A Collection of Essays in Empirical Finance

by

Kenneth Deon Roskelley

A Dissertation Submitted to the Faculty of the

Committee on Business Administration

In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy
with a Major in Management

In the Graduate College

The University of Arizona

2002

UMI Number: 3053911

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Kenneth Deon Roskelley

entitled A Collection of Essays in Empirical Finance

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Chris Lamoureux	5-21-02 Date
George Jiang	5/16/02 Date
Price Fishback	5/14/02 Date
Stanley Reynolds	5/16/02 Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director
Chris Lamoureux

5-21-02
Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Signed: [signature]

ACKNOWLEDGMENTS

I would like to thank Price Fishback, Bill Horrace, George Jiang, and Stan Reynolds for serving on my committee and giving me many valuable suggestions and support. In particular I would also like to thank Chris Lamoureux for serving as my advisor and spending innumerable hours helping me to see the forest instead of the trees.

Table of Contents

List of Tables

ABSTRACT

This dissertation consists of three papers. The first assesses the ability of bivariate distribution models to explain the contemporaneous and autocorrelation between volume and volatility. GMM is used to fit first and second moments of the model to the data and analyze the model's fit. The second paper looks at the uncertainty surrounding cost recovery in regulated utilities. Stock market data is used to ascertain the market's perception about the deregulation of electricity in the United States. The third and final paper looks at the economic evidence for a stochastic opportunity set from an investor's point of view. A Bayesian investor must allocate her wealth between a risky and a risk free asset after observing market data when the model for asset returns is unknown and returns are potentially predictable.

Chapter 1

A Generalized Method of Moments Test of the Mixture of Distributions Hypothesis when the Latent Variable is Autoregressive

1.1 Introduction

The empirical correlation between volume and return is a well known characteristic of stock market data. yet there have been few theoretical explanations as to why volume and returns should be correlated. Traditional asset pricing models such as the CAPM and APT rely primarily on some form of absence of arbitrage to price assets but are silent on the role trading volume plays in the market. The importance of understanding the relation between volume and return is explained in Karpoff (1987). who also provides a review of the evidence on the contemporaneous correlation between return. return volatility. and volume.

Of course the principal reason for trying to understand the relation between volume and return is to have a more general understanding of financial markets. In particular it is of interest why the underlying price of an asset should be related to the number of times the asset changes hands. In more practical terms. the effect of volume on prices could have implications about the costs and benefits of speculative trades and the desirability of market regulations such as circuit breakers. Likewise. returns are notoriously noisy and understanding the joint distribution between volume and returns could possibly improve the power of event studies or tests for predictability in returns.

Despite the importance of understanding the relationship between return and volume, models such as Admati and Pfleiderer (1988), Foster and Viswanathan (1990), Anderson (1996), and Tauchen and Pitts (1983) have either been impossible to test directly, found little empirical support, or have not been well enough developed to supplant current models such as the CAPM or APT in terms of understanding asset returns and financial markets in general. This has resulted in much of the empirical work taking an atheoretical approach to exploring relationships in the data. Granger and Morganstern (1970), for instance, report that the squared difference between the daily open and close is correlated with volume while Crouch (1970) finds that the correlation between daily volume and the absolute value of daily price changes is positive. Such results suggest that understanding how and why volume is related to return volatility is essential to form a better understanding of option pricing.

The finding that daily volume is correlated with squared price changes has led to the development of Mixture of Distributions Models, or MODM. One advantage of the MODM class is that the models also provide an explanation for why returns do not appear to be normally distributed, but instead exhibit fat tails in their empirical distributions. Fat tails in asset returns are potentially troublesome given that many financial models imply that asset returns have a discrete-time normal distribution derived from their continuous-time properties. Understanding the deviation between traditional financial models and the observed data is essential for the development of more accurate models. Since the MODM has at its heart a mixing over distributions. the models generate fat tailed distributions for returns and are candidates for explaining the observed deviation from theoretical predictions.

