Online appendix for "Maximum likelihood estimation of the equity premium" *

Efstathios Avdis
University of Alberta

Jessica A. Wachter
University of Pennsylvania
and NBER

September 11, 2016

*Avdis: avdis@ualberta.ca; Wachter: jwachter@wharton.upenn.edu.

A. Derivation of the maximum likelihood estimators

A.1. *Benchmark*

We denote the maximum likelihood estimate of parameter q as $\hat{q}$. Here we derive the estimators for μ_r, μ_x, β, θ, σ_u^2, σ_v^2 and σ_{uv}. We note in particular that $\hat{\sigma}_u^2$ is the estimator of σ_u^2, not the square of the estimator of σ_u, and similarly for $\hat{\sigma}_v^2$. Maximizing the exact log likelihood function is the same as minimizing the function $\mathcal{L}$:

$$\begin{aligned}\mathcal{L}(\beta,\theta,\mu_r,\mu_x,\sigma_{uv},\sigma_u,\sigma_v) &= \log(\sigma_v^2) - \log(1-\theta^2) + \frac{1-\theta^2}{\sigma_v^2}(x_0-\mu_x)^2 \\ &+ T\log(|\Sigma|) + \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t^2 - 2\frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} u_t v_t + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T} v_t^2, \quad \text{(A.1)}\end{aligned}$$

where $|\Sigma| = \sigma_u^2\sigma_v^2 - \sigma_{uv}^2$. The function $\mathcal{L}$ is -2 times the logarithm of the likelihood function (6) modulo constants. The first-order conditions arise from setting the following partial derivatives of $\mathcal{L}$ to zero:

$$0 = \frac{\partial}{\partial\beta}\mathcal{L} = 2\left[\frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t(\mu_x - x_{t-1}) - \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T}(\mu_x - x_{t-1})v_t\right] \quad \text{(A.2a)}$$

$$\begin{aligned}0 = \frac{\partial}{\partial\theta}\mathcal{L} = 2\Bigg[&\frac{\theta}{1-\theta^2} - \theta\frac{(x_0-\mu_x)^2}{\sigma_v^2} \\ &- \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} u_t(\mu_x - x_{t-1}) + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T} v_t(\mu_x - x_{t-1})\Bigg]\end{aligned} \quad \text{(A.2b)}$$

$$0 = \frac{\partial}{\partial\mu_r}\mathcal{L} = 2\left[-\frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t + \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} v_t\right] \quad \text{(A.2c)}$$

$$0 = \frac{\partial}{\partial\mu_x}\mathcal{L} = 2\Bigg[-\frac{1-\theta^2}{\sigma_v^2}(x_0-\mu_x) + \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T}\beta u_t$$

$$- \frac{\sigma_{uv}}{|\Sigma|} \sum_{t=1}^{T} (\beta v_t - (1-\theta) u_t) - \frac{\sigma_u^2}{|\Sigma|} \sum_{t=1}^{T} (1-\theta) v_t \Bigg] \tag{A.2d}$$

$$0 = \frac{\partial}{\partial \sigma_{uv}} \mathcal{L} = -T \frac{2\sigma_{uv}}{|\Sigma|} + 2 \frac{\sigma_{uv}\sigma_v^2}{|\Sigma|^2} \sum_{t=1}^{T} u_t^2 - 2 \frac{\sigma_u^2 \sigma_v^2 + \sigma_{uv}^2}{|\Sigma|^2} \sum_{t=1}^{T} u_t v_t + 2 \frac{\sigma_{uv}\sigma_u^2}{|\Sigma|^2} \sum_{t=1}^{T} v_t^2 \tag{A.2e}$$

$$0 = \frac{\partial}{\partial \sigma_u^2} \mathcal{L} = T \frac{\sigma_v^2}{|\Sigma|} - \frac{\sigma_v^4}{|\Sigma|^2} \sum_{t=1}^{T} u_t^2 + 2 \frac{\sigma_{uv}\sigma_v^2}{|\Sigma|^2} \sum_{t=1}^{T} u_t v_t - \frac{\sigma_{uv}^2}{|\Sigma|^2} \sum_{t=1}^{T} v_t^2 \tag{A.2f}$$

$$0 = \frac{\partial}{\partial \sigma_v^2} \mathcal{L} = \frac{1}{\sigma_v^2} + T \frac{\sigma_u^2}{|\Sigma|} - (1-\theta^2)(x_0 - \mu_x)^2 \frac{1}{\sigma_v^4} - \frac{\sigma_{uv}^2}{|\Sigma|^2} \sum_{t=1}^{T} u_t^2 + 2 \frac{\sigma_{uv}\sigma_u^2}{|\Sigma|^2} \sum_{t=1}^{T} u_t v_t - \frac{\sigma_u^4}{|\Sigma|^2} \sum_{t=1}^{T} v_t^2. \tag{A.2g}$$

Define the residuals

$$\hat{u}_t = r_t - \hat{\mu}_r - \hat{\beta}(x_{t-1} - \hat{\mu}_x) \tag{A.3a}$$

$$\hat{v}_t = x_t - \hat{\mu}_x - \hat{\theta}(x_{t-1} - \hat{\mu}_x). \tag{A.3b}$$

We now outline the algebra that allows us to solve these first-order conditions.

Step 1: Express $\hat{\mu}_x$ in terms of $\hat{\theta}$ and the data.

Combining the first-order conditions (A.2c) and (A.2d) gives

$$\sum_{t=1}^{T} \hat{v}_t = \left(1 + \hat{\theta}\right) \left(\hat{\mu}_x - x_0\right), \tag{A.4}$$

which we can write as

$$\hat{\mu}_x = \frac{\left(1 + \hat{\theta}\right) x_0 + \sum_{t=1}^{T} \left(x_t - \hat{\theta} x_{t-1}\right)}{\left(1 + \hat{\theta}\right) + \left(1 - \hat{\theta}\right) T}. \tag{A.5}$$

Step 2: Express the covariance matrix in terms of $\hat{\mu}_x$, $\hat{\theta}$, $\hat{\mu}_r$, $\hat{\beta}$ and the data.

The first-order conditions (A.2e), (A.2f) and (A.2g) give the relations

$$T\hat{\sigma}_u^2 = -\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\hat{\sigma}_{uv} + (1-\hat{\theta}^2)(x_0 - \hat{\mu}_x)^2 \left(\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\right)^2 + \sum_{t=1}^{T} \hat{u}_t^2, \tag{A.6}$$

$$(T+1)\hat{\sigma}_v^2 = (1-\hat{\theta}^2)(x_0 - \hat{\mu}_x)^2 + \sum_{t=1}^{T} \hat{v}_t^2, \tag{A.7}$$

$$\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2} = \frac{\sum_{t=1}^{T} \hat{u}_t \hat{v}_t}{\sum_{t=1}^{T} \hat{v}_t^2}. \tag{A.8}$$

Step 3: Solve for $\hat{\theta}$ in terms of the data. This also gives $\hat{\mu}_x$ and $\hat{\sigma}_v^2$ in terms of the data.

Combining the first-order conditions (A.2a) and (A.2b) gives

$$0 = \sum_{t=1}^{T}(\hat{\mu}_x - x_{t-1})\hat{v}_t + \hat{\sigma}_v^2 \frac{\hat{\theta}}{1-\hat{\theta}^2} - \hat{\theta}(x_0 - \hat{\mu}_x)^2. \tag{A.9}$$

Here $\hat{\mu}_x$ and $\hat{v}_t$ are functions of only $\hat{\theta}$ and the data, so if we combine (A.27) and (A.7) we can get an equation for $\hat{\theta}$:

$$0 = (T+1)\sum_{t=1}^{T}(\hat{\mu}_x - x_{t-1})\hat{v}_t + \frac{\hat{\theta}}{1-\hat{\theta}^2}\sum_{t=1}^{T} \hat{v}_t^2 - T\hat{\theta}(x_0 - \hat{\mu}_x)^2. \tag{A.10}$$

Because we require that $-1 < \hat{\theta} < 1$, we can multiply this by

$$\left((T+1) - (T-1)\hat{\theta}\right)^2 \left(1 - \hat{\theta}^2\right) \tag{A.11}$$

and rearrange to obtain

$$0 = T\left(\hat{\theta} - 1\right)\left((T+1)\left(1-\hat{\theta}^2\right) + 2\hat{\theta}\right)\left(\sum_{t=0}^{T} x_t - \hat{\theta}\sum_{t=1}^{T-1} x_t\right)^2$$
$$+ \left((T+1) - (T-1)\hat{\theta}\right)\left(\hat{\theta} - 1\right)\left(\sum_{t=0}^{T} x_t - \hat{\theta}\sum_{t=1}^{T-1} x_t\right)$$

$$
\times\left[2T\hat{\theta}(1+\hat{\theta})\left(\sum_{t=1}^{T-1}x_t\right)-\left((T+1)+(T-1)\hat{\theta}\right)\left(\sum_{t=0}^{T}x_t+\sum_{t=1}^{T-1}x_t\right)\right]
$$

$$
+\left((T+1)-(T-1)\hat{\theta}\right)^2
$$

$$
\times\left[\hat{\theta}\left(\left(1-\hat{\theta}^2\right)T+1\right)\left(\sum_{t=1}^{T-1}x_t^2\right)+\left(\hat{\theta}^2(T-1)-(T+1)\right)\sum_{t=1}^{T}x_tx_{t-1}+\hat{\theta}\sum_{t=0}^{T}x_t^2\right]. \tag{A.12}
$$

This is a fifth-order polynomial in $\hat{\theta}$ where the coefficients are determined by the sample. As a consequence, it is very hard to establish analytical results on existence and uniqueness of solutions that would be accepted as estimators of θ. Nevertheless, in lengthy experimentation and simulation runs we have always found that this polynomial only has one root within the unit circle of the complex plane and that this root is real. Therefore this root is a valid MLE of θ. Given this solution for $\hat{\theta}$, (A.5) gives the estimator for μ_x and (A.7) gives the estimator for σ_v^2.

