Temporary Components of Stock Returns: What Do the Data Tell Us?

Christopher G. Lamoureux
University of Arizona

Guofu Zhou
Washington University

Within the past few years several articles have suggested that returns on large equity portfolios may contain a significant predictable component at horizons 3 to 6 years. Subsequently, the tests used in these analyses have been criticized (appropriately) for having widely misunderstood size and power, rendering the conclusions inappropriate. This criticism however has not focused on the data, it addressed the properties of the tests. In this article we adopt a subjectivist analysis — treating the data as fixed — to ascertain whether the data have anything to say about the permanent/temporary decomposition. The data speak clearly and they tell us that for all intents and purposes, stock prices follow a random walk.

Whether there exists a predictable component to the returns on the stock market is a fundamental question in finance. The existence of such a component would

This article contains some concepts from (and subsumes) our earlier paper, "Exact Densities for the Decomposition of Time Series into Permanent and Transitory Components Using the Gibbs Sampler" (first draft June 1992). We thank an anonymous referee, Sid Chib, John Cochrane, Phil Dybvig, John Geweke, Ravi Jagannathan (the editor), Bill Lastrapes, Dale Poirier, and Robert Whitelaw, along with seminar participants at the Universities of Arizona, Minnesota, California at Los Angeles, and Toronto, as well as Wharton, Washington University, MIT, Goldman, Sachs, & Co., and the 1994 WFA meetings in Santa Fe for useful discussions. We retain responsibility for all errors. Please address request for reprints to Professor Lamoureux, Department of Finance, School of Business and Public Administration, The University of Arizona, Tuscon, AZ 85721.

The Review of Financial Studies Winter 1996 Vol. 9, No. 4, pp. 1033–1059

not be a prima facie cause to reject the notion of market efficiency, but it would have strong implications about the nature of the time-series properties of the investment opportunity set. If there were a nontrivial predictable component in long-horizon stock returns, it would also affect the way we test for long-run relationships between stock returns and other variables. Samuelson (1988) demonstrates that if stock returns are mean reverting, then investors with a relative risk aversion coefficient greater than 1 should optimally invest proportionally more in the stock market as their planning horizon lengthens. Fama and French (1988) examine the serial dependencies in long-horizon stock returns. They find correlation coefficients on the order of magnitude of 28% to 60% for 3- to 6-year returns on large portfolios. Fama and French (1988) found this evidence compelling enough to revise a prior belief that long-horizon market returns were unpredictable. Similarly, Poterba and Summers (1988) use classical (or frequentist) statistical inference and find that long-horizon stock returns show significant negative serial correlations. They issue a clarion call for a research agenda to discern whether this predictability is due to irrational trading (which generates long swings away from "intrinsic value") or time-varying risk premia.

Criticisms of the statistical tests of Fama and French and Poterba and Summers were forthcoming. For example, Richardson (1993) notes that Fama and French's results could have been obtained if the data were, in fact, generated by a pure random walk. In light of this, Fama (1991), in surveying what we know about the data, suggests that 60 years worth of data is insufficient to learn about the long-term properties of the data (i.e., the spectral density at frequency 0). Note however that Richardson's critique focuses primarily on a particular statistical technique, as opposed to the actual data.

Recent studies on return predictability have recognized the lack of power of the univariate tests and have shifted to multivariate analysis of predictability [Fama (1991)]. Common predictors include lagged interest rates and the dividend price ratio [see, for example, Campbell (1991)]. Despite the shift, a fresh look at the univariate case is valuable for several reasons. First, the importance of the univariate findings is still an open issue; as noted above, mean reverting behavior of the stock market returns is of interest in its own right. Second, many of the statistical problems that plague the univariate studies have analogs in the multivariate setting. Third, this article sheds new light on the earlier findings — both the original set of estimates, as well as their lack of power. Finally, these predictor variables are themselves highly autocorrelated, and data snooping concerns are certainly reduced when historical prices comprise the set of predictors. We use Bayesian methods to shed light on these puzzling results. We also focus directly on

predictability per se, as opposed to a particular null hypothesis that suggests that there is no predictability.

The distributions of frequentist test statistics are derived by assuming that the particular 60 years worth of data on hand is just one possible such draw. For the purpose of addressing how much information is in the data *on hand*, such a stochastic specification is unattractive. It involves integrating over possible draws (which may bear little resemblance to the actual data) from the null hypothesis. To refocus attention on the data itself, rather than the statistical technique, this article adopts a Bayesian framework that treats the data as fixed. The parameters of the model are considered random variables. Our analysis starts with a prior belief about the parameters and analyzes the extent to which the data cause us to revise those beliefs. Put differently, frequentist analysis conditions on the null hypothesis and performs a hypothesis test that is designed to ascertain whether the data on hand are unusual under the null; the subjectivist analysis conditions on the data and a prior belief about the parameters. Thus, the subjectivist approach, which we use in this article focuses on the effect that the data have on our prior beliefs. In order to focus attention on the data, we will evaluate the effect of the prior by using several different priors.

The model that we specify is an unobserved components model which allows stock prices to be generated by a process that includes both a random walk and a stationary process. The two processes are econometrically identified by assuming that they are mutually orthogonal [as in Watson (1986)]. We consider a variety of prior beliefs about the parameter space. The priors that we specify are proper so that we can analyse explicitly the implication of the prior for the relevant functions of interest. In this case these functions of interest include the ratio of the random walk variance to the total variance, and measures of the persistence of shocks to the stationary process and impulse response functions, as well as the regression coefficient of the temporary component of long-horizon returns on their lags (ρ).

The combined effect of the predictability in the temporary component and the relative magnitudes of the two components is summarized by the regression coefficient of long-horizon returns on their lags (β). Although Fama and French (1988) used β as an indication of the amount of predictability in returns, this autoregression coefficient gives a limited picture of maximum predictability. Since our model fully specifies the time-series properties of returns, we can examine the maximum reduction in forecast error variance attainable with n lags of returns. We will also present prior and posterior probabilities for measures of forecast error variance conditioned on various lengths of historical returns. The posterior densities of both of the regression

coefficients, half-lives, variance ratios, and forecast error variance are obtained directly, since our estimation procedure provides draws from the conditional densities of the structural parameters of the unobserved components model. The regression coefficients, for example, are simply (deterministic) functions of the structural parameters. For each prior, we combine the prior with the data and the likelihood to construct the posterior densities of the parameters as well as the functions of interest. This procedure provides posterior *densities* for the functions of interest, such as the variance ratio. Thus, the estimation procedure generates information about the location and precision of parameters and functions directly, instead of obtaining a point estimate and then having to evaluate the precision (or informativeness) of that point estimate as a separate step.

We thus see exactly the sense in which the data on hand cause us to revise our prior beliefs about these functions of interest. Despite the frequentist conjecture that 60 some years is not an adequate sample size to address these fundamental questions of interest, the data do significantly cause us to revise our priors, in all cases. Specifically, we find that unless the prior is very informative that the stationary component is large, the data shift most of the probability mass into the random walk component. More importantly, in all cases, the posterior belief about the half-life of shocks to the stationary component is such that this component is virtually nonstationary. The bottom line is that the data do speak clearly about long-horizon predictability.

Finally, we use a diffuse prior over the hyperparameters of the model to verify and shed light on the frequentist results. With a diffuse prior, the results are similar to those obtained in the literature. This provides evidence that the likelihood function is not driving the results. We demonstrate that the problem with the diffuse prior on the hyperparameters is that it has an unattractive implication for the prior on β. The remainder of the article is organized as follows. Section 1 contains the model, a characterization of the estimation procedure, and a summary of the functions of interest. The data are described and the results of estimation are presented in Section 2. Section 3 provides a summary of the motivation and findings of the study.

1. Analytics

1.1 The model

Our analysis is conducted for quarterly stock returns. The model we specify is an unobserved components model, which consists of a nonstationary random walk (with drift) process and an orthogonal station-

ary, autoregressive process for log price:

$$p_t = q_t + z_t, \tag{1}$$

$$q_t = q_{t-1} + \mu + u_t, \tag{2}$$

where q is a random walk with Gaussian white noise, $u \sim N(0, \sigma_u^2)$, and z is a stationary process that is independent of u. Equation (1) states that the stock price has two components: the random walk component and the stationary component. Shocks to the random walk component persist indefinitely, whereas shocks to the stationary component are irrelevant for the construction of forecasts in the *infinite* future. Notice that only p is observable; q and z are unobservable. Nevertheless, by assuming mutual orthogonality, the parameters are econometrically identified.[1,2]

In addition to the identifying restrictions noted above, complete specification requires a model for the stationary component of returns. Assume it has an AR(m) representation

$$z_t = \psi_1 z_{t-1} + \cdots + \psi_m z_{t-m} + \epsilon_t, \tag{3}$$

where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.[3] The implied model for returns (prices are adjusted to include dividends paid and adjusted for stock splits) is

$$\begin{aligned} r_t &= p_{t+1} - p_t \\ &= (q_{t+1} - q_t) + (z_{t+1} - z_t) \\ &= \mu + u_{t+1} + z_{t+1} - z_t. \end{aligned} \tag{4}$$

Because z_t is stationary, so is $z_{t+1} - z_t$, which we denote x_t. Equation (4) states that the return is the sum of a white noise process (first differenced random walk) and a stationary process.