Clark (1973) attempts to explain the leptokurtosis in returns as the result of the randomness of the number of within-day transactions which are themselves linked to the number of price changes. In his model the asset price's change is the sum of the within-day price changes. The trading volume is thus related to the number of within-day transactions. causing trading volume to be correlated to the volatility. Harris (1986) provides circumstantial evidence supporting the MODM by comparing sample conditional and unconditional moments to those predicted by the MODM. finding that both kurtosis and skewness in the data are adequately explained.

In the Clark (1973) model. however. information flow must be made simultaneously to all investors. This contrasts to the model developed in Tauchen and Pitts (1983) where the model is motivated by a latent process that enters into both the return and volume generating process. resulting in their being jointly distributed. The model consists of general information flows to the market and information flows to individual investors. but the information flows are not necessarily simultaneous. The model implies both conditional and unconditional moments for the data and also accounts for growing markets by positing that volume increases linearly with the number of traders. Tauchen and Pitts (1983) estimate the model's parameters using maximum likelihood on data from T-bill futures markets and find the parameters to be consistent with economic intuition.

Yet another stylized fact that can be justified by MODM is the relationship between volume and squared return. The MODM of Epps and Epps (1976) gives the price change for each within-day transaction as the average change in investors' reservation prices. They assume that the magnitude of disagreement about prices amongst investors is positively related to the magnitude of the change in the asset's price. The implication is that the volume-volatility relationship is caused by heterogenous beliefs about the asset's value. and the more disparate those beliefs are the more volume and the larger price changes will be.

The volume-volatility relationship has received considerable empirical attention over the last decade stemming from the development of ARCH and GARCH models. For instance Nelson (1990) illustrates that Clark's model can be re-written as the discrete time version of a continuous time exponential ARCH model. Lamoureux and Lastrapes (1990) analyze the question of endogeneity between return and volume. presenting evidence that volume is a sufficient statistic for squared returns in a GARCH specification. More recently GARCH models have encountered competition from stochastic volatility models where the return volatility follows a stochastic specification. The difficulty in using stochastic volatility models has been that they are random error models. but the errors are not distributed iid. This means the likelihood function is a T-dimensional integral which is prohibitively difficult to analyze either analytically or numerically. Jacquier. Polson. and Rossi (1994) use a Monte

Carlo Markov Chain method to form posterior distributions for a stochastic volatility model in a Bayesian approach. While their paper illustrates the promise of using stochastic volatility models in explaining and understanding stock returns and volatility, it is still uncertain what approach is the most appropriate since we must always be concerned with over-fitting the data.

One of the appeals to MODM models is that they have a natural interpretation as stochastic volatility models when volume is not conditioned upon. Due to the presence of the latent process which serves as the mixing variable, however, estimation is not straightforward since the model is like a random error model. Given the difficulty of working with the models' likelihood, little has been done to test the MODM directly. Tauchen and Pitts (1983) is one example of such a MODM. Though the model has circumstantial support, in order to test the model directly the latent variable must be extracted or integrated over. Lamoureux and Lastrapes (1994) employ a signal extraction technique to capture the latent process and then test if conditioning on the extracted variable does away with GARCH effects in the data. After extracting the latent variable and conditioning on it they conclude that the MODM fails to fully account for the GARCH effects in the data. Other empirical studies of MODM include Anderson (1996) who tests a MODM using GMM and finds a good fit to the data, while Watanabe (2000) analyzes the same class of models using Bayesian techniques and finds strong evidence that the models are rejected by the data. Similar to Anderson (1996), Richardson and Smith (1994) use GMM to estimate the properties of the latent process and test the fit of the MODM to the data. They conclude, unlike Anderson (1996), that the MODM does a poor job at fitting the data and that the latent variable seems to be positively skewed and kurtotic.