Step 4: Solve for $\hat{\mu}_r$ and $\hat{\beta}$ in terms of the data. This also gives the solution for $\hat{\sigma}_{uv}$ and $\hat{\sigma}_u^2$.

The first-order condition (A.2c) gives

$$
\sum_{t=1}^{T}\hat{u}_t=\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\sum_{t=1}^{T}\hat{v}_t. \tag{A.13}
$$

Combining this with the first-order condition (A.2a) yields

$$
\hat{\beta}=\beta^{\text{OLS}}+\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}\left(\hat{\theta}-\theta^{\text{OLS}}\right), \tag{A.14}
$$

where

$$
\theta^{\text{OLS}}=\frac{1}{\frac{1}{T}\sum_{t=1}^{T}x_{t-1}^2-\left(\frac{1}{T}\sum_{t=1}^{T}x_{t-1}\right)^2}\left[\frac{1}{T}\sum_{t=1}^{T}x_{t-1}x_t-\left(\frac{1}{T}\sum_{t=1}^{T}x_{t-1}\right)\left(\frac{1}{T}\sum_{s=1}^{T}x_s\right)\right] \tag{A.15}
$$

is the OLS coefficient of regressing x_t on x_{t-1} and

$$\beta^{\text{OLS}} = \frac{1}{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2 - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right)^2} \left[\frac{1}{T}\sum_{t=1}^{T} x_{t-1} r_t - \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right)\left(\frac{1}{T}\sum_{s=1}^{T} r_s\right)\right] \tag{A.16}$$

is the OLS coefficient of regressing r_t on x_{t-1}.

Equations (A.8), (A.13) and (A.14) constitute a system of three equations in the three unknowns $\hat{\mu}_r$, $\hat{\beta}$ and $\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2}$. The solution is

$$\begin{aligned}\hat{\mu}_r = \frac{1}{J}\Bigg[&\frac{1}{T}\sum_{t=1}^{T} r_t - \left(\frac{1}{T}\sum_{t=1}^{T} x_t - \hat{\mu}_x\right)\frac{F - \beta^{\text{OLS}} H}{1 + (\hat{\theta} - \theta^{\text{OLS}})H} \\ &- \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1} - \hat{\mu}_x\right)\frac{\beta^{\text{OLS}}(1 + \hat{\theta} H) - \theta^{\text{OLS}} F}{1 + (\hat{\theta} - \theta^{\text{OLS}})H}\Bigg]\end{aligned} \tag{A.17}$$

$$\hat{\beta} = \frac{\beta^{\text{OLS}} + (\hat{\theta} - \theta^{\text{OLS}})F}{1 + (\hat{\theta} - \theta^{\text{OLS}})H} - \frac{(\hat{\theta} - \theta^{\text{OLS}})G}{1 + (\hat{\theta} - \theta^{\text{OLS}})H}\hat{\mu}_r \tag{A.18}$$

$$\frac{\hat{\sigma}_{uv}}{\hat{\sigma}_v^2} = \frac{F - \beta^{\text{OLS}} H}{1 + (\hat{\theta} - \theta^{\text{OLS}})H} - \frac{G}{1 + (\hat{\theta} - \theta^{\text{OLS}})H}\hat{\mu}_r, \tag{A.19}$$

where

$$J = 1 - \frac{G}{1 + (\hat{\theta} - \theta^{\text{OLS}})H}\left[\frac{1}{T}\sum_{t=1}^{T} x_t - \hat{\mu}_x - \theta^{\text{OLS}}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1} - \hat{\mu}_x\right)\right] \tag{A.20a}$$

$$F = \frac{\sum_{t=1}^{T} r_t \hat{v}_t}{\sum_{t=1}^{T} \hat{v}_t^2} \tag{A.20b}$$

$$G = \frac{\sum_{t=1}^{T} \hat{v}_t}{\sum_{t=1}^{T} \hat{v}_t^2} \tag{A.20c}$$

$$H = \frac{\sum_{t=1}^{T} (x_{t-1} - \hat{\mu}_x)\hat{v}_t}{\sum_{t=1}^{T} \hat{v}_t^2}. \tag{A.20d}$$

Expressions (A.17) and (A.18) provide the estimators for μ_r and β because they depend only on the data and $\hat{\mu}_x$ and $\hat{\theta}$, which we have already expressed in terms of the data. Finally, (A.19) gives the estimator the estimator of σ_{uv} via (A.7), which further yields the estimator of σ_u^2 via (A.6).

A.2. Restricted maximum likelihood

We consider maximum likelihood estimation under the restriction $\beta = 0$. We denote the restricted maximum likelihood estimate of parameter q as $\check{q}$. This case turns out to be less tractable than the unrestricted case, and for this reason, we fix the entries of the variance-covariance matrix Σ. We implement the estimator in two stages; in the first stage we run OLS to find Σ under the assumption of $\beta = 0$. In the second stage, we solve the equations that follow.

Consider (A.1) with the restriction of $\beta = 0$. The first-order conditions are as follows:

$$0 = \frac{\partial}{\partial \theta}\mathcal{L} = 2\Bigg[\frac{\theta}{1-\theta^2} - \theta\frac{(x_0 - \mu_x)^2}{\sigma_v^2} - \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} u_t(\mu_x - x_{t-1}) + \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T} v_t(\mu_x - x_{t-1})\Bigg] \tag{A.21a}$$

$$0 = \frac{\partial}{\partial \mu_r}\mathcal{L} = 2\left[-\frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} u_t + \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T} v_t\right] \tag{A.21b}$$

$$0 = \frac{\partial}{\partial \mu_x}\mathcal{L} = 2\Bigg[-\frac{1-\theta^2}{\sigma_v^2}(x_0 - \mu_x) + \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T} \beta u_t - \frac{\sigma_{uv}}{|\Sigma|}\sum_{t=1}^{T}(\beta v_t - (1-\theta)u_t) - \frac{\sigma_u^2}{|\Sigma|}\sum_{t=1}^{T}(1-\theta)v_t\Bigg] \tag{A.21c}$$

Define the residuals

$$\check{u}_t = r_t - \check{\mu}_r \tag{A.22a}$$

$$\check{v}_t = x_t - \check{\mu}_x - \check{\theta}(x_{t-1} - \check{\mu}_x). \tag{A.22b}$$

We now outline the algebra that allows us to solve these first-order conditions.

Step 1: Express $\check{\mu}_x$ and $\check{\mu}_r$ in terms of $\check{\theta}$ and the data.

The first-order condition (A.21b) gives

$$\sum_{t=1}^{T} \check{u}_t = \frac{\sigma_{uv}}{\sigma_v^2} \sum_{t=1}^{T} \check{v}_t. \tag{A.23}$$

Combining this with the first-order condition (A.21c) gives

$$\sum_{t=1}^{T} \check{v}_t = \left(1+\check{\theta}\right) \left(\check{\mu}_x - x_0\right), \tag{A.24}$$

which we can write as

$$\hat{\mu}_x = \frac{\left(1+\check{\theta}\right) x_0 + \sum_{t=1}^{T} \left(x_t - \check{\theta} x_{t-1}\right)}{\left(1+\check{\theta}\right) + \left(1-\check{\theta}\right) T}. \tag{A.25}$$

Combining (A.24) and (A.23) yields

$$\check{\mu}_r = \frac{1}{T} \sum_{t=1}^{T} r_t - \frac{1}{T} \frac{\sigma_{uv}}{\sigma_v^2} \left(1+\check{\theta}\right) \left(\check{\mu}_x - x_0\right). \tag{A.26}$$

Step 2: Solve for $\check{\theta}$ in terms of the data.

Substituting (A.23), (A.24) and (A.26) into the first-order condition (A.21a) gives

$$\begin{aligned} 0 = \sigma_v^2 \frac{\check{\theta}}{1-\check{\theta}^2} - \check{\theta}(x_0 - \check{\mu}_x)^2 + \left(1+\check{\theta}\right) \check{\mu}_x \left(\check{\mu}_x - x_0\right) \\ + \frac{1}{|\Sigma|} \left(\sum_{t=1}^{T} x_{t-1} \right) \left[\frac{\sigma_{uv}^2}{T} \left(1+\check{\theta}\right) \left(\check{\mu}_x - x_0\right) + \sigma_u^2 \sigma_v^2 \left(1-\check{\theta}\right) \check{\mu}_x \right] \\ + \frac{1}{|\Sigma|} \left[\sigma_{uv} \sigma_v^2 \sum_{t=1}^{T} x_{t-1} \left(r_t - \frac{1}{T} \sum_{s=1}^{T} r_s \right) - \sigma_u^2 \sigma_v^2 \sum_{t=1}^{T} x_{t-1} \left(x_t - \check{\theta} x_{t-1} \right) \right] \end{aligned} \tag{A.27}$$

Here $\check{\mu}_x$ is a function of only $\check{\theta}$ and the data, so given σ_u^2, σ_v^2 and σ_{uv} the above an equation for $\check{\theta}$. Similarly to Appendix A, multiplying through by

$$\left((T+1) - (T-1)\check{\theta}\right)^2 \left(1 - \check{\theta}^2\right) \tag{A.28}$$

and carrying out the algebra gives a fifth-order polynomial in $\check{\theta}$ where the coefficients are determined by the sample. As for the exact ML estimator in Appendix A, in lengthy experimentation and simulation runs we have always found that this polynomial only has one root within the unit circle of the complex plane and that this root is real. Therefore this root is a valid MLE of θ. Given this solution for $\check{\theta}$, (A.25) gives the estimator for μ_x and (A.26) gives the estimator for μ_r.

A.3. The multivariate case

Our model is

$$\begin{aligned} r_{t+1} - \mu_r &= \sum_{i=1}^{N} \beta_i (x_{it} - \mu_{xi}) + u_{t+1} \\ x_{1t+1} - \mu_{x1} &= \theta_1 (x_{1t} - \mu_{x1}) + v_{1t+1} \\ &\vdots \\ x_{Nt+1} - \mu_{xN} &= \theta_N (x_{Nt} - \mu_{xN}) + v_{Nt+1} \end{aligned} \tag{A.29}$$

where, with $v_t = (v_{1t}, \dots, v_{Nt})^\top$, the vector $(u_t, v_t^\top)^\top$ is Gaussian and iid over time with covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv}^\top \\ \sigma_{uv} & \Sigma_v \end{bmatrix}. \tag{A.30}$$

Let Σ_x denote the covariance matrix of the vector $x_t = (x_{1t}, \dots, x_{Nt})^\top$. Element (i, j) of matrix Σ_x equals

$$\frac{\sigma_{ij}}{1 - \theta_i \theta_j}, \tag{A.31}$$

where σ_{ij} is element (i, j) of matrix Σ_v. Let μ_x denote the vector $(\mu_{x1}, \dots, \mu_{xN})^\top$, β denote the vector $(\beta_1, \dots, \beta_N)^\top$, θ denote the vector $(\theta_1, \dots, \theta_N)^\top$, and Θ denote the $N \times N$ diagonal matrix with the vector θ as its diagonal.

We denote the maximum likelihood estimate of parameter q as $\breve{q}$. Here we derive the estimators for μ_r, μ_x, β, and θ, taking σ_u^2, Σ_v, and σ_{uv} as given. Maximizing the exact log likelihood function is the same as minimizing the function $\mathcal{L}$:

$$\mathcal{L}(\beta,\theta,\mu_r,\mu_x) = \log|\Sigma_x| + (x_0-\mu_x)^\top\Sigma_x^{-1}(x_0-\mu_x) + T\log(|\Sigma|) + \sum_{t=1}^{T}\begin{pmatrix} u_t & v_t^\top \end{pmatrix}\Sigma^{-1}\begin{pmatrix} u_t \\ v_t \end{pmatrix} \quad \text{(A.32)}$$

where $|Q|$ is notation for the determinant of matrix Q.

Let e_i denote a column vector with one as its ith element and zeros everywhere else. The first-order conditions arise from setting the partial derivatives of the likelihood function to zero.

$$0 = \frac{\partial}{\partial\beta_i}\mathcal{L} \Rightarrow 0 = \frac{\sigma_v^2}{|\Sigma|}\sum_{t=1}^{T}(\mu_x - x_{it-1})\left(u_t - \sigma_{uv}^\top\Sigma_v^{-1}v_t\right) \quad \text{(A.33a)}$$

$$0 = \frac{\partial}{\partial\theta_i}\mathcal{L} \Rightarrow 0 = \operatorname{tr}\left(\Sigma_x^{-1}\frac{\partial}{\partial\theta_i}\Sigma_x\right) - (x_0-\mu_x)^\top\Sigma_x^{-1}\left(\frac{\partial}{\partial\theta_i}\Sigma_x\right)\Sigma_x^{-1}(x_0-\mu_x) + 2\sum_{t=1}^{T}(x_{it-1}-\mu_{xi})e_i^\top\left[\frac{1}{\sigma_\varepsilon^2}\Sigma_v^{-1}\sigma_{uv}u_t - \left(\Sigma_v^{-1} + \frac{1}{\sigma_\varepsilon^2}\Sigma_v^{-1}\sigma_{uv}\sigma_{uv}^\top\Sigma_v^{-1}\right)v_t\right] \quad \text{(A.33b)}$$

$$0 = \frac{\partial}{\partial\mu_r}\mathcal{L} \Rightarrow \sum_{t=1}^{T}u_t = \sigma_{uv}^\top\Sigma_v^{-1}\sum_{t=1}^{T}v_t \quad \text{(A.33c)}$$

$$0 = \frac{\partial}{\partial\mu_{xi}}\mathcal{L} \Rightarrow e_i^\top\Sigma_x^{-1}(x_0-\mu_x) = (\theta_i - 1)\begin{pmatrix} 0 & e_i^\top \end{pmatrix}\Sigma^{-1}\begin{pmatrix} \sum_{t=1}^{T}u_t \\ \sum_{t=1}^{T}v_t \end{pmatrix}, \quad \text{(A.33d)}$$

where

$$\sigma_\varepsilon^2 = \sigma_u^2 - \sigma_{uv}^\top\Sigma_v^{-1}\sigma_{uv}. \quad \text{(A.34)}$$

Define the residuals

$$\breve{u}_t = r_t - \breve{\mu}_r - \breve{\beta}^\top\left(x_{t-1} - \breve{\mu}_x\right) \quad \text{(A.35a)}$$

$$\breve{v}_t = x_t - \breve{\mu}_x - \breve{\Theta}\left(x_{t-1} - \breve{\mu}_x\right). \tag{A.35b}$$

We now outline the algebra that allows us to solve these first-order conditions.

Step 1: Express $\breve{\mu}_x$ in terms of $\breve{\Theta}$ and the data.

Stacking the first-order conditions for μ_{xi} in a vector, we get, after carrying out the algebra,

$$(\breve{\Theta} - \mathbb{I})\Sigma_v^{-1}\left[\sum_{t=1}^{T}\breve{v}_t + \frac{1}{\sigma_\varepsilon^2}\sigma_{uv}\left(\sigma_{uv}^\top\Sigma_v^{-1}\sum_{t=1}^{T}\breve{v}_t - \sum_{t=1}^{T}\breve{u}_t\right)\right] = \Sigma_v^{-1}\left(x_0 - \breve{\mu}_x\right). \tag{A.36}$$

Using (A.33c) we can simplify this to

$$\left(\breve{\Theta} - \mathbb{I}\right)\Sigma_v^{-1}\sum_{t=1}^{T}\breve{v}_t = \breve{\Sigma}_x^{-1}\left(x_0 - \breve{\mu}_x\right), \tag{A.37}$$

where $\breve{\Sigma}_x$ is a matrix with

$$\frac{\sigma_{ij}}{1 - \breve{\theta}_i\breve{\theta}_j} \tag{A.38}$$

as its (i, j)th element. We can write (A.37) as

$$\begin{aligned}\breve{\mu}_x = &\left[\mathbb{I} + T\breve{\Sigma}_x\left(\breve{\Theta} - \mathbb{I}\right)\Sigma_v^{-1}\left(\breve{\Theta} - \mathbb{I}\right)\right]^{-1} \\ &\times\left[x_0 - \breve{\Sigma}_x\left(\breve{\Theta} - \mathbb{I}\right)\Sigma_v^{-1}\left(\sum_{t=1}^{T}x_t - \breve{\Theta}\sum_{t=1}^{T}x_{t-1}\right)\right].\end{aligned} \tag{A.39}$$

Given σ_u^2, Σ_v, and σ_{uv}, this equation expresses $\breve{\mu}_x$ in terms of the data and $\breve{\Theta}$.

Step 2: Solve for $\breve{\theta}$ in terms of the data. This also gives $\breve{\mu}_x$ in terms of the data.

Using (A.33a) in (A.33b) gives

$$0 = \operatorname{tr}\left(\breve{\Sigma}_x^{-1}\frac{\partial}{\partial\theta_i}\breve{\Sigma}_x\right) - (x_0 - \mu_x)^\top\breve{\Sigma}_x^{-1}\left(\frac{\partial}{\partial\theta_i}\breve{\Sigma}_x\right)\breve{\Sigma}_x^{-1}(x_0 - \mu_x)$$

$$- 2e_i^\top \Sigma_v^{-1} \sum_{t=1}^{T} (x_{it-1} - \mu_{xi}) \breve{v}_t, \quad \text{(A.40)}$$

for $i = 1, \ldots, N$. From (A.39) we have $\breve{\mu}_x$ in terms of $\breve{\theta}$ and the data, so if we combine (A.39) and (A.40) we get a system of N nonlinear equations for $\breve{\theta}_1, \ldots, \breve{\theta}_N$. Given the solution of this system for $\breve{\theta}_1, \ldots, \breve{\theta}_N$, (A.39) gives the estimator for μ_x.

Step 3: Solve for $\breve{\mu}_r$ and $\breve{\beta}$ in terms of the data.

The first-order condition (A.33c) gives

$$\breve{\mu}_r = \frac{1}{T} \sum_{t=1}^{T} r_t - \sigma_{uv}^\top \Sigma_v^{-1} \frac{1}{T} \sum_{t=1}^{T} \breve{v}_t - \breve{\beta}^\top \left(\frac{1}{T} \sum_{t=1}^{T} x_{t-1} - \breve{\mu}_x \right). \quad \text{(A.41)}$$

Using this in (A.33a) and carrying out the algebra we get

$$\begin{aligned}
&\left[\frac{1}{T} \sum_{t=1}^{T} x_{it-1} r_t - \left(\frac{1}{T} \sum_{t=1}^{T} x_{it-1} \right) \left(\frac{1}{T} \sum_{t=1}^{T} r_t \right) \right] \\
&\qquad - \breve{\beta}^\top \left[\frac{1}{T} \sum_{t=1}^{T} x_{it-1} x_{t-1} - \left(\frac{1}{T} \sum_{t=1}^{T} x_{it-1} \right) \left(\frac{1}{T} \sum_{t=1}^{T} x_{t-1} \right) \right] \\
&\qquad = \sigma_{uv}^\top \Sigma_v^{-1} \left\{ \frac{1}{T} \sum_{t=1}^{T} x_{it-1} x_t - \left(\frac{1}{T} \sum_{t=1}^{T} x_{it-1} \right) \left(\frac{1}{T} \sum_{t=1}^{T} x_t \right) \right. \\
&\qquad \left. - \breve{\Theta} \left[\frac{1}{T} \sum_{t=1}^{T} x_{it-1} x_{t-1} - \left(\frac{1}{T} \sum_{t=1}^{T} x_{it-1} \right) \left(\frac{1}{T} \sum_{t=1}^{T} x_{t-1} \right) \right] \right\}, \quad \text{(A.42)}
\end{aligned}$$

for $i = 1, \ldots, N$. Recall that we have solved for $\breve{\Theta}$ in terms of the data, so (A.42) constitutes a system of linear equations in $\breve{\beta}_1, \ldots, \breve{\beta}_N$. Given the solution of this system for $\breve{\beta}$, (A.41) gives the estimator for μ_r.

A.4. Asymptotic standard errors

Here we derive asymptotic standard errors for our maximum likelihood estimates using the methodology described in Hayashi (2000). Let q denote the vector

$$(\mu_r, \mu_x, \beta, \theta, \sigma_u^2, \sigma_v^2, \sigma_{uv})^\top, \tag{A.43}$$

and let s_t denote the score vector for observation t. In addition, let

$$p\left(x_0|q\right) = \left(2\pi\sigma_x^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{x_0 - \mu_x}{\sigma_x}\right)^2\right\} \tag{A.44}$$

denote the likelihood of the initial draw x_0, and let

$$p\left(u_t, v_t|q\right) = |2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{\sigma_v^2}{|\Sigma|}u_t^2 - 2\frac{\sigma_{uv}}{|\Sigma|}u_t v_t + \frac{\sigma_u^2}{|\Sigma|}v_t^2\right)\right\} \tag{A.45}$$

denote the likelihood of the shock vector $(u_t, v_t)^\top$. We specify our objective function as $1/T$ times our exact likelihood function,

$$\frac{1}{T}\log p\left(r_1, \ldots, r_T; x_0, \ldots, x_T|q\right) = \frac{1}{T}\sum_{t=1}^{T}\left[\log p\left(u_t, v_t|q\right) + \frac{1}{T}p\left(x_0|q\right)\right], \tag{A.46}$$

where the equality follows by independence of the shocks over t, and by writing $p\left(x_0|q\right) = \sum_{t=1}^{T}\frac{1}{T}p\left(x_0|q\right)$. The score s_t is

$$s_t = \frac{\partial}{\partial q}\left[\log p\left(u_t, v_t|q\right) + \frac{1}{T}p\left(x_0|q\right)\right]. \tag{A.47}$$

We can see that the exact score is the conditional score $\frac{\partial}{\partial q}\log p\left(u_t, v_t|q\right)$ plus the "correction" term $\frac{\partial}{\partial q}\frac{1}{T}p\left(x_0|q\right)$.

The usual approach of obtaining the asymptotic covariance matrix is to derive a "sandwich estimator." Hayashi (2000, section 7.3) shows that, under maximum likelihood, the sandwich estimator simplifies due to the information matrix equality. One particularly convenient estimator of the asymptotic covariance matrix is

$$\text{Avar}\left(\hat{q}\right) = \left[\frac{1}{T}\sum_{t=1}^{T} s_t\ s_t^\top\right]^{-1}. \tag{A.48}$$

Hayashi notes that this estimator often has better finite-sample performance than the more complicated sandwich estimator, due the ease with which it is computed. The standard errors for our parameter estimates are given by the square root of the diagonal elements of Avar $(\hat{q})$ divided by $\sqrt{T}$.

It is straightforward to adopt the method above for restricted MLE; we set $\beta = 0$ and we drop the element of the score corresponding to β.

B. Further properties of maximum likelihood

B.1. The equity premium in levels

In this section we discuss how to translate our results for log returns into levels. For simplicity, assume that the log returns $\log(1 + R_t)$ are normally distributed. Then

$$E[R_t] = E\left[e^{\log(1+R_t)}\right] - 1 = e^{E[\log(1+R_t)]+\frac{1}{2}\mathrm{Var}(\log(1+R_t))} - 1. \tag{B.1}$$

Using the definition of the excess log return, $E[\log(1 + R_t)] = E[r_t]+E[\log(1+R_t^f)]$, so the above implies that

$$E[R_t - R_t^f] = e^{E[r_t]}e^{E\left[\log\left(1+R_t^f\right)\right]+\frac{1}{2}\mathrm{Var}(\log(1+R_t))} - 1 - E[R_t^f]. \tag{B.2}$$

Our maximum likelihood method provides an estimate of $E[r_t]$ and all other quantities above can be easily calculated using sample moments. Taking the sample mean of the series $R_t - R_t^f$ for the period 1953-2011 yields a risk premium that is 0.530% per month, or 6.37% per annum. On the other hand, using the above calculation and our maximum likelihood estimate of the mean of r_t gives an estimate of $\mathbb{E}[R_t - R_t^f]$ of 0.422% per month, or 5.06% per annum.[1] Thus our estimate of the risk premium in return levels is 131 basis lower than taking the sample average, in line with our results for log returns.

[1] In the data, in monthly terms for the period 1953-2011, the sample mean of R_t is 0.918%, the sample mean of R_t^f is 0.387%, the sample mean of $\log(1 + R_t^f)$ is 0.386% and the variance of $\log(1 + R_t)$ is 0.194%.

B.2. Comparison with Fama and French (2002)

Fama and French (2002) also propose an estimator that takes the time series of the dividend-price ratio into account in estimating the mean return. Noting the following return identity:

$$R_t = \frac{D_t}{P_{t-1}} + \frac{P_t - P_{t-1}}{P_{t-1}}, \tag{B.3}$$

and taking the expectation:

$$E[R_t] = E\left[\frac{D_t}{P_{t-1}}\right] + E\left[\frac{P_t - P_{t-1}}{P_{t-1}}\right], \tag{B.4}$$

they propose replacing the capital gain term $E[(P_t - P_{t-1})/P_{t-1}]$ with dividend growth $E[(D_t - D_{t-1})/D_{t-1}]$. They argue that, because prices and dividends are cointegrated, their mean growth rates should be the same. They find that the resulting expected return is less than half the sample average, namely 4.74% rather than 9.62%.

While their argument seems intuitive, a closer look reveals a problem. Let $X_t = D_t/P_t$, and let lower-case letters denote natural logs. Then

$$d_{t+1} - d_t = x_{t+1} - x_t + p_{t+1} - p_t. \tag{B.5}$$

Because X_t is stationary, $E[x_{t+1} - x_t] = 0$ and it is indeed the case that

$$E[d_{t+1} - d_t] = E[p_{t+1} - p_t]. \tag{B.6}$$

However, exponentiating (B.5) and subtracting 1 implies

$$\frac{D_{t+1} - D_t}{D_t} = \frac{X_{t+1}}{X_t}\frac{P_{t+1}}{P_t} - 1. \tag{B.7}$$

That is, stationarity of X_t implies (B.6), but not $E[(P_t - P_{t-1})/P_{t-1}] = E[(D_t - D_{t-1})/D_{t-1}]$. Namely it does not imply that the average level growth rates are equal.

For expected growth rates to be equal in levels, (B.7) shows that it must be the case that $E\left[\frac{X_{t+1}}{X_t}\frac{P_{t+1}}{P_t}\right] = E\left[\frac{P_{t+1}}{P_t}\right]$. It seems unlikely that there are general conditions under which this holds. Note that it follows from $E[\log(X_{t+1}/X_t)] = 0$ and Jensen's inequality that $E[X_{t+1}/X_t] > 1$.[2] This implies that the estimator proposed by Fama and French (2002) is inconsistent for the equity premium, and thus it is not necessary (or possible) to evaluate efficiency.

Nonetheless, our results show that assuming cointegration of prices and dividends can be very informative for estimation of the mean return.[3] Indeed, the intuition that we will develop in the next section is closely related to that conjectured by Fama and French (2002): The sample average of realized returns is "too high" because shocks to discount rates (proxied for by the dividend-price ratio) were negative on average over the sample period.

[2]Indeed, if we assume that growth rates of dividends and prices are log-normal, a necessary and sufficient condition for equality of expected (level) growth rates is that the variances of the log growth rates are equal:

$$\mathrm{Var}(d_{t+1} - d_t) = \mathrm{Var}(p_{t+1} - p_t). \tag{B.8}$$

To see this, note that (B.6), combined with log-normality, implies that

$$E\left[\frac{D_{t+1}}{D_t}\right] e^{-\frac{1}{2}\mathrm{Var}(d_{t+1}-d_t)} = E\left[\frac{P_{t+1}}{P_t}\right] e^{-\frac{1}{2}\mathrm{Var}(p_{t+1}-p_t)}. \tag{B.9}$$

If (B.8) holds, then the second terms on the right and left hand side cancel, yielding the result. This is a knife-edge result in which the variance of the log dividend-price ratio x_t and the covariance of x_t with log price changes cancel out. However, it is well-known that prices are more volatile than dividends (Shiller, 1981).

[3]This point is also made by Constantinides (2002), who suggests adjusting the mean return by the difference in the valuation ratio between the first and last observation. Constantinides derives conditions such that the resulting estimator has lower variance than the average return.

C. Properties of the time series of returns under the benchmark data generating process

C.1. *Mean reversion in returns*

Consider the effect of a series of shocks on excess returns (in this subsection, we will assume, for expositional reasons, that the mean excess return is zero):

$$\begin{aligned} r_t &= \beta x_{t-1} + u_t \\ r_{t+1} &= \beta\theta x_{t-1} + \beta v_t + u_{t+1} \\ r_{t+2} &= \beta\theta^2 x_{t-1} + \beta\theta v_t + \beta v_{t+1} + u_{t+2} \end{aligned} \tag{C.1}$$

and so on. Thus, for $k \geq 1$, the autocovariance of returns is given by

$$\text{Cov}\left(r_t, r_{t+k}\right) = \theta^k \beta^2 \text{Var}(x_t) + \theta^{k-1}\beta\sigma_{uv}, \tag{C.2}$$

where $\text{Var}(x_t) = \sigma_v^2/(1-\theta^2)$. An increase in θ increases the variance of the predictor variable. In the absence of covariance between the shocks u and v, this effect would increase the autocovariance of returns through the term $\theta^k\beta^2\text{Var}(x_t)$. However, because u and v are negatively correlated, the second term in (C.2), $\theta^{k-1}\beta\sigma_{uv}$ is also negative. We show below that this second term dominates the first for all positive values of θ up until a critical value, at which point the first comes to dominate.

Assume $\theta > 0$, $\beta > 0$ and $\sigma_{uv} < 0$, as we estimate the case to be in our data. Substituting in $\text{Var}(x_t) = \sigma_v^2/(1-\theta^2)$, multiplying by $(1-\theta^2) > 0$ and dividing through by $\theta^{k-1}\beta > 0$ shows that the autocovariance of returns is negative whenever

$$-\sigma_{uv}\theta^2 + \beta\sigma_v^2\theta + \sigma_{uv} < 0. \tag{C.3}$$

The left-hand side is a quadratic polynomial in θ with a positive leading coefficient. As a result, whenever this polynomial has two real roots in θ, the entire expression is negative if and only if θ lies in between those roots. Indeed, the polynomial has two real roots because its discriminant equals $\beta^2\sigma_v^4 + 4\sigma_{uv}^2 > 0$. Let θ_1 be the smaller of the two roots and let θ_2 be the larger one, that is,

$$\theta_2 = \frac{-\beta\sigma_v^2 + \sqrt{\beta^2\sigma_v^4 + 4\sigma_{uv}^2}}{-2\sigma_{uv}}. \tag{C.4}$$

Under our assumptions it is straightforward to prove that $\theta_1 < -1$ and $-1 < \theta_2 < 1$, so the only possible change of sign of the return autocovariance happens at θ_2. In particular, $\text{Cov}\left(r_t, r_{t+k}\right) < 0$ whenever $\theta < \theta_2$ and $\text{Cov}\left(r_t, r_{t+k}\right) > 0$ whenever $\theta > \theta_2$.

C.2. The variance of the sample mean return

By definition

$$\frac{1}{T}\sum_{t=1}^{T} r_t = \mu_r + \beta\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1} - \mu_x\right) + \frac{1}{T}\sum_{t=1}^{T} u_t, \tag{C.5}$$

thus

$$\begin{aligned}\text{Var}\left(\frac{1}{T}\sum_{t=1}^{T} r_t\right) = \beta^2\text{Var}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\right) + \text{Var}\left(\frac{1}{T}\sum_{t=1}^{T} u_t\right) \\ + 2\beta\text{Cov}\left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}, \frac{1}{T}\sum_{t=1}^{T} u_t\right).\end{aligned} \tag{C.6}$$

The variance of the average predictor is available and it depends on θ. The variance of the average residual does not depend on θ. Finally, the covariance of the average predictor and the average predictor depends on θ and ρ_{uv}. It is not a trivial quantity because even though u_t is uncorrelated with x_{t-1}, it is correlated with x_t via v_t whenever $\rho_{uv} \neq 0$ and thus it is also correlated with $x_{t+1}, x_{t+2}, \ldots, x_{T-1}$ whenever $\theta \neq 0$. In particular,

$$\text{Var}\left(\frac{1}{T}\sum_{t=1}^{T} u_t\right) = \sigma_u^2\frac{1}{T}, \tag{C.7}$$

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}x_{t-1}\right)=\frac{\sigma_v^2}{1-\theta^2}\left[\frac{1}{T}\left(1+2\frac{\theta}{1-\theta}\right)+\frac{2}{T^2}\frac{\theta(\theta^T-1)}{(1-\theta)^2}\right], \tag{C.8}$$

$$\mathrm{Cov}\left(\frac{1}{T}\sum_{t=1}^{T}x_{t-1},\frac{1}{T}\sum_{t=1}^{T}u_t\right)=\sigma_{uv}\left[\frac{1}{T}\frac{1}{1-\theta}+\frac{1}{T^2}\frac{\theta^T-1}{(1-\theta)^2}\right], \tag{C.9}$$

so that

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}r_t\right)=&\frac{1}{T}\left(\sigma_u^2+2\beta\frac{\sigma_{uv}}{1-\theta}+\beta^2\frac{\sigma_v^2}{1-\theta^2}\right)\\&-\frac{1}{T^2}2\beta\frac{1-\theta^T}{(1-\theta)^2}\left(\beta\theta\frac{\sigma_v^2}{1-\theta^2}+\sigma_{uv}\right).\end{aligned} \tag{C.10}$$

It follows that

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}r_t\right)=\frac{1}{T}\left(\sigma_u^2+\beta^2\frac{\sigma_v^2}{1-\theta^2}+2\beta\frac{\sigma_{uv}}{1-\theta}\right)+O\left(\frac{1}{T^2}\right). \tag{C.11}$$

The term $\sigma_u^2+\beta^2\sigma_v^2/(1-\theta^2)$ measures the contribution of the return shocks and the predictor to the variability of the sample-mean return. The term $\beta\sigma_{uv}/(1-\theta)$ measures the contribution of the covariance of the return shocks and the predictor shocks to the variability of the sample-mean return. The former term increases as θ increases, which says that the sample-mean return is more variable because the predictor is more variable. At the same time, the latter term becomes more negative as θ increases, so that in fact the overall variability of the sample-mean return can decrease.

D. Omitted tables and figures

Table D.1. Small-sample distribution of estimators: t-distributed shocks

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95 %
μ_r	0.322	Sample	0.323	0.138	0.098	0.320	0.552
		MLE	0.322	0.072	0.204	0.322	0.440
μ_x	−3.504	Sample	−3.504	0.578	−4.454	−3.498	−2.543
		MLE	−3.504	0.549	−4.404	−3.498	−2.589
β	0.090	OLS	0.746	0.634	−0.007	0.601	1.947
		MLE	0.683	0.594	0.040	0.533	1.836
θ	0.998	OLS	0.991	0.007	0.978	0.993	0.999
		MLE	0.992	0.006	0.980	0.993	0.998
σ_u	4.430	OLS	4.419	0.185	4.136	4.411	4.727
		MLE	4.419	0.185	4.136	4.410	4.727
σ_v	0.046	OLS	0.046	0.002	0.043	0.045	0.049
		MLE	0.046	0.002	0.043	0.045	0.049
ρ_{uv}	−0.961	OLS	−0.961	0.004	−0.967	−0.961	−0.954
		MLE	−0.961	0.004	−0.967	−0.961	−0.954

Notes: We simulate 10,000 monthly samples from

$$\begin{aligned} r_{t+1} - \mu_r &= \beta(x_t - \mu_x) + u_{t+1} \\ x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1}, \end{aligned}$$

where $[u_t, v_t]$ has a bivariate t-distribution. The sample length is as in postwar data. Parameters are set to their maximum likelihood estimates (assuming normally distributed shocks) where β and θ are adjusted for bias. We conduct benchmark maximum likelihood estimation (MLE) for each sample path (this assumes normality and is therefore mis-specified). As a comparison, we take sample means to estimate μ_r and μ_x (Sample) and use ordinary least squares to estimate the slope coefficients and the variance and correlations of the residuals (OLS). The table reports the means, standard deviations, and 5th, 50th, and 95th percentile values across simulations. We set the degrees of freedom for the t-distribution to 5.96. This matches the average kurtosis of the estimated residuals for returns and the dividend-price ratio, and takes into account that the kurtosis is downward biased.

Table D.2. Small-sample distribution of estimators: Calibration to OLS estimates and sample means

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95 %
μ_r	0.433	Sample	0.432	0.082	0.297	0.431	0.565
		MLE	0.432	0.049	0.352	0.432	0.513
μ_x	−3.545	Sample	−3.550	0.192	−3.865	−3.551	−3.232
		MLE	−3.550	0.184	−3.854	−3.552	−3.242
β	0.828	OLS	1.414	0.715	0.512	1.276	2.801
		MLE	1.372	0.689	0.515	1.241	2.675
θ	0.992	OLS	0.986	0.007	0.971	0.987	0.995
		MLE	0.986	0.007	0.972	0.988	0.995
σ_u	4.414	OLS	4.410	0.118	4.215	4.410	4.603
		MLE	4.408	0.118	4.214	4.408	4.601
σ_v	0.046	OLS	0.046	0.001	0.044	0.046	0.048
		MLE	0.046	0.001	0.044	0.046	0.048
ρ_{uv}	−0.961	OLS	−0.961	0.003	−0.965	−0.961	−0.956
		MLE	−0.961	0.003	−0.965	−0.961	−0.956

Notes: We simulate 10,000 monthly samples from

$$\begin{aligned} r_{t+1} - \mu_r &= \beta(x_t - \mu_x) + u_{t+1} \\ x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1}, \end{aligned}$$

where u_t and v_t are Gaussian and iid over time with standard deviations σ_u and σ_v and correlation ρ_{uv}. The sample length is as in postwar data. Parameters μ_r and μ_x are set to their sample averages, and parameters β, θ and variances and correlations are set to their OLS estimates. We conduct maximum likelihood estimation (MLE) for each sample path. We also report sample averages for μ_r and μ_x (Sample) and OLS estimates for the remaining parameters.

Table D.3. Small-sample distribution of estimators: calibration to 1927–2011 sample

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95 %
μ_r	0.391	Sample	0.390	0.080	0.258	0.389	0.522
		MLE	0.391	0.058	0.295	0.390	0.485
μ_x	−3.383	Sample	−3.383	0.196	−3.710	−3.385	−3.063
		MLE	−3.384	0.190	−3.701	−3.384	−3.074
β	0.650	OLS	1.039	0.547	0.336	0.941	2.063
		MLE	1.018	0.530	0.345	0.923	2.007
θ	0.991	OLS	0.987	0.006	0.976	0.988	0.995
		MLE	0.987	0.006	0.977	0.989	0.994
σ_u	5.464	OLS	5.460	0.119	5.265	5.459	5.655
		MLE	5.458	0.119	5.263	5.458	5.653
σ_v	0.057	OLS	0.057	0.001	0.055	0.057	0.059
		MLE	0.057	0.001	0.055	0.057	0.059
ρ_{uv}	−0.953	OLS	−0.953	0.003	−0.958	−0.953	−0.948
		MLE	−0.953	0.003	−0.958	−0.953	−0.948

Notes: We simulate 10,000 monthly samples from

$$\begin{aligned} r_{t+1} - \mu_r &= \beta(x_t - \mu_x) + u_{t+1} \\ x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1}, \end{aligned}$$

where u_t and v_t are Gaussian and iid over time with standard deviations σ_u and σ_v and correlation ρ_{uv}. The sample length is set to match the 1927–2011 sample, and parameters are set to their maximum likelihood estimates over this period. We conduct maximum likelihood estimation (MLE) for each sample path. As a comparison, we take sample means to estimate μ_r and μ_x (Sample) and use ordinary least squares to estimate the slope coefficients and the variance and correlations of the residuals (OLS). The table reports the means, standard deviations, and 5th, 50th, and 95th percentile values across simulations.

Table D.4. Small-sample distribution of MLE_0

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95%
		Sample	0.312	0.169	0.040	0.309	0.591
μ_r	0.312	MLE	0.312	0.090	0.164	0.312	0.458
		MLE_0	0.312	0.089	0.164	0.312	0.460
		Sample	−3.439	1.078	−5.226	−3.450	−1.675
μ_x	−3.437	MLE	−3.436	1.051	−5.172	−3.438	−1.713
		MLE_0	−3.436	1.044	−5.156	−3.435	−1.718
		OLS	0.678	0.601	−0.048	0.550	1.845
β	0	MLE	0.602	0.558	0.012	0.450	1.694
		MLE_0					
		OLS	0.9920	0.0063	0.9798	0.9933	0.9996
θ	0.9992	MLE	0.9928	0.0058	0.9812	0.9944	0.9988
		MLE_0	0.9982	0.0012	0.9959	0.9985	0.9995

Notes: We simulate 10,000 monthly data samples from

$$\begin{aligned} r_{t+1} - \mu_r &= u_{t+1} \\ x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1}. \end{aligned}$$

where u_t and v_t are Gaussian and iid over time with correlation ρ_{uv}. The sample length is as in postwar data. The parameters are set to their restricted maximum likelihood estimates in Table 1. For each sample path, we compute sample averages for μ_r and μ_x (Sample), OLS estimates of β and θ (OLS), unrestricted maximum likelihood (MLE, mis-specified in this case), and restricted maximum likelihood (MLE_0, correctly specified).