---

[1] This model was used by Watson (1986) to decompose GNP from a frequentist perspective. Earlier examples of classical identification and estimation of unobserved components models include Harvey (1985) and Nerlove et al. (1979).

[2] Quah (1992) demonstrates that there are infinitely many identification schemes for the decomposition of a nonstationary series into permanent and temporary components, depending on the model for the nonstationary portion and the correlation between the two processes. For stock prices, the random walk model is theoretically motivated, as is orthogonality between the random walk and temporary component. An alternative tack to decompose the series employs additional time series (e.g., dividend data) to identify structurally the components [see, e.g., Cochrane (1992) and Cochrane and Sbordone (1988)].

[3] The identifying restrictions in our model are identical to those of Fama and French (1988). Fama and French do not, however, have to specify a model for the stationary process, since they directly estimate the function of interest β. This generality is costly though, because direct estimation of β is accomplished using a rolling-overlapping procedure that has a nontrivial effect, including bias, on the estimation itself.

A classical approach obtains the parameter estimates by maximizing the log-likelihood function (apart from a constant):

$$\begin{aligned}\log \mathcal{L}(\mu, \sigma_u^2, \psi_1, \ldots, \psi_m, \sigma_\epsilon^2) \\ = -\frac{1}{2}\log|\mathbf{\Gamma}_{rr}| - \frac{1}{2}(\mathbf{r} - \mu\mathbf{1})'\mathbf{\Gamma}_{rr}^{-1}(\mathbf{r} - \mu\mathbf{1}),\end{aligned} \tag{5}$$

where $\mathbf{\Gamma}_{rr}$ is the covariance matrix of $\mathbf{r} = (r_1, \ldots, r_T)'$:

$$\mathbf{\Gamma}_{rr} = \begin{pmatrix} \sigma_u^2 + \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \sigma_u^2 + \gamma_0 & \cdots & \gamma_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \sigma_u^2 + \gamma_0 \end{pmatrix}, \tag{6}$$

with $\gamma_j = \mathrm{cov}(x_t, x_{t+j})$. Notice that $\mathbf{\Gamma}_{rr}$ is a $T \times T$ matrix.

1.2 Estimation

This article exploits a set of tools that are most naturally couched in a Bayesian context to deal with this problem. Not only have efficient procedures been developed to implement the measurement problem, but we are also able to construct the posterior density for all of our functions of interest (since these may be expressed as deterministic functions of the underlying parameters of the process). Although this posterior density is obtained numerically, its error is controllable. In contrast, the asymptotic error of the classical approach depends on the sample size and it is difficult to determine whether or not the sample size is large enough to justify the asymptotic theory. For the problem at hand, considerations of power and size are somewhat controversial. The question of how to do the asymptotics in the case of Fama and French's (1988) regression statistic, β, is controversial. Richardson and Stock (1989) argue that the asymptotics should accommodate a concurrent growth in both the maximum return horizon (K) and the length of the data set (T). The *asymptotic* properties of these estimators is shown to be sensitive to whether K is allowed to grow along with T or is held fixed [as in Fama and French (1988)]. There is no room for such controversy with the posterior densities derived in this analysis.

In a frequentist approach geared toward *testing* there is also a Davies (1977, 1987) problem in this context, as discussed by Watson (1986). Specifically a classical test of whether σ_x^2 is zero involves parameters that are not identified under the null (the vector of ψ). By assigning zero probability (in our prior) to the state of nature — σ_x^2 is exactly 0 — our approach finesses this problem.

The Bayesian analysis starts with a prior distribution on the param-

eter space: $P_0(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\mu, \sigma_u^2, \psi_1, \ldots, \psi_m, \sigma_\epsilon^2)$ denotes all the parameters of the model. By Bayes' theorem, the posterior density is proportional to the prior density times the likelihood function:

$$P(\boldsymbol{\theta}) \propto P_0(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{\theta}). \tag{7}$$

The key in applications is to compute functions of interest which can often be summarized as the computation of the expected value of certain functions:

$$Eg(\boldsymbol{\theta}) = \int g(\boldsymbol{\theta})P(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{8}$$

As noted in the introduction, the functions of interest in understanding predictability on long-horizon stock returns include the regression coefficient of long-horizon returns on their lags, the ratio of the temporary component variance to the total variance, the half-life of shocks to the stationary component of prices, the impulse response function of shocks to the stationary component, the regression coefficient of the temporary component of long-horizon returns on their lags, and the forecast error variance as a function of the size of the conditioning set. For example, if $g(\boldsymbol{\theta}) = \beta$, $Eg(\boldsymbol{\theta})$ tells us the posterior mean of β, indicating the extent to which the data cause us to revise our beliefs about the size of the autoregression coefficient of long-horizon returns. However, it is difficult if not impossible to obtain $g(\boldsymbol{\theta})$ analytically. Furthermore, standard quadrature methods require a nontrivial amount of computational time, making them intractable in our applications. Fortunately, Monte Carlo integration with the Gibbs sampler offers an attractive solution.

The simple, intuitive idea behind Gibbs sampling is the use of all conditional densities to obtain the joint density, bypassing the difficulty of drawing samples directly from the marginal posterior distribution in Monte Carlo integrations. From a Bayesian perspective, latent variables and parameters are handled in the same manner. Gibbs sampling provides a tractable method of generating random draws from the latent variables and parameters conditional on the data and the prior. The tractability arises as the latent variables and parameters are drawn *sequentially*, conditioned on previous draws from the other parameters and latent variables. The sequential structure of the Gibbs sampler implies that we cannot draw u directly. Instead, define $v_t = u_t/\sigma_u$. We augment the data with $\mathbf{v}$, where $\mathbf{v} = (v_1, \ldots, v_T)'$. Note that the joint distribution of $\mathbf{v}$ and r is

$$\begin{pmatrix} \mathbf{v} \\ \mathbf{r} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mu\mathbf{1} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \sigma_u\mathbf{I} \\ \sigma_u\mathbf{I} & \boldsymbol{\Gamma}_{rr} \end{pmatrix}\right]. \tag{9}$$

The class of priors that we consider is

$$\psi \sim N(\bar{\psi}, \delta_0 \mathbf{I}_m), \tag{10}$$

$$\sigma_\epsilon^2 \sim \frac{\upsilon_0 s_0^2}{\chi_{\nu_0}^2}, \tag{11}$$

$$\sigma_u \sim N(\bar{\sigma}_u, \kappa_0), \sigma_u > 0. \tag{12}$$

(Thus the prior on σ_u is a truncated normal.) A diffuse prior is specified for μ. None of the results or functions of interest depend on μ. The *prior* density of the m vector ψ is complicated by the fact that the z process must be stationary.

The log likelihood of the model is

$$\mathcal{L}(\mu, \sigma_u^2, \psi_1, \ldots, \psi_m, \sigma_\epsilon^2) \propto -\frac{1}{2}\log|\mathbf{\Gamma}_{rr}| - \frac{1}{2}(\mathbf{r} - \mu\mathbf{1})'\mathbf{\Gamma}_{rr}^{-1}(\mathbf{r} - \mu\mathbf{1}), \tag{13}$$

(recall $\mathbf{\Gamma}_{rr}$ is the covariance matrix of $\mathbf{r}$, defined in Equation (6) above). Given the prior density on the parameters and the likelihood, Bayes' theorem [Equation (7)] provides the posterior density. For the class of priors used, the following posterior densities result:

$$\mu \sim N(\hat{\mu}, \sigma_\mu^2), \tag{14}$$

$$\psi \sim N(\tilde{\psi}, \mathrm{var}(\psi)), \tag{15}$$

$$\sigma_\epsilon^2 \sim \frac{\upsilon_1 s_1^2}{\chi_{\nu_1}^2}, \tag{16}$$

$$\sigma_u \sim N(\tilde{\sigma}_u, \kappa_1), \sigma_u > 0, \tag{17}$$

where

$$\hat{\mu} = \mathbf{1}'\mathbf{\Gamma}_{\mathbf{xx}}^{-1}(\mathbf{r} - \sigma_u\mathbf{v})/\mathbf{1}'\mathbf{\Gamma}_{\mathbf{xx}}^{-1}\mathbf{1}, \tag{18}$$

$$\sigma_\mu^2 = (\mathbf{1}'\mathbf{\Gamma}_{\mathbf{xx}}^{-1}\mathbf{1})^{-1}, \tag{19}$$

$$\hat{\sigma}_u = \mathbf{v}'\mathbf{\Gamma}_{\mathbf{xx}}^{-1}(\mathbf{r} - \mu\mathbf{1})/\mathbf{v}'\mathbf{\Gamma}_{\mathbf{xx}}^{-1}\mathbf{v}, \tag{20}$$