This paper re-evaluates the GMM test of the Tauchen and Pitts (1983) model taken by Richardson and Smith (1994) by using different moment conditions to over-identify the model. It is known that GMM is sensitive to the moments used to over-identify the model and it is important to understand in what dimension the MODM provides a poor fit. While Richardson and Smith (1994) use only contemporaneous moments in fitting the MODM. I over-identify the model using the implied contemporaneous moment conditions and certain time-series moments. The over-identifying moments are then used to evaluate the fit of MODM to the data. Even though the GMM Q-statistic on the over-identifying moments is known to be asymptotically distributed as a chi-square. the estimation is highly non-linear and potentially poorly behaved in finite samples. To ascertain the small sample properties of the Q-statistic I run simulations calibrated to the data in order to form the finite sample distribution of the statistic.

In the next section I discuss the motivation behind choosing different moments and why my results will be potentially different from those of Richardson and Smith (1994). In Section 1.3 I discuss the general estimation problem. including the choice of the weighting matrix and the test statistic for the over-identifying moments. I then present the results in Section 1.4 and discuss the implications of the results in Section

1.5.

1.2 Motivation

As pointed out in Lamoureux and Lastrapes (1994), Tauchen and Pitts' MODM can be written as

$$r_t = \sigma_1 Z_{1t} \sqrt{F_t}$$

$$V_t = \mu_2 F_t + \sigma_2 Z_{2t} \sqrt{F_t}$$

where r_t is the return on the stock at time t, V_t is the stock's trading volume at time t, μ_2 represents the unconditional mean of volume, Z_{1t} and Z_{2t} are random variables with zero mean and unit variance, and F_t is a mixing variable that accounts for the contemporaneous correlation between squared return and volume. Z_{1t}, Z_{2t}, and F_t are all assumed to be mutually independent, and hence the covariance between squared return and volume conditional on the unobserved mixing variable F_t is zero. This implies that volume and squared return are uncorrelated when we condition on F_t, meaning that the true value of an asset is not in fact determined by the number of times it trades. At the same time, though, the model allows squared return to be unconditionally correlated to volume. It still remains, however, to explain the nature of the latent process F_t.

The most intuitive explanation of the latent process is that of an information flow. It seems plausible that volume and price would move jointly because of information shocks about the future earnings or prospects of the stock. In this context the variable F_t represents only the amount of information arriving to the market. explaining why F_t must be non-negative. The determination as to the nature of the information itself, i.e. if it's good or bad and its magnitude, will be determined by the variable Z_{it}. This also explains why Z_{it} is mean zero since one would expect new information to be as equally likely to be good as bad. The model implies specific conditional and unconditional moments which can be tested. Since the mixing variable F_t is not directly observable. it is convenient to work with the model's unconditional moments. Richardson and Smith (1994) derive and use the first, second, and third moments and cross moments of the model to over-identify the model's parameters. They do, nevertheless, make one deviation from the MODM presented in Tauchen and Pitts by allowing return to have a non-zero mean. They do so by re-writing returns as

$$r_t = \mu_1 F_t + \sigma_1 Z_{1t}\sqrt{F_t}$$

Writing returns in this manner has the advantage that the moments now over-identify the parameters without requiring any specific distributional assumptions on Z_{1t}, Z_{2t}, or F_t. The nine moment conditions[1] used are in fact

[1] Tauchen and Pitts (1983) note that standardizing F_t to have unit mean is observationally equivalent to the specified MODM. The only difference will be the scale of the other parameters. The following moment conditions are for the transformed model.