Table D.5. Estimates using multiple predictors

	returns	d/p	dsfp	tmsp
		Panel A: ML estimates		
μ_r	0.338			
μ_{xi}		−3.493	0.903	−0.871
β_i		0.893	−0.524	−0.143
θ_i		0.994	0.969	0.972
RMSE	4.569			
		Panel B: Sample and OLS estimates		
μ_r	0.441			
μ_{xi}		−3.548	0.904	−0.871
β_i		1.239	−0.157	−0.480
θ_i		0.991	0.968	0.973
RMSE	4.581			
		Panel C: Covariance matrix		
σ	4.391	0.046	0.101	0.246
ρ_{ui}		−0.957	−0.058	−0.115
ρ_{1i}			0.067	0.133
ρ_{2i}				−0.130

Notes: Estimates of

$$\begin{aligned} r_{t+1} - \mu_r &= \sum_{i=1}^{N} \beta_i (x_{it} - \mu_{x_i}) + u_{t+1} \\ x_{1,t+1} - \mu_{x_1} &= \theta_1 (x_{1t} - \mu_{x1}) + v_{1,t+1} \\ &\vdots \\ x_{N,t+1} - \mu_{x_N} &= \theta_N (x_{Nt} - \mu_{xN}) + v_{N,t+1} \end{aligned}$$

where u_t and $v_{1t}, \ldots, v_{Nt}$ are Gaussian and iid over time with covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \rho_{u1}\sigma_u\sigma_1 & \ldots & \rho_{uN}\sigma_u\sigma_N \\ \rho_{u1}\sigma_u\sigma_1 & \sigma_1^2 & \ldots & \rho_{1N}\sigma_1\sigma_N \\ \vdots & \vdots & \ddots & \\ \rho_{uN}\sigma_u\sigma_N & \rho_{1N}\sigma_1\sigma_N & & \sigma_N^2 \end{bmatrix},$$

where r_t is the continuously-compounded CRSP return minus the 30-day Treasury Bill return, x_{1t} is the log dividend-price ratio, x_{2t} is the default spread, and x_{3t} is the term spread. Data are monthly, April 1953 – December 2011. Means and standard deviations of returns are in percentage terms. In Panel A, parameters are estimated using maximum likelihood. In Panel B, μ_r and μ_{x_i} are estimated by sample averages, and β_i and θ_i are estimated by ordinary least squares. Panel C gives the standard deviations of the shocks (top row) and the correlations between the shocks estimated using OLS residuals. Variables are the dividend-price ratio (d/p), the continuously-compounded yield of BAA-rated bonds minus the continuously-compounded yield of AAA rated bonds (dfsp), and the continuously-compounded yield of ten-year treasury bonds minus the continuously-compounded yield of one-year treasury bonds (tmsp).

Table D.6. Annual estimates using repurchase-adjusted dividend-price ratios

	Treasury-stock adjusted d/p				Cash-flow adjusted d/p			
	OLS	Sample	MLE	MLE_0	OLS	Sample	MLE	MLE_0
μ_r		5.718	4.252	4.092		5.718	4.806	4.558
μ_x		−3.352	−3.334	−3.318		−3.258	−3.240	−3.221
β	19.556		17.221		21.343		19.868	
θ	0.897		0.923	0.977	0.865		0.883	0.958
σ_u	16.164		16.185	17.195	16.167		16.113	17.195
σ_v	0.125		0.126	0.125	0.130		0.130	0.130
ρ_{uv}	−0.700		−0.708	−0.658	−0.668		−0.674	−0.628
RMSE		17.233	16.470	16.598		17.233	16.581	16.606
p(ΔMSE)			0.021	0.102			0.023	0.094

Notes: Estimates of

$$\begin{aligned} r_{t+1} - \mu_r &= \beta(x_t - \mu_x) + u_{t+1} \\ x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1}, \end{aligned}$$

where r_t is the continuously-compounded CRSP return minus the annual Treasury Bill return and x_t is the logarithm of the dividend yield, adjusted for repurchases. Two such adjusted dividend-price ratios are considered: the cash-flow based yield (cfby) and the Treasury-stock based yield (tsby). Shocks u_t and v_t are mean zero and iid over time with standard deviations σ_u and σ_v and correlation ρ_{uv}. Return data and dividend-yield data are annual, 1953–2003. Means and standard deviations of returns are in percentage terms. Under the OLS columns, parameters are estimated by ordinary least squares, with σ_u, σ_v, and ρ_{uv} estimated from the residuals. In the Sample column, μ_r is the average excess return over the sample and μ_x is the average of the log dividend-price ratio. In the MLE column parameters are estimated using maximum likelihood. In the MLE_0 columns, parameters are estimated using maximum likelihood assuming $\beta = 0$. RMSE denotes the root-mean-squared error from monthly out-of-sample return forecasts.

Table D.7. Estimation of a predictive regression with heteroskedasticity

Panel A: Means and coefficients		Panel B: Volatility parameters		Panel C: Covariance matrix	
μ_r	0.335	ω_u	4.763	σ_u^*	4.351
μ_x	-3.569	α_u	0.029	σ_v^*	0.045
β	0.688	δ_u	0.719	ρ_{uv}	-0.959
θ	0.993	ω_v	1.855×10^{-4}		
		α_v	0.016		
		δ_v	0.892		

Notes: We estimate the bivariate process

$$
\begin{aligned}
r_{t+1} - \mu_r &= \beta(x_t - \mu_x) + u_{t+1} \\
x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1},
\end{aligned}
$$

where, conditional on information available up to and including time t,

$$
\begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{u,t+1} & \rho_{uv}\sigma_{u,t+1}\sigma_{v,t+1} \\ \rho_{uv}\sigma_{u,t+1}\sigma_{v,t+1} & \sigma^2_{v,t+1} \end{bmatrix}\right),
$$

and

$$
\begin{aligned}
\sigma^2_{u,t+1} &= \omega_u + \alpha_u u_t^2 + \delta_u \sigma^2_{u,t}, \\
\sigma^2_{v,t+1} &= \omega_v + \alpha_v v_t^2 + \delta_v \sigma^2_{v,t}.
\end{aligned}
$$

Here, r_t is the continuously compounded return on the value-weighted CRSP portfolio in excess of the return on the 30-day Treasury Bill and x_t is the log of the dividend-price ratio. Starred parameters are implied by other estimates, namely $\sigma_u^* = \sqrt{\omega_u/(1-\alpha_u-\delta_u)}$ and $\sigma_v^* = \sqrt{\omega_v/(1-\alpha_v-\delta_v)}$. Parameters are estimated using a two-stage process by which the means and coefficients (Panel A) are treated as fixed and the volatility parameters (Panels B and C) are estimated using conditional maximum likelihood in the first stage, and the volatility parameters are treated as fixed, while the means and coefficients are re-estimated in the second stage. Data are monthly, from January 1953 to December 2011. Means and standard deviations of returns are in percentage terms.

Table D.8. Small-sample distribution of estimators when the dividend-price ratio follows a random walk

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95 %
μ_r	0.322	Sample	0.325	0.166	0.050	0.327	0.599
		MLE	0.322	0.047	0.246	0.323	0.401
μ_x	−3.504	Sample	−2.988	0.699	−4.130	−2.996	−1.845
		MLE	−2.986	0.637	−4.006	−2.997	−1.971
θ	0.993	OLS	0.992	0.006	0.980	0.994	1.000
		MLE	0.993	0.006	0.981	0.995	0.999
σ_u	4.416	OLS	4.413	0.117	4.221	4.414	4.605
		MLE	4.415	0.117	4.223	4.417	4.607
σ_v	0.046	OLS	0.046	0.001	0.044	0.046	0.048
		MLE	0.046	0.001	0.044	0.046	0.048
ρ_{uv}	−0.961	OLS	−0.962	0.003	−0.967	−0.962	−0.957
		MLE	−0.962	0.003	−0.967	−0.962	−0.957

Notes: We simulate 10,000 monthly data samples from

$$r_{t+1} - \mu_r = u_{t+1}$$
$$x_{t+1} = x_t + v_{t+1}$$

where u_t and v_t are Gaussian and iid over time with correlation ρ_{uv}. For each sample path we conduct (mis-specified) maximum likelihood estimation (MLE) of

$$\begin{aligned} r_{t+1} - \mu_r &= \beta(x_t - \mu_x) + u_{t+1} \\ x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1}. \end{aligned}$$

For comparison, we take sample means to estimate μ_r and μ_x (Sample) and use ordinary least squares to estimate the slope coefficients and the variance and correlations of the residuals (OLS). The table reports the means, standard deviations, and 5th, 50th, and 95th percentile values across simulations.

Table D.9. Small-sample distribution of estimators when the dividend-price ratio has a time trend

	True Value	Method	Mean	Std. Dev.	5 %	50 %	95 %
μ_r	0.322	Sample	0.322	0.168	0.044	0.321	0.599
		MLE	0.280	0.145	0.044	0.280	0.516
μ_x	−3.504	Sample	−3.682	0.234	−4.066	−3.682	−3.292
		MLE	−3.663	0.223	−4.028	−3.661	−3.296
β	0	OLS	0.590	0.684	−0.255	0.460	1.880
		MLE	0.514	0.660	−0.270	0.375	1.756
θ	0.993	OLS	0.987	0.007	0.974	0.988	0.996
		MLE	0.988	0.007	0.975	0.989	0.996
σ_u	4.416	OLS	4.410	0.117	4.219	4.410	4.602
		MLE	4.409	0.117	4.218	4.410	4.601
σ_v	0.046	OLS	0.046	0.001	0.044	0.046	0.048
		MLE	0.046	0.001	0.044	0.046	0.048
ρ_{uv}	−0.961	OLS	−0.961	0.003	−0.965	−0.961	−0.956
		MLE	−0.961	0.003	−0.965	−0.961	−0.956

Notes: We simulate 10,000 monthly data samples from

$$\begin{aligned} r_{t+1} - \mu_r &= u_{t+1} \\ x_{t+1} - \mu_x &= \Delta + \theta(x_t - \mu_x) + v_{t+1} \end{aligned}$$

where u_t and v_t are Gaussian and iid over time with correlation ρ_{uv}. We set μ_r, μ_x, θ, σ_u, σ_v and ρ_{uv} to their benchmark maximum likelihood estimates (Table 1) and Δ to the mean residual $(1/T)\sum_{t=1}^{T}\hat{v}_t = -0.14868$. For each sample path we conduct (mis-specified) maximum likelihood estimation (MLE) of

$$\begin{aligned} r_{t+1} - \mu_r &= \beta(x_t - \mu_x) + u_{t+1} \\ x_{t+1} - \mu_x &= \theta(x_t - \mu_x) + v_{t+1}. \end{aligned}$$

For comparison, we take sample means to estimate μ_r and μ_x (Sample) and use ordinary least squares to estimate the slope coefficients and the variance and correlations of the residuals (OLS). The table reports the means, standard deviations, and 5th, 50th, and 95th percentile values across simulations.

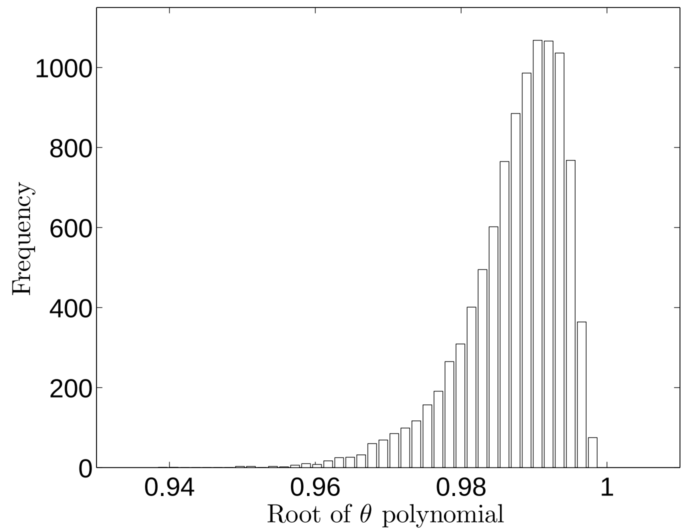

Fig. D.1. Histogram of maximum likelihood estimates of θ, the autocorrelation of the dividend-price ratio from simulated data. We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series.

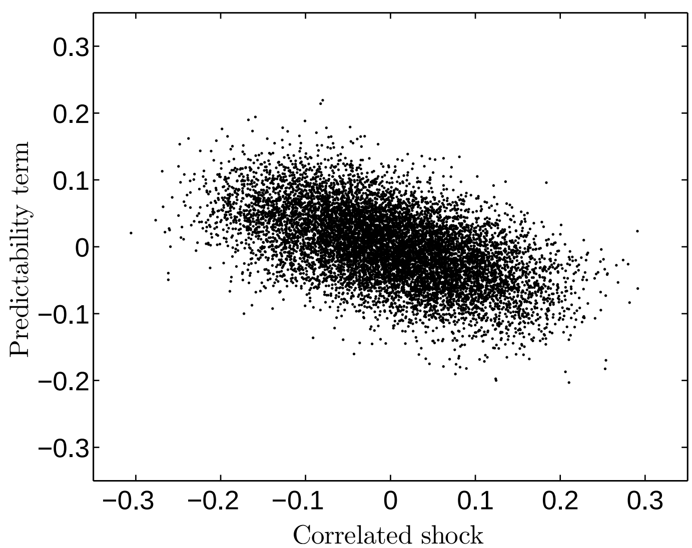

Fig. D.2. We simulate 10,000 monthly data samples from (1) with length and parameters as in the postwar data series. The figure shows the joint distribution of the predictability term $\hat{\beta}\frac{1}{T}\sum_{t=1}^{T}(x_{t-1}-\hat{\mu}_x)$ and the correlated shock term $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t$ that sum to the difference between the maximum likelihood estimate and the sample mean.

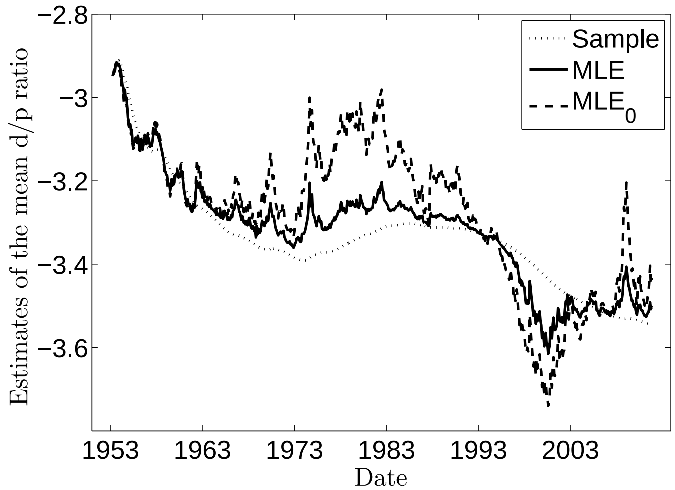

Fig. D.3. For each month, beginning in January 1953, we estimate the mean of the dividend-price ratio using maximum likelihood (MLE), maximum likelihood with the restriction $\beta = 0$ (MLE_0), and the sample mean (Sample), using data from January 1953 up until that month.

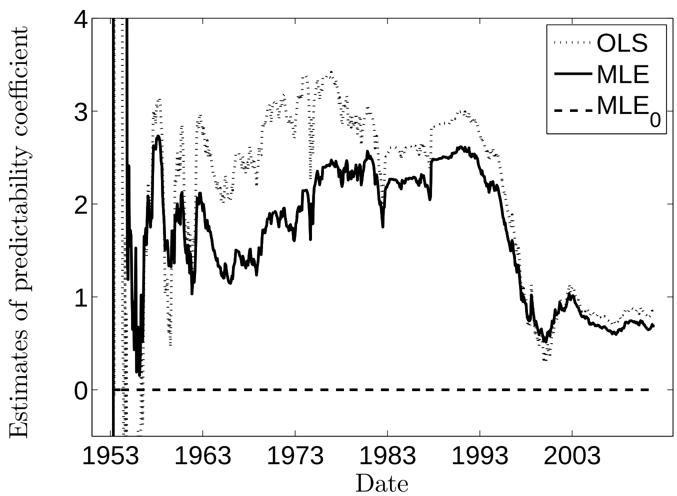

Fig. D.4. For each month, beginning in January 1953, we estimate the coefficient of predictability (β) using maximum likelihood (MLE), and Ordinary Least Squares (OLS), using data from January 1953 up until that month. For our restricted maximum likelihood method (MLE_0), $\beta = 0$ by assumption.

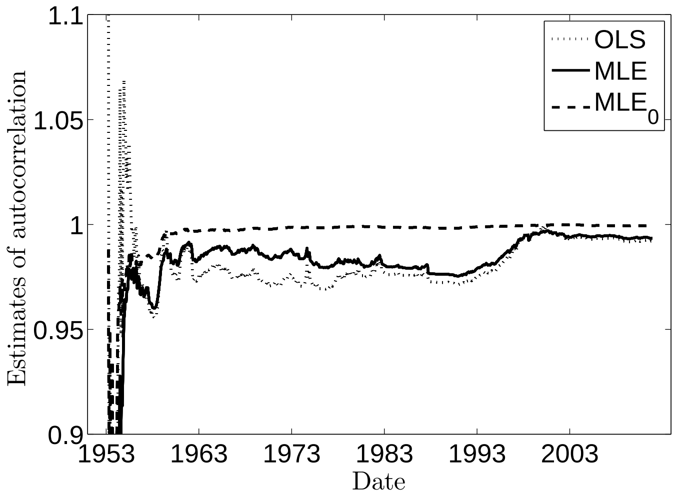

Fig. D.5. For each month, beginning in January 1953, we estimate the autocorrelation coefficient of the dividend-price ratio using maximum likelihood (MLE), maximum likelihood with the restriction $\beta = 0$ (MLE_0), and Ordinary Least Squares (OLS), using data from January 1953 up until that month.

References

Constantinides, G. M., 2002. Rational asset prices. The Journal of Finance 57, 1567–1591.

Fama, E. F., French, K. R., 2002. The equity premium. The Journal of Finance 57, pp. 637–659.

Hayashi, F., 2000. Econometrics. Princeton University Press, Princeton, NJ.

Shiller, R. J., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review 71, 421–436.