$\mathbf{\Gamma}_{\mathbf{xx}}$ is the $(T \times T)$ covariance matrix of x; and

$$\mathbf{z} = \begin{pmatrix} z_2 \\ z_3 \\ \vdots \\ z_{T+1} \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} z_1 & \cdots & z_{2-m} \\ z_2 & \cdots & z_{3-m} \\ \vdots & \ddots & \vdots \\ z_T & \cdots & z_{T+1-m} \end{pmatrix}, \tag{21}$$

$$\kappa = 1/\mathbf{z}'\mathbf{\Gamma}_{rr}^{-1}z, \tag{22}$$

$$\mathrm{var}(\psi) = (\mathbf{I}_m/\delta_0 + \mathbf{Z}'\mathbf{Z}/\sigma_\epsilon^2)^{-1}, \tag{23}$$

$$\tilde{\psi} = \mathrm{var}(\psi)(\bar{\psi}/\delta_0 + \mathbf{Z}'\mathbf{z}/\sigma_\epsilon^2), \tag{24}$$

and $\nu_1 = \nu_0 + T$, $s^2 = \frac{1}{T}(\mathbf{z} - \mathbf{Z}\psi)'(\mathbf{z} - \mathbf{Z}\psi)$, $s_1^2 = (\nu_0 s_0^2 + Ts^2)/\nu_1$, $\kappa_1 = \kappa_0\kappa/(\kappa_0 + \kappa)$, and $\tilde{\sigma}_u = (\kappa\bar{\sigma}_u\kappa_0\hat{\sigma}_u)/(\kappa_0 + \kappa)$. The conditional distribution of $\mathbf{v}$ is

$$\mathbf{v}|\mathbf{r} \sim N[\sigma_u \mathbf{\Gamma}_{rr}^{-1}(\mathbf{r} - \mu\mathbf{1}), \mathbf{I} - \sigma_u^2 \mathbf{\Gamma}_{rr}^{-1}]. \tag{25}$$

The way the Gibbs sampler works is that we draw from each of the posteriors — *conditional on the data and the current draw from each of the other parameters and* $\mathbf{v}$ — as above and recognize that the marginal density is the average of the conditional densities. Given the current draw on $\mathbf{v}$, σ_u, μ, and the augmented initial values of the z process (i.e., $z_1, \ldots, z_{2-m}$), the values of $\mathbf{z}$ may be formed directly ($z_{t+1} = z_t + r_t - \mu - \sigma_u v_t$) (where v_t is a pseudorandom draw from a standard unit normal density). At each draw we also construct a new set of the augmented (or starting values of) $\mathbf{z}$ by projecting the autoregressive model backwards. This eliminates a need to condition on arbitrary starting values for $\mathbf{z}$. In this manner we integrate over these starting values, in the same sense that we integrate over the $\mathbf{v}$ in forming the posterior densities, which is natural given the sequential structure of the Gibbs sampler. An identifying restriction of the unobserved components model is that the vector ψ obtained at each draw from the Gibbs sampler be such that z is stationary. This condition is verified at each step. Values of ψ that are inconsistent with stationarity are discarded.

In practice, the starting condition is arbitrary. We take 11,000 draws from the posterior densities, and to define the posterior density, we discard the first 1,000 draws. The entire process is summarized by the densities of the parameters. Any functions of interest can now be evaluated by constructing their densities from the underlying parameters. Thus, for example, the density of the variance ratio of the random walk to the total ($\frac{\sigma_\mu^2}{\sigma_r^2}$) is constructed by taking this ratio in each of the 10,000 draws. These 10,000 values of the ratio comprise the posterior density of the variance ratio.

This procedure amounts to Monte Carlo simulation (to compute the marginal densities), and as such there is some numerical error in the computations. The numerical accuracy of the Gibbs sampler may be assessed using the procedure developed by Geweke (1992). All of the information regarding the numerical efficiency of the Gibbs procedure is contained in the spectral density at frequency 0 of the series of Gibbs draws on a particular function of interest. Denote this $S_G(0)$. The periodogram of the 10,000 Gibbs draws on a particular function of interest may be estimated using a fast Fourier transform. Next, the spectral density at frequency 0 is approximated by smoothing the periodogram using both a Daniell window [following Geweke (1992)]

and a Parzen window with spectral window parameter M of $(N^{.5} \cdot .3)$, which is 30 in our case (recall that N refers to the number of Gibbs draws after eliminating the preliminary sample).[4]

Having identified $S_G(0)$ (which will be different for each parameter or function of interest), we can define two measures of the numerical efficiency of the Gibbs sampler. First, the numerical standard error is $[\frac{1}{N} \cdot S_G(0)]^{.5}$. The relative numerical efficiency of the mean of the function of interest is $\text{var}[g(\boldsymbol{\theta})]/S_G(0)$ [a general discussion of assessing numerical accuracy in Bayesian estimation is Geweke (1989)]. As Geweke (1992, p. 8) describes: "The number of drawings required to achieve a given degree of numerical accuracy is inversely related to the relative numerical efficiency of the Gibbs sampling process for the function of interest." We provide the reader with the numerical standard error and the relative numerical efficiency of the parameters as well as the functions of interest, when we discuss the results, to give a feel for the technical attributes of the Gibbs sampler.

1.3 Functions of interest

The Gibbs sampler provides a series of draws on the structural parameters of the model. These draws provide the posterior densities for the parameters. All of the functions that will characterize the predictability in long-horizon returns may be written in terms of these structural parameters, and therefore for each draw from the structural parameters we can construct a draw from any of these functions and construct their posterior densities accordingly. The functions $\frac{\sigma_u^2}{\sigma_r^2}$ and $\frac{\sigma_x^2}{\sigma_r^2}$ summarize the contributions of the differenced random walk and the stationary process, respectively, to returns. As noted above, isolating these variance ratios may be misleading, however (as we will see below), because the stationary process may be virtually nonstationary, that is characterized by very persistent shocks.

As noted in the introduction, we measure the persistence of shocks to the stationary z process using the half-life. The notion of half-life is not characterized by the same sort of knife-edge tension as the notion of a unit root. Half-life is defined, and finite, for any stationary process. A stationary process may have its largest eigenvalue very close to 1 so that the half-life is very long. This is important to note, because a test for a unit root may be devoid of economic content. The half-life of a shock indicates the number of periods (in this case quarters) that it takes for one-half of the impact of the shock to dissipate. (For a

[4] Geweke (1992) does not explore the effect on estimates of numerical accuracy of different window choices, and this is certainly of tertiary order of importance in the present article. For a discussion of the Daniell and Parzen windows see Priestley (1981, pp. 437–449).

nonstationary process, half-life is infinite.) Half-life is defined as:

$$\text{half-life} = 1 - \frac{\log 2}{\log \lambda}; \quad \left(\lambda = \sum_{j=1}^{m} \psi_j\right).$$

An alternative way of presenting the damping pattern of shocks to a stationary process over time is the impulse response function. Popularized by Sims (1980), the impulse response functions are the coefficients in the moving average representation of the stationary component. They are empirical comparative statics, measuring how much impact a unit standard deviation shock to an exogenous process today will have in the future. Given the parameters of the AR process of the stationary component, the impulse response functions are computed using polynomial division:

$$c_j = \sigma_\epsilon \mathbf{J}\mathbf{\Phi}^j\mathbf{J}', \tag{26}$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \psi_1 & \psi_2 & \cdots & \psi_{m-1} & \psi_m \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{27}$$

and $\mathbf{J} = (1, 0, \ldots, 0)$ [see Baillie (1987)]. A moving average representation for the stationary component is

$$x_t = c_0\eta_t + c_1\eta_{t-1} + \cdots = \sum_{j=0}^{\infty} c_j\eta_{t-j}. \tag{28}$$

Another way to characterize the predictability of returns is to isolate the regression coefficient of long-horizon returns on their lags. First, consider the autoregression of the temporary component of long-horizon returns ρ_K, and the slope of regression of $z_{t+K} - z_t$ on $z_t - z_{t-K}$:

$$\rho_K = \frac{\text{cov}[z_{t+K} - z_t, z_t - z_{t-K}]}{\text{var}[z_{t+K} - z_t]}. \tag{29}$$

As shown by Fama and French, (1988), ρ_K goes to -0.5 as K becomes large.[5] However, for small values of K, ρ_K may be close to 0. To compute ρ_K from a draw of the parameters, it is enough to express

[5] This follows because z is a stationary process. Therefore, the covariance between z_{t+K} and z_t approaches 0 as K increases. Intuitively, for large K, this amounts to a regression of a random variable on its negative plus an orthogonal random variable with the same variance.

it as a function of the covariances of the x_t's, since these covariances are already simple functions of the parameters:

$$\rho_K = \frac{\sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \operatorname{cov}(x_i, x_{j-K})}{\sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \operatorname{cov}(x_i, x_j)}. \tag{30}$$

Clearly the predictability of long-horizon returns will depend on ρ as well as $\frac{\sigma_x^2}{\sigma_r^2}$. This may be summarized by the β function, which is the slope of the regression of $r_{t+K} - r_t$ on $r_t - r_{t-K}$:

$$\begin{aligned} \beta_K &= \frac{\operatorname{cov}[r_{t+K} - r_t, r_t - r_{t-K}]}{\operatorname{var}[r_{t+K} - r_t]} \\ &= \frac{\rho_K \operatorname{var}[z_{t+K} - z_t]}{\operatorname{var}[z_{t+K} - z_t] + \operatorname{var}[q_{t+k} - q_t]}. \end{aligned} \tag{31}$$

Note in our model that $\operatorname{var}[z_{t+K} - z_t] = \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} \operatorname{cov}(x_i, x_j)$ and $\operatorname{var}[q_{t+K} - q_t] = K\sigma_u^2$. β_K has interesting alternative interpretations. If the stock price does not have a stationary component, $\beta_K = 0$. If the stock price does not have a random walk component, $\beta_K = \rho_K$. If the stock price has both a stationary and a random walk component, the behavior of β_K is more complex. Fama and French (1988) suggest a U-shape of β_K over K.

β is only a partial measure of the predictability of long-horizon returns. We might also be interested in the optimal linear forecast of k-step ahead returns. A measure of predictability then is the ratio of the conditional forecast error variance to the total variance. The minimum forecast error variance for any horizon would condition on the infinite past. Consider the 1-step (quarter) ahead forecast error variance. We can define

$$\text{maximum forecastability} = \frac{\sigma_r^2 - (\sigma_u^2 + \sigma_\epsilon^2)}{\sigma_r^2},$$

which indicates the percentage reduction in the forecast error variance obtained by conditioning on the infinite past relative to an unconditional forecast.

Since the infinite past is not available when constructing forecasts, we also examine the prior and posterior densities for the 1-step ahead forecast error variance conditional on n lags, as n varies from 1 to 300

quarters. These forecast errors may be obtained recursively using the innovations algorithm:[6]

$$\xi_0 = \mathbf{\Gamma}_{rr}(1, 1), \tag{32}$$

$$\begin{aligned}\zeta_{n,n-k} &= \xi_k^{-1}\{\mathbf{\Gamma}_{rr}(n+1, k+1) \\ &\quad - \sum_{j=0}^{k-1} \zeta_{k,k-j}\zeta_{n,n-j}\xi_j\}, \; k = 0, 1, \ldots, n-1,\end{aligned} \tag{33}$$

$$\xi_n = \mathbf{\Gamma}_{rr}(n+1, n+1) - \sum_{j=0}^{n-1} \zeta_{n,n-j}^2 \xi_j, \tag{34}$$

where $\mathbf{\Gamma}_{rr}(j, k)$ represents the (j, k) element of $\mathbf{\Gamma}_{rr}$. Here, ξ_n represents the 1-step ahead forecast error variance, conditional on n lags. Thus, ξ_0 corresponds to the variance of r; ζ is an intermediate variable in the recursions.

2. Data and Results

The stock return data is taken from the CRSP tapes. We use the monthly returns file on the CRSP value-weighted NYSE index (with dividends). Quarterly returns are constructed by compounding the monthly returns. This series covers the period 1926:I through 1990:IV, providing a total of 260 observations.[7] Quarterly data are used in this article for two reasons. First, quarterly data is more consistent with the unconditional Gaussian assumptions used throughout than is higher frequency data. Second, the computational burden would be insurmountable with monthly data.

In the next five subsections we present results for two proper (informative) priors. In Section 2.6, we use a diffuse prior to clarify the findings and shed light on earlier findings.

2.1 Prior 1

The results from the first prior are presented in Table 1. For each parameter or function of interest, the first row corresponds to the prior and the second row to the posterior. Here, the prior on the variance

[6] Conceptually, this amounts to solving a series of difference equations. This algorithm exploits the special structure of the autocovariance matrix; see [Brockwell and Davis (1991), p. 172].

[7] The data differ slightly from that used by Fama and French (1988). Fama and French's data go through 1985. They use monthly returns to construct the long-horizon returns, and they deflate returns using the U.S. CPI. To demonstrate the similarity between the two data sets, we replicate the Fama and French OLS regressions for 1- through 6-year returns. The following are the Fama and French (1988, p. 258) (our) OLS estimates for these horizons: 1. −.05 (−.12); 2. −.24 (−.23); 3. −.32 (−.29); 4. −.19 (−.25); 5. −.07 (−.21); 6. .09 (−.09).

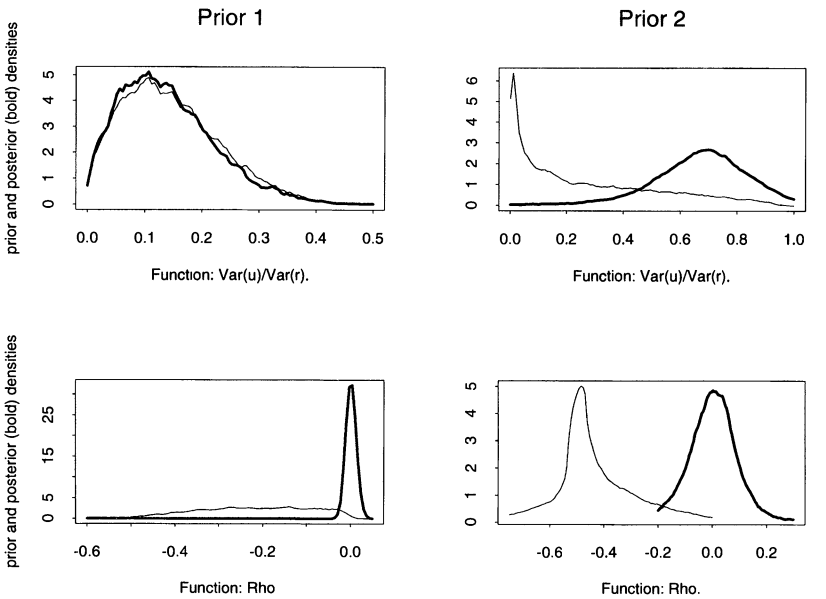

Figure 1
Prior and posterior densities for the two priors
$\mathrm{Var}(u)/\mathrm{var}(r)$ is the ratio of the variance of the permanent component to the total variance of returns on the CRSP value-weighted index. Rho is the autocorrelation coefficient of the *temporary* component of returns. The top row reports the prior densities and posterior densities for the variance ratio. The lower row reports the prior and posterior densities for Rho. The data tend to pull ρ toward 0, although the separate components of predictability are not separately well identified.

ratio of the random walk to the total variance is centered at 14.64%. The interquartile range for this variance ratio is 8% to 20%. The lighter curve of the panel entitled "Prior 1," in the top panel of Figure 1, is a graphical representation of the prior on this variance ratio.[8] Recall that the prior $(P_0(\boldsymbol{\theta}))$ is fully specified by the vector $(\nu_0, s_0^2, \bar{\sigma}_u, \kappa_0, \delta_0, \psi)$. The implications of this vector for the prior density are not so straightforward. Within the class of informative priors considered [Equations (10) through (12)] the effect of the choice of this vector on the functions of interest is not obvious. Generally speaking, δ_0 characterizes the precision of the prior on the autoregressive coefficients (i.e., lower δ_0 implies a tighter prior). Similarly, η_0 and κ_0 characterize the confidence the prior places in knowledge of σ_ϵ and σ_u, respectively. However, in this application attention is focused on functions of the

[8] The prior is constructed by taking 10,000 draws from the prior densities [Equations (10) through (12)]. For purposes of the plots, all densities are approximated using a rectangular window to smooth the histogram.

Table 1
Distribution of parameters for prior 1 260 quarterly stock returns

Variable	Mean	Std dev.	Percentile 5	25	50	75	95
σ_u	0.0402	0.0138	0.0171	0.0310	0.0403	0.0496	0.0630
	0.0420	0.0138	0.0187	0.0326	0.0422	0.0513	0.0645
σ_ϵ	0.1015	0.0174	0.0773	0.0893	0.0992	0.1110	0.1334
	0.1092	0.0074	0.0966	0.1045	0.1095	0.1141	0.1208
σ_r	0.1115	0.0173	0.0870	0.0993	0.1094	0.1213	0.1426
	0.1181	0.0050	0.1101	0.1146	0.1179	0.1213	0.1222
ψ_1	0.9633	0.0288	0.9158	0.9442	0.9631	0.9829	1.0108
	0.9718	0.0250	0.9301	0.9550	0.99719	0.9881	0.9923
ψ_2	−0.0017	0.0288	−0.0496	−0.0212	−0.0018	0.0180	0.0454
	0.0212	0.0258	−0.0212	0.0038	0.0210	0.0387	0.0636
ψ_3	−0.0019	0.0287	−0.0493	−0.0212	−0.0017	0.0175	0.0451
	0.0278	0.0079	−0.0146	0.0108	0.0277	0.0450	0.0704
ψ_4	−0.0115	0.0285	−0.0584	−0.0307	−0.0111	0.0078	0.0350
	−0.0208	0.0247	−0.0619	−0.0372	−0.0209	−0.0041	−0.0198
β_{12} (%)	−17.68	10.32	−35.60	−25.28	−17.15	−9.24	−2.33
	0.13	1.71	−1.26	−0.46	0.12	0.76	1.76
ρ_{12} (%)	−21.61	12.49	−42.97	−31.22	−21.19	−11.32	−2.77
	0.12	2.04	−1.51	−0.55	0.15	0.88	1.98
Half-life	66.5	511.3	6.1	10.0	16.3	33.4	157.6
(qtrs)	1.9×10^{12}	9.8×10^{13}	13,672	39,609	81,547	198,043	1,155,246
$\frac{\sigma_u^2}{\sigma_r^2}$	0.1464	0.0872	0.0258	0.0795	0.1350	0.2000	0.3082
	0.1400	0.0842	0.0251	0.0762	0.1282	0.1899	0.2959
$\frac{\sigma_x^2}{\sigma_r^2}$	0.8536	0.0872	0.6918	0.7999	0.8650	0.9205	0.9742
	0.8600	0.0842	0.7038	0.8101	0.8717	0.9238	0.9749

For each parameter or function, the first row represents the prior density and the second row represents the posterior.
Model:

$$P_t = q_t + z_t$$

$q \sim$ random walk (with drift); $u = \Delta(q)$; $x = \Delta(z)$.

$$z_t = \psi_1 z_{t-1} + \psi_2 z_{t-2} + \psi_3 z_{t-3} + \psi_4 z_{t-4} + \epsilon_t$$

Estimation:
The posterior density is constructed from 10,000 Gibbs draws as described in the text. A start-up sample of 1,000 draws is discarded to remove dependence on initial values. The data are 260 quarterly returns on the CRSP NYSE value weighted index.
σ_u, σ_ϵ, σ_r, and ψ_j, $(j = 1, \ldots, m)$ are the parameters of the likelihood function. The remaining functions may all be written in terms of these parameters: β_{12} is the slope of a regression of returns compounded over 12 quarters on lagged 12-quarter returns (i.e., the predictability in 3-year stock returns); ρ_{12} is the slope of a regression coefficient of the stationary component of returns compounded over 12 quarters; half-life represents the number of quarters that it takes for one-half of a shock to the stationary component of returns to dissipate: (half life $= 1 - \frac{\log 2}{\log(\sum_{j=1}^{m} \psi_j)}$).

hyperparameters, not the hyperparameters themselves. Thus, to select the priors on the hyperparameters we simulate from the prior to examine the prior densities of the functions of interest. For example, the parameters used to generate the prior densities of Table 1 are as follows: $\nu_0 = 20$, $s_0^2 = 0.0095$, $\bar{\sigma}_u = 0.04$, $\kappa_0 = 0.0002$, $\delta_0 = 0.000975$, $\psi_1 = 0.975$, $\psi_2 = 0.01$, $\psi_3 = 0.01$, and $\psi_4 = 0$.

The model specified here is an AR(4) for z. (We used several priors with an AR(2) which produced qualitatively similar posteriors.) The

Table 2
Distribution of parameters for prior 2 260 quarterly stock returns

Variable	Mean	Std dev.	Percentile 5	25	50	75	90
σ_u	0.0847	0.0629	0.0071	0.0350	0.0724	0.1212	0.2057
	0.0990	0.0131	0.0774	0.0940	0.1007	0.1068	0.1151
σ_ϵ	0.0994	0.0415	0.0565	0.0726	0.0899	0.1142	0.1751
	0.0514	0.0167	0.0290	0.0401	0.0490	0.0599	0.0826
σ_r	0.2221	0.2207	0.0933	0.1362	0.1797	0.2431	0.4516
	0.1192	0.0054	0.1107	0.1155	0.1190	0.1227	0.1287
ψ_1	0.6599	0.3990	−0.0182	0.3953	0.6725	0.9300	1.3024
	0.6274	0.2173	0.2586	0.4936	0.6344	0.7690	0.9617
ψ_2	−0.0743	0.4064	−0.7442	−0.3532	−0.0756	0.2038	0.5985
	0.2160	0.1624	−0.0517	0.1131	0.2195	0.3245	0.4782
ψ_3	−0.1475	0.3819	−0.7442	−0.3532	−0.0756	0.2038	0.5985
	0.4485	0.2160	0.0785	0.3609	0.4805	0.5874	0.7125
ψ_4	−0.0274	0.3305	−0.5819	−0.2610	−0.0207	0.2064	0.5147
	−0.2920	0.1711	−0.5430	−0.4066	−0.3089	−0.1965	0.0193
β_{12} (%)	−22.26	20.74	−59.61	−35.44	−17.72	−6.83	−0.83
	0.95	2.57	−1.02	−0.46	0.05	1.43	6.04
ρ_{12} (%)	−41.99	20.62	−70.97	−50.79	−46.67	−34.12	−4.80
	0.24	9.50	−15.72	−5.40	0.34	5.75	15.32
Half-life	8.0	78.4	1	2	2	4	18
(qtrs)	2.5×10^{10}	1.4×10^{12}	2,972	11,073	26,059	67,296	364,815
$\frac{\sigma_u^2}{\sigma_r^2}$	0.2554	0.2486	0.0013	0.0373	0.1713	0.4256	0.7553
	0.7009	0.1506	0.4259	0.6322	0.7202	0.7963	0.9064
$\frac{\sigma_\epsilon^2}{\sigma_r^2}$	0.7446	0.2486	0.2447	0.5744	0.8285	0.9626	0.9987
	0.2991	0.1506	0.0936	0.2036	0.2798	0.3677	0.5740

See notes to Table 1. Specification:

$$z_t = \psi_1 z_{t-1} + \psi_2 z_{t-2} + \psi_3 z_{t-3} + \psi_4 z_{t-4} + \epsilon_t.$$

prior is relatively tight on the AR coefficients.[9] The median half-life of shocks to z is 16.3 quarters, and the interquartile range is 10 to 33 quarters. (As can be noted from the reported percentiles in the tables, half-life is potentially very skewed to the right. Therefore, mean and variance may be poor estimates of central tendency and scale, respectively.)

The data speak very loudly about the total return standard deviation. Here the prior is centered at .112, and the posterior is centered at .118. The interquartile range for the prior is .09 to .12; for the posterior, it is .115 to .121. For this prior, the prior dominates the variance ratio and the posterior ratios closely resemble those of the prior. This can be seen in the top panel entitled "Prior 1" of Figure 1. As noted above, predictability has two components: one is measured by the variance ratio, the other by ρ. As can be seen in the lower panel of Figure 1 for

[9] Stationarity is imposed on the AR process. Therefore, in order to have a prior that puts a lot of mass on the largest eigenvalue close to 1, the standard deviation of the prior (i.e., δ) must be small.

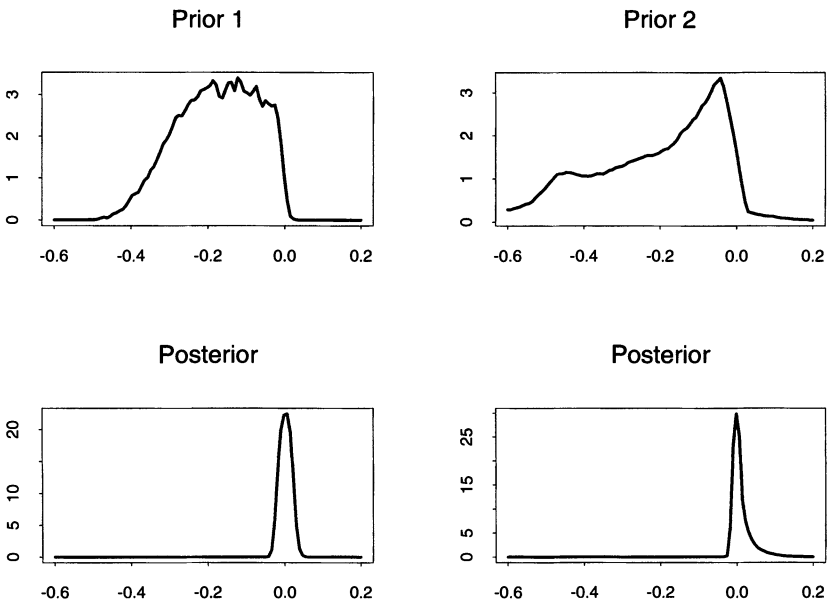

Figure 2
Prior and posterior densities for β_{12}
β_{12} is the 12-quarter autoregression coefficient for returns on the CRSP value-weighted index. Each column reports a prior density for this parameter and the corresponding posterior density. Whereas the priors are fairly diffuse, the posteriors are tightly distributed around 0.

"Prior 1," the prior for ρ_{12} is flat over its range, whereas the posterior is tightly concentrated at 0. Although the large temporary component is stationary by design, shocks to this process are much more persistent than the prior. The median half-life of a shock to z is 81,547 quarters, and the interquartile range spans 39,609 to 198,043 quarters. For purposes of prediction, a stationary process with a half-life of 81,000 quarters (over 20,000 years) may as well be non-stationary.

The combined effect of the variance ratio and the persistence to the stationary process on long-horizon return predictability is seen in Figure 2 (and in the pair of rows labeled β_{12} in Table 1). Here, because the posterior places so much mass very close to 0, the prior and posterior must be placed in different panels because of the scaling of the densities. By examining the top panel of Figure 2, we see exactly the implication of our prior beliefs over the structural parameters on the predictability of returns. For this particular prior, *although the prior forces the temporary component to be very large*, the data suggest that price shocks are essentially permanent.

2.2 Prior 2

Prior 2 roughly corresponds to the alternative evaluated by Richardson (1993, p. 205) [and credited to Poterba and Summers (1988)], where three-fourths of the return variance comes from a predictable component. The parameters used to generate this prior are as follows: $\nu_0 = 5$, $s_0^2 = 0.007$, $\bar{\sigma}_u = 0.014$, $\kappa_0 = 0.01$, $\delta_0 = 0.3$, $\psi_1 = 1.1$, $\psi_2 = 0.1$, $\psi_3 = -0.22$, and $\psi_4 = 0$.

Looking at the lower panel for Prior 2 in Figure 1, we see that the ρ coefficient for the temporary component of 3-year returns is concentrated around $-.5$. Despite this, the data combine with the likelihood to shift our prior so that it is centered over 0. From the upper panel of this figure, we also note that the data shift our beliefs about the proportion of total return variance that is attributable to the differenced random walk from a median of 17% in the prior to 72% in the posterior. Once again, the variance ratio may be misleading, as can be discerned from the panel "Prior 2" in Figure 2. The data shift our beliefs about β from a very diffuse distribution over $-.5$ to 0 in the prior to a tight concentration around 0 in the posterior. Here, the posterior on β (and, of course, ρ) places some mass in the positive line, which is not the case with the prior. As noted above, this is due to the very long half-life of the stationary process. Three years is not a long enough period of time to achieve the limiting results for ρ.

To further isolate this phenomenon, Table 3 contains the prior and posterior densities for β_j and ρ_j as j goes from 4 to 40 quarters. The median ρ in the prior is -42% for 4 quarters and it approaches $-.5$ monotonically as j increases to 40 quarters. The posterior behaves very differently. The median ρ_4 is $-.08$, and at 8 to 40 quarters this value is very close to 0. Notice too from Table 3 that, even in the prior, β_j does not exhibit a U-shaped pattern in j. Rather, β_j in the prior has a median of $-.24$ at 4 quarters, and this declines monotonically in j to $-.11$ at 40 quarters. Also notice that for $j \geq 8$ quarters, the priors for both β and ρ do not include positive values. For j in the range 4 to 40 quarters, in the posterior, β_j is largest in absolute value for j of 4 (where its value is -1.12%). At all other values of j, the median β_j is virtually 0 and the posterior densities are fairly tight around 0.

By comparing the results from priors 1 and 2, we note that a tight prior on the variance ratio can cause the posterior to closely resemble the prior, whereas a less precise prior results in a material shift in the posterior of this ratio. However, for both priors, ρ is shifted toward 0 by the data and the summary measure of predictability, β, is tightly centered on 0. This suggests that the two components of predictability are not separately well identified, but that the combined effect is.

Table 3
Distribution of regression parameters (ρ and β) for prior 2 as a function of the compounding interval

			Percentile				
Variable	Mean	Std dev.	5	25	50	75	95
β_4 (%)	−27.32	28.82	−79.06	−45.91	−23.56	−7.82	14.16
	−0.12	4.55	−5.38	−2.82	−1.12	1.34	9.12
β_8 (%)	−22.82	18.31	−64.76	−37.10	−18.39	−6.93	3.20
	1.26	3.40	−1.70	−0.71	0.08	2.06	8.32
β_{12} (%)	−22.96	20.74	−59.61	−35.44	−17.72	−6.83	−0.83
	0.94	2.57	−1.02	−0.46	0.05	1.43	6.04
β_{16} (%)	−21.28	19.59	−54.03	−34.39	−16.50	−6.30	−1.04
	0.64	2.08	−0.90	−0.42	−0.02	1.01	4.48
β_{20} (%)	−20.68	19.09	−52.08	−33.73	−15.60	−5.76	−1.08
	0.53	1.77	−0.67	−0.31	0.00	0.81	3.57
β_{24} (%)	−20.04	18.72	−51.18	−32.88	−14.58	−5.22	−1.03
	0.41	1.58	−0.57	−0.27	−0.01	0.66	2.90
β_{28} (%)	−19.45	18.31	−49.97	−32.08	−13.93	−4.82	−0.99
	0.34	1.47	−0.50	−0.23	−0.01	0.56	2.45
β_{32} (%)	−18.70	18.10	−49.76	−31.32	−12.89	−4.33	−0.90
	0.28	1.41	−0.43	−0.21	−0.01	0.48	2.12
β_{36} (%)	−18.32	17.96	−49.56	−30.63	−12.32	−4.10	−0.89
	0.23	1.38	−0.39	−0.19	−0.01	0.42	1.84
β_{40} (%)	−17.67	17.61	−49.37	−29.49	−11.39	−3.79	−0.79
	0.19	1.37	−0.35	−0.17	−0.01	0.37	1.64
ρ_4 (%)	−39.53	31.95	−86.15	−62.03	−41.77	−22.27	19.73
	−8.16	20.92	−42.30	−22.75	−8.18	6.35	26.15
ρ_8 (%)	−39.26	24.33	−75.21	−52.03	−43.39	−28.03	6.17
	0.65	13.54	−21.35	−7.55	0.58	8.86	22.64
ρ_{12} (%)	−41.99	20.62	−70.97	−50.79	−46.67	−34.12	−4.80
	0.23	9.50	−15.72	−5.40	0.34	5.75	15.32
ρ_{16} (%)	−43.36	18.55	−67.35	−50.46	−48.33	−38.60	−9.11
	−0.72	8.04	−14.56	−4.97	−0.17	3.97	11.01
ρ_{20} (%)	−44.64	16.99	−65.05	−50.23	−49.27	−41.44	−12.89
	−0.48	6.40	−11.57	−3.75	−0.01	3.22	8.66
ρ_{24} (%)	−45.34	15.64	−63.10	−50.12	−49.63	−43.42	−15.46
	−0.62	5.53	−10.22	−3.28	−0.07	2.59	6.94
ρ_{28} (%)	−45.98	14.59	−61.46	−50.09	−49.85	−45.35	−18.10
	−0.62	4.83	−9.06	−2.87	−0.08	2.19	5.83
ρ_{32} (%)	−46.14	14.26	−59.31	−50.03	−49.92	−46.34	−18.59
	−0.62	4.31	−8.14	−2.54	−0.09	1.87	5.00
ρ_{36} (%)	−46.66	13.33	−58.46	−50.02	−49.96	−47.26	−22.27
	−0.62	3.92	−7.39	−2.29	−0.11	1.63	4.36
ρ_{40} (%)	−46.70	12.99	−57.16	−50.01	−49.98	−47.82	−21.94
	−0.62	3.61	−6.80	−2.09	−0.12	1.44	3.88

See notes to Table 1. β_j refers to the slope of the autoregression of j-quarter returns. ρ_j refers to the slope of the autoregression of the stationary component of j-quarter returns. For each function, the first row refers to the prior and the second row refers to the posterior.

2.3 Optimal forecasts

As noted in the introduction, while β is a measure of return predictability, it is somewhat arbitrary. The maximum predictability (in terms of percentage reduction in the forecast error variance) of k-step ahead forecasts is obtained with $k = 1$. Consider the 1-step ahead forecast error variance reduction constructed by conditioning on the infinite past. This value is plotted and tabulated for prior 2 in Figure 3. This figure shows the prior and posterior densities for the maximum

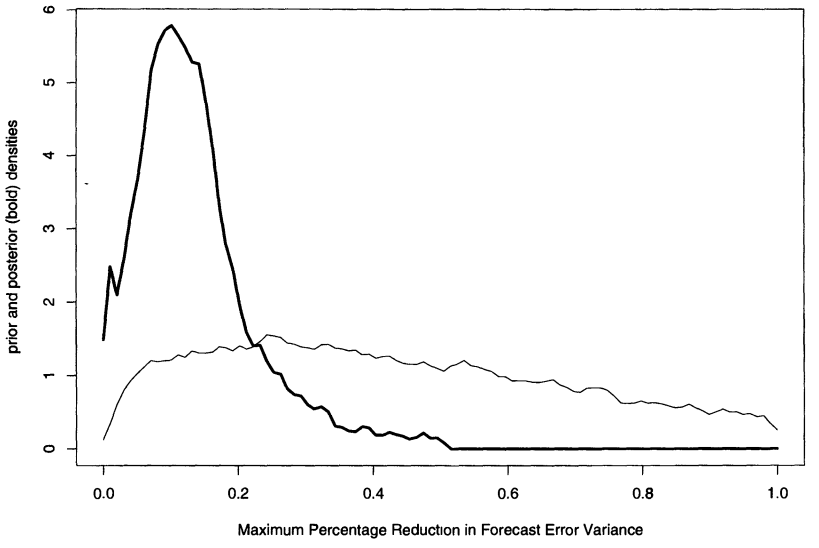

Figure 3
Prior and posterior densities for maximum forecastability under prior 2
Maximum forecastability indicates the forecast error variance, conditioning on the infinite history of returns on the CRSP value-weighted index, as a percent of the unconditional variance. Given that the long-horizon autoregression coefficient of returns is small, we are interested in a more general characterization of predictability. Whereas the prior is relatively flat over the possible range of this ratio, the posterior is more concentrated at 0.

percentage reduction in forecast error variance. We see that the prior is fairly flat over the range 0 to 1, with median .4, whereas the posterior is much more concentrated close to 0, with a median of .04.