$$
\begin{aligned}
E[r_t] &= \mu_1 && (1.1)\\
E[V_t] &= \mu_2 && (1.2)\\
E[(V_t-\mu_2)^2] &= \sigma_2^2+\mu_2^2\sigma_f^2 && (1.3)\\
E[(r_t-\mu_1)^2] &= \sigma_1^2+\mu_1^2\sigma_f^2 && (1.4)\\
E[(V_t-\mu_2)(r_t-\mu_1)] &= \mu_1\mu_2\sigma_f^2 && (1.5)\\
E[(V_t-\mu_2)^3] &= 3\mu_1\sigma_1^2\sigma_f^2+\mu_1 m_3 && (1.6)\\
E[(r_t-\mu_1)^3] &= 3\mu_2\sigma_2^2\sigma_f^2+\mu_2 m_3 && (1.7)\\
E[(V_t-\mu_2)^2(r_t-\mu_1)] &= \mu_1\mu_2^2 m_3+\sigma_2^2\mu_2\sigma_f^2 && (1.8)\\
E[(V_t-\mu_2)(r_t-\mu_1)^2] &= \mu_1\mu_2^2 m_3+\sigma_2^2\mu_1\sigma_f^2 && (1.9)
\end{aligned}
$$

As can easily be corroborated. if the non-zero mean term of $\mu_1 F_t$ is left off. then the right hand side of equations 1 and 5 through 8 become either zero or unknown unless some type of distributional assumptions are imposed on Z_{lt}. Z_{2t}. and/or F_t. This. however. would prohibit us from testing the MODM directly. and instead we would be testing the MODM jointly with our distributional assumptions. On the other hand if we don't include the non-zero mean we are left with only four variables in four equations. meaning we have no over-identifying restrictions to test.

Before proceeding either way it is important to first understand what economic interpretation can be given to the term $\mu_1 F_t$ that Richardson and Smith add to the model. At first glance the new term allows for a non-zero average return which is consistent with theory and empirics. However. continuing with the intuition that the mixing variable is an unobserved information flow. it becomes apparent that in the specification $\mu_1 F_t$ information can only positively effect an asset's mean return. even if the information received were adverse. Whereas information flows. good or bad.

could increase the mean trading volume, the same is likely not true for an asset's return. This makes the addition of the term to the model rather unappealing.

Re-specifying the mean to be μ_1, independent of the latent variable, does not solve our problem. Such a specification does not result in over-identifying restrictions on the parameters[2]. Failure to add the term will force the specification of the distribution of the latent process in order to obtain sufficient moments to carry out inference since otherwise we are left with four moments and four unknowns. With only four equations and four parameters it is not possible to use the Wald statistic developed by Hansen (1982) to test the MODM since there are no over-identifying restrictions to be tested. Though Richardson and Smith present evidence for the most part contrary to the MODM, it is not clear that the model they test is consistent with that of Tauchen and Pitts.

One other possibility for obtaining over-identifying restrictions, however, is to somehow include the well documented persistence in conditional variance of return and volume. First documented by Mandelbrot (1963), evidence on persistence in conditional variance has of late been supplied from ARCH/GARCH models (Bollerslev, Chou, and Kroner (1992)). Persistence in volatility is a well known empirical occurrence, but explaining the reasoning for its existence has been as troublesome as explaining the contemporaneous relation between volume and return. Despite the usefulness of ARCH models in modeling market phenomena, ARCH itself is not the-

[2] That is precisely why Richardson and Smith (1994) and Harris (1986) specify mean return as $\mu_1 F_t$.

oretically motivated and provides no insight into why such persistence in conditional variance is evident in market data. Any model that proposes to meaningfully explain the data generating process of variance and return should be able to explain both the contemporaneous correlation between volume and return and the persistence in the variance of returns and volume.

If we assume that the mixing variable F_t follows a first-order auto-regressive process because of information lags we could explain the serial dependence in return volatility and volume. Such a specification will maintain the conditional independence of volume and volatility while providing an explanation of persistence. This of course means that a rejection of the model could be due purely to misspecification of the latent process. Given that over-identification of the model parameters is impossible without further assumptions there is little else but to specify the information flow's distribution or the latent process itself. Following Lamoureux and Lastrapes (1994) and Watanabe (2000) who specify the mixing variable as AR(1). I construct moment conditions given that F_t follows such a process. This allows me to avoid specifying restrictive distributