp}(0) \end{bmatrix} \begin{bmatrix} v_{dp,t} \end{bmatrix}$$

Since there is only one underlying shock, however, how we express the resulting dynamics is somewhat arbitrary. We can express the exactly the same dynamics as driven by $v_{\mu,t}/a_{r,dp}(0)$, or by $v_{g,t}/a_{d,dp}(0)$. (Of course in both cases if one of $a_{r,dp}(0)$ or $a_{d,dp}(0)$ is zero, then the arbitrariness is removed.)

In sum, we know how to construct a representation of the form (69), but it is not in general uniquely identified, as must be the case since (69) has a singular covariance matrix.

The same feature means that we cannot in general express the representation (69) that we recover with impact multipliers that are all one, i.e. $A(0)$ with one on the diagonal and zeros elsewhere. To achieve that result would require the opposite inverse relation

$$I_4 = \begin{bmatrix} a_{r,dp}(0) & a_{r,z}(0) & a_{r,d}(0) \\ a_{d,dp}(0) & a_{d,z}(0) & a_{d,d}(0) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Q_{dp,\mu} & Q_{dp,g} & Q_{dp,z} & Q_{dp,d} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$I_4 = \begin{bmatrix} a_{r,dp}(0)Q_{dp,\mu} & Q_{dp,g}a_{r,dp}(0) & Q_{dp,z}a_{r,dp}(0) & Q_{dp,d}a_{r,dp}(0) \\ a_{d,dp}(0)Q_{dp,\mu} & Q_{dp,g}a_{d,dp}(0) & a_{d,dp}Q_{dp,z}(0) & Q_{dp,d}a_{d,dp}(0) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

which in general is not possible. We cannot choose $Q_{dp,\mu}$ to make the top left element one *and* to make the 1,2 element zero.

## 4.4 From state space to VAR

The state-space model is

$$\begin{aligned} \begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \\ z_t \end{bmatrix} &= \tilde{A}(L) \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix} \\ \Delta d_{t+1} &= \hat{g}_t + v_{d,t+1} \end{aligned} \tag{82}$$

$$dp_t = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \frac{L\tilde{A}(L) - \rho\tilde{A}(\rho)}{L-\rho} \begin{bmatrix} v_{dp,t} \\ v_{z,t} \\ v_{d,t} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \tilde{A}(\rho)$$

Again we find MA and VAR representations for the observables $\{dp_t, \Delta d_t, r_t, z_t\}$ by eliminating the latent variables $\hat{\mu}$, $\hat{g}$.

*Moving average.* $dp_t$ and $z_t$ are already in moving average representation. We find dividends and returns simply by (82) and its return counterpart. Thus, the moving average representation is

$$\begin{bmatrix} dp_{t+1} \\ z_{t+1} \\ \Delta d_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{matrix} \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \frac{L\tilde{A}(L)-\rho\tilde{A}(\rho)}{L-\rho} \\ \tilde{A}_z(L) \\ L\tilde{A}_g(L) + \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \\ L\tilde{A}_\mu(L) + \begin{bmatrix} -\rho & 0 & 1 \end{bmatrix} \end{matrix} \begin{bmatrix} v_{dp,t+1} \\ v_{z,t+1} \\ v_{d,t+1} \end{bmatrix}$$

Where I use $\tilde{A}_g$, etc. to denote the rows of $\tilde{A}$.

### 4.5 Example

Suppose we have a general first-order VAR, with no z variable

$$\begin{bmatrix} dp_{t+1} \\ \Delta d_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{bmatrix} a_{dp,dp} & a_{dp,d} \\ a_{d,dp} & a_{d,d} \\ a_{r,dp} & a_{r,d} \end{bmatrix} \begin{bmatrix} dp_t \\ \Delta d_t \end{bmatrix} + \begin{bmatrix} v_{dp,t+1} \\ v_{d,t+1} \\ v_{r,t+1} \end{bmatrix}.$$

As usual the return identity means that one row is redundant

$$\begin{aligned} a_{r,dp} &= -\rho a_{dp,dp} + a_{d,dp} \\ a_{r,d} &= -\rho a_{dp,d} + a_{d,d} \end{aligned}$$

Equation (79) gives us the observable state-space representation

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \end{bmatrix} = \begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix} \left[ I - \begin{bmatrix} a_{dp,dp} & a_{dp,d} \\ a_{d,dp} & a_{d,d} \end{bmatrix} L \right]^{-1} \begin{bmatrix} v_{dp,t} \\ v_{d,t} \end{bmatrix}. \tag{83}$$

This result is more intuitive in AR form.

$$\left[ I - \begin{bmatrix} a_{dp,dp} & a_{dp,d} \\ a_{d,dp} & a_{d,d} \end{bmatrix} L \right] \begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \end{bmatrix} = \begin{bmatrix} v_{dp,t} \\ v_{d,t} \end{bmatrix}.$$

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \end{bmatrix} = \begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix} \begin{bmatrix} a_{dp,dp} & a_{dp,d} \\ a_{d,dp} & a_{d,d} \end{bmatrix} \begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mu}_{t-1} \\ \hat{g}_{t-1} \end{bmatrix} + \begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix} \begin{bmatrix} v_{dp,t} \\ v_{d,t} \end{bmatrix}.$$

Using the identities

$$\begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix} = \begin{bmatrix} -\rho a_{dp,dp} + a_{d,dp} & -\rho a_{dp,d} + a_{d,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix} = \begin{bmatrix} -\rho & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{dp,dp} & a_{dp,d} \\ a_{d,dp} & a_{d,d} \end{bmatrix},$$

and hence

$$\begin{bmatrix} -\frac{1}{\rho} & \frac{1}{\rho} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix} = \begin{bmatrix} a_{dp,dp} & a_{dp,d} \\ a_{d,dp} & a_{d,d} \end{bmatrix},$$

we can simplify the AR form to

$$\begin{bmatrix} \hat{\mu}_t \\ \hat{g}_t \end{bmatrix} = \begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix} \begin{bmatrix} -\frac{1}{\rho} & \frac{1}{\rho} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mu}_{t-1} \\ \hat{g}_{t-1} \end{bmatrix} + \begin{bmatrix} a_{r,dp} & a_{r,d} \\ a_{d,dp} & a_{r,d} \end{bmatrix} \begin{bmatrix} v_{dp,t} \\ v_{d,t} \end{bmatrix}$$

and finally

$$\left[\begin{array}{c}\hat{\mu}_t\\ \hat{g}_t\end{array}\right]=\frac{1}{\rho}\left[\begin{array}{cc}-a_{r,dp} & a_{r,dp}+\rho a_{r,d}\\ -a_{d,dp} & a_{d,dp}+\rho a_{r,d}\end{array}\right]\left[\begin{array}{c}\hat{\mu}_{t-1}\\ \hat{g}_{t-1}\end{array}\right]+\left[\begin{array}{cc}a_{r,dp} & a_{r,d}\\ a_{d,dp} & a_{r,d}\end{array}\right]\left[\begin{array}{c}v_{dp,t}\\ v_{d,t}\end{array}\right] \tag{84}$$

We recognize a variation of the original simple example, which specified a diagonal transition matrix.

From (80) we can define expected return and dividend shocks

$$\left[\begin{array}{c}v_{\mu,t}\\ v_{g,t}\end{array}\right]=\left[\begin{array}{cc}a_{r,dp} & a_{r,d}\\ a_{d,dp} & a_{d,d}\end{array}\right]\left[\begin{array}{c}v_{dp,t}\\ v_{d,t}\end{array}\right]$$

and thus write simply

$$\left[\begin{array}{c}\hat{\mu}_t\\ \hat{g}_t\end{array}\right]=\frac{1}{\rho}\left[\begin{array}{cc}-a_{r,dp} & a_{r,dp}+\rho a_{r,d}\\ -a_{d,dp} & a_{d,dp}+\rho a_{r,d}\end{array}\right]\left[\begin{array}{c}\hat{\mu}_{t-1}\\ \hat{g}_{t-1}\end{array}\right]+\left[\begin{array}{c}v_{\mu,t}\\ v_{g,t}\end{array}\right]$$

This example has a sad lesson – all of the interesting moving average dynamics studied in the context of the simple example comes from ruling out off-diagonal elements of the $\mu, g$ transition matrix. Here is an example of almost the same form, whose AR representation is simply a first-order VAR.

# 5 References


van Binsbergen, Jules H. and Ralph S. J. Koijen, 2008, "Predictive Regressions: A Present-Value Approach" http://ssrn.com/abstract=968053

Campbell, John Y. and Robert J. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *The Review of Financial Studies,* 1, 195-228.

Campbell, John Y. and John Ammer, 1993, "What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns," *The Journal of Finance* 48, 3-37.

Campbell, John Y. and John H. Cochrane, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior" *Journal of Political Economy* 107, 205-251.

Cochrane, John H., 2008, "The Dog That Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies* forthcoming (RFS Advance Access September 2007)

Cochrane, John H., 1994, "Permanent and Transitory Components of GNP and Stock Prices" *Quarterly Journal of Economics* 109, 241-266

Cochrane, John H., 1992, "Explaining the Variance of Price-Dividend Ratios" *Review of Financial Studies* 5, 243-280.

Cochrane, John H., 2004, *Asset Pricing*, Revised Edition Princeton NJ: Princeton University Press.