We also examine the behavior of the 1-step ahead forecast error variance, as a function of the size of the conditioning set, for prior 2 (as discussed in Section 1.3 above). Figure 4 contains prior and posterior densities for the ratio of the 1-step ahead forecast error variance to the unconditional return variance, constructed by conditioning on n lags, for n equal to 1,2,4, and 8 (quarters). Thus, we see that by conditioning on 2 years of data, the 1-quarter ahead forecast error variance is about 6% smaller than the (unconditional) variance in the posterior, whereas the prior is flat. In particular, with 8 quarter lags, the mean of the posterior density of the ratio of the forecast error variance to the unconditional variance is 92.6%, and the standard deviation of this ratio 5.6%. The corresponding mean and standard deviation in the prior are 50.9% and 31.6%, respectively. We can also see from Figure 4 that most of the information comes from the first two lags. In fact, there is virtually no effect on the posterior mean of this variance ratio of increasing the size of the conditioning set to 300 quarters. For

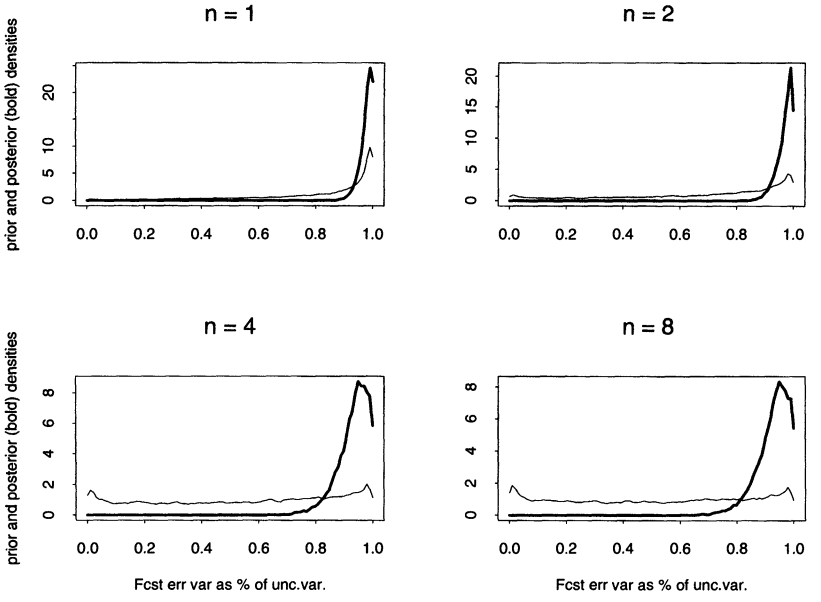

Figure 4
Prior and posterior densities for forecastability as a function of the size of the conditioning set
This figure plots the maximum percent reduction in the 1-step ahead forecast error variance as a function of the conditioning set for prior 2. The conditioning set is either 1, 2, 4, or 8 lags of quarterly returns on the CRSP value-weighted index. The priors are fairly flat, and the posteriors tightly concentrated at 1.

each of the four lag lengths presented in Figure 4, the posteriors are tightly distributed and suggest very little predictability, even though the priors are essentially flat using this metric.

Figure 5 characterizes the prior and posterior densities of the impulse response function for prior 2. This figure contains the ninetieth, fiftieth, and tenth quantiles of the marginal distributions of the impulse response function from both the prior and the posterior. We note that the effects of shocks shrinks over the first 50 quarters monotonically in the prior. On the other hand, there is no indication of any damping over this period from the posterior. This complements the half-life function discussed above. Here we see that after 20 quarters, for example, looking at the median impulse response functions in the prior, more than half of the shock has vanished. In the posterior, there is no indication of any damping. The initial oscillations in the posterior result from the much larger absolute values of the second through fourth autoregressive parameters relative to the prior.

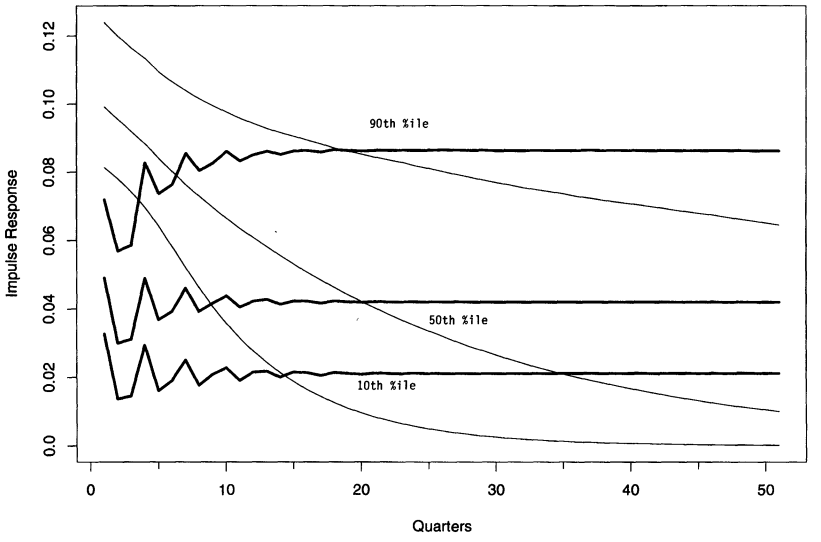

Figure 5
Prior and posterior (bold) densities for impulse response functions
The impulse response function indicates the extent to which an exogenous shock to the CRSP value-weighted index quarterly returns damps with the passage of time. Here we see that in prior 2 after 50 quarters, a shock has almost no residual effect. In the posterior however, there is no indication of any damping after 50 quarters. The posterior indicates that there is little predictability in returns because putatively temporary shocks persist almost indefinitely.

2.4 Inference

Since we have assigned zero point mass to the event $\frac{\sigma_x^2}{\sigma_u^2} = 0$, we are not in a position to make inference in the formal sense that this event is true. As we can see from the posterior densities, and as noted by Sims (1988) in the context of the unit root test, there is an arbitrariness associated with this event, which in any case is not worthy of special attention. Furthermore, as noted by Sowell (1991, p. 257), we have considered "only a small set of priors [and hence,] interpretations can only be claimed to hold conditional on restricting attention to the priors considered. Because this set is not exhaustive, the conclusions cannot be claimed in general." With these caveats in mind, several conclusions appear reasonable. First, the relative importance of the two components is not well identified. However, the persistence of shocks to the stationary component in all cases considered is shifted by the likelihood to make these processes virtually nonstationary. This provides an example of Sims's (1988) point about the knife-edge nature of classical unit root tests. A stationary process with a half-life of

26,000 quarters hardly gives rise to predictable patterns in our data.[10] This lack of predictability can also be observed in the flat impulse response functions of the posterior. Once again this is manifest in the posterior densities of the β_{12} coefficients, which are tightly distributed around 0 in all cases. Unlike the priors, these β densities do suggest that small positive values are possible, which is further indication of the extreme persistence in the posteriors. Finally, although the posterior distributions of β suggest that long-horizon returns are not mean reverting, there is evidence that the prediction error of next quarter's return may be reduced somewhat by using the parameters from the model (a finding that is inconsistent with a random walk).

2.5 Numerical accuracy

As noted above, it is possible to characterize the numerical efficiency of the Gibbs sampler (i.e., the quadrature) by evaluating the series of draws for a particular function of interest. To demonstrate this and to provide the reader with a sense of the numerical accuracy of our procedure, we present the numerical standard errors (nse) and relative numerical efficiencies (rne) as described in Section 1 above, for the posterior densities obtained for both σ_u and $\frac{\sigma_u^2}{\sigma_r^2}$.[11] Using the Daniell window to approximate the spectral density at frequency 0 (of draws 1,001 to 11,000) from prior 1, the nse (nse_D) on σ_u is .000094 and the rne (rne_D) is .35. The corresponding values using a Parzen window are $\text{nse}_P = .000083$ and $\text{rne}_P = .4476$. The posteriors from prior 2 yield $\text{nse}_D = .00026$, $\text{rne}_D = .25$, $\text{nse}_P = .00022$, and $\text{rne}_P = .34$. Numerical efficiency is somewhat worse when we examine the function of interest, $\frac{\sigma_u^2}{\sigma_r^2}$. Here from prior 1 we obtain $\text{nse}_D = .00069$, $\text{rne}_D = .24$, $\text{nse}_P = .00059$, and $\text{rne}_P = .33$. Finally, from prior 2, $\text{nse}_D = .0031$, $\text{rne}_D = .24$, $\text{nse}_P = .0026$, and $\text{rne}_P = .32$.

These values show that the preceding analysis of the posterior densities is not affected by numerical inaccuracy, and quantify exactly the sense to which the results are affected by numerical imprecision.

2.6 Reconciliation with earlier findings

Using a variety of proper priors we have seen that the posterior density of β is tightly centered around zero. This may seem inconsistent

[10] Note that the nature of the posterior does not depend on whether we specify an AR(2) or an AR(4) process for z. We do not explore the question of model specification (vis-à-vis m) as the posteriors do not seem to depend on m.

[11] Estimation in this context is a nontrivial computational matter. The 11,000 draws from the posterior require approximately 40 hours of CPU time on a Sparc 10. Note however that once the draws have been obtained (and stored on disk), analyzing any functions of interest, including the Markov properties of the draws themselves, is a simple matter.

with earlier results in the literature that find point estimates for β_{12} around $-.29$. In this subsection, we specify a diffuse prior over the hyperparameters. For this prior the posterior density closely resembles the sampling distribution for β [Berger (1985), p. 137]. We show that the diffuse prior on the hyperparameters does not translate into a diffuse prior on β however. The diffuse prior states

$$P_0(\boldsymbol{\theta}) \propto \frac{1}{\sigma_\epsilon \sigma_u}. \tag{35}$$

Analysis and estimation proceed as above, with different conditional posterior densities resulting from the application of Bayes' rule. Here, we specify an AR(4) as with the two proper priors reported above. The posterior mean of β_{12} in this case is $-.23$, with a standard deviation of .14. The fifth and ninety-fifth percentiles of the marginal posterior are $-.48$ and $-.05$, respectively. The posterior mean (standard deviation) of $\frac{\sigma_u^2}{\sigma_r^2}$ is .35 (0.22), and of $\rho(12)$ is $-.485$ (0.02). These results are similar to frequentist findings and suggest that the data have virtually no information about β (the summary measure of predictability). This result suggests that the findings using informative priors above are not the result of a likelihood function that fails to characterize the properties of the data, rather they result from a specific difference between the informed priors and the diffuse prior.

In order to understand this set of results, we need to analyze what effect the diffuse prior has on the functions of interest. Although our intuition might suggest that this diffusiveness carries over to functions of the hyperparameters, this may not be so. To analyze the effect of a diffuse prior on β we can simulate from the priors (as above), although here, this analysis is suggestive and not exact, as the integrals do not formally exist. Conceptually, we could draw from diffuse priors on σ_ϵ and σ_u by exponentiating a draw from a uniform distribution over $-\infty$ to ∞. We draw ψ from the uniform (closed) interval $[-3, 3]$ independently, and discard those draws where the largest eigenvalue exceeds 1. The effect of this on β^{12} is presented in Figure 6. Here we see that the diffuse prior on the hyperparameters has a peculiar effect on the prior of β. Roughly half of the distribution is at 0 and the remaining half of the prior is at $-.5$. Clearly, this prior is far from diffuse on β and it seems unappealing for the basis of decision making. The intuition for this outcome is clear once we see the figure. The uniform densities on the variances suggests that half of the time, one of the two variances will be infinitely large relative to the other.

As we would expect, the classical findings may be replicated by an appropriate choice of prior. The results reported above using proper priors are materially different from earlier results using frequentist

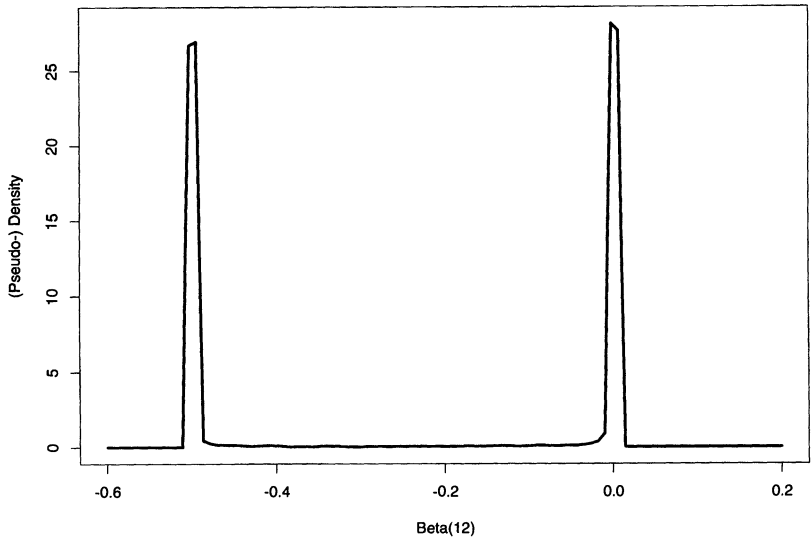

Figure 6
The implication of a diffuse prior on model hyperparameters for β_{12}
When a diffuse prior is specified for the hyperparameters, the implications for the autoregression coefficient for 3-year returns is not obvious. In fact, half of the mass is placed at the minimum value of −0.5 and half the mass at the maximum value of 0. Not surprisingly, then, when a diffuse prior is used, the data appear to have little information about long-horizon predictability.

tools. This subsection makes the case that this difference is not the result of an inappropriate model or likelihood function, rather it arises from the peculiar (and somewhat unnatural) prior implicit in frequentist analysis.

Conclusions

The question of whether shocks to a time series persist indefinitely, or whether significant mean reversion is present, is clearly important for both statistical and economic interpretation of data. This question has been explored using classical (frequentist) tools, and these tools have been criticized. In particular, conclusions were reached that the available data is inadequate to address the question of whether returns consist of a material stationary component (which is quickly damping). Bayesian or subjective statistical analysis provides a rigorous means of examining the question of how much information is in a particular sample. The subjectivist treats the data as fixed and the model parameters as random variables. In fact, by using Bayesian analysis, we are able to identify the source of imprecision of classical estimation. The diffuse prior is tantamount to saying that half of the

time the stationary component is infinitely large relative to the nonstationary component, and the other half of the time the opposite is true.

Moreover, a significant appeal of the Bayesian time-series decomposition used herein is that the entire posterior distribution of any function of interest is available, conditional on the data that were observed. Thus, we observe precisely the manner in which the data and the likelihood function combine to yield precision about a particular parametric restriction. This is in sharp contrast to frequentist estimation, in the presence of a small number of *independent* observations, where statements about the power and size of various tests are ad hoc. We also demonstrated another attractive feature of a Bayesian approach to this problem: easy computation of the posterior densities of functions of interest. Not only is this generally infeasible in a frequentist setting, there is controversy about the manner in which the limits should be taken.

For the class of priors we used, we found that the data are not mute on the question of long-horizon predictability. The summary measure of predictability in long-horizon returns (β) is generally very tightly concentrated around 0 in the posteriors whereas it is more diffuse in the priors. This results from the fact that the half-life of shocks to the stationary component, for all priors considered, is so long as to render the distinction between predictable and unpredictable over reasonable forecast horizons moot, a phenomenon that may also be gleaned from the posterior densities of the impulse response functions, which show no damping over the first 50 quarters. Finally, while β is an ad hoc measure of predictability, we note that the data cause us to revise our priors on the maximum reduction in forecast error variance toward 0 (and also tighten our beliefs) for all sizes of conditioning sets.

References

Baillie, R. T., 1987, "Inference in Dynamic Models Containing 'Surprise' Variables," *Journal of Econometrics*, 35, 101–117.

Berger, J. O., 1985, *Statistical Decision Theory and Bayesian Analysis* (2nd ed.), Springer-Verlag, New York.

Brockwell, P. J., and R. A. Davis, 1991, *Time Series: Theory and Methods*, Springer-Verlag, New York.

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

Cochrane, J. H., 1992, "Permanent and Transitory Components of GNP and Stock Prices," working paper, University of Chicago.

Cochrane, J. H., and A. M. Sbordone, 1988, "Multivariate Estimates of the Permanent Components of GNP and Stock Prices," *Journal of Economic Dynamics and Control*, 12, 255–296.

Davies, R. B., 1977, "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 64, 247–254.

Davies, R. B., 1987, "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 74, 33–43.

Fama, E. F., 1991, "Efficient Capital Markets: II," *Journal of Finance*, 46, 1575–1617.

Fama, E. F., and K. R. French, 1988, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246–273.

Geweke, J. F., 1989, "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* 57, 1317–1339.

Geweke, J. F., 1992, "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), *Bayesian Statistics 4*, Oxford University Press, Oxford.

Harvey, A. C., 1985, "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics*, 3, 216–227.

Nerlove, M., D. M. Grether, and J. L. Carvalho, 1979, *Analysis of Economic Time Series*, Academic Press, San Diego, Calif.

Poterba, J., and L. Summers, 1988, "Mean Reversion in Stock Prices: Evidence and Implications," *Journal of Financial Economics* 22, 27–59.

Priestley, M. B., 1981, *Spectral Analysis and Time Series*, Academic Press, San Diego, Calif.

Quah, D., 1992, "The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds," *Econometrica*, 60, 107–118.

Richardson, M., 1993, "Temporary Components of Stock Prices: A Skeptic's View," *Journal of Business and Economic Statistics*, 11, 199–207.

Richardson, M., and J. H. Stock, 1989, "Drawing Inferences from Statistics Based on Multiyear Asset Returns," *Journal of Financial Economics*, 25, 323–348.

Samuelson, P., 1988, "Longrun Risk Tolerance When Equity Returns are Mean Regressing: Pseudoparadoxes and Vindication of 'Business Man's Risk," mimeo, Massachusetts Institute of Technology.

Sims, C. A., 1980, "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Sims, C. A., 1988, "Bayesian Skepticism of Unit Root Econometrics," *Journal of Economic Dynamics and Control*, 12, 463–474.

Sowell, F., 1991, "On DeJong and Whiteman's Bayesian Inference for the Unit Root Model," *Journal of Monetary Economics*, 28, 255–263.

Watson, M., 1986, "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, 18, 49–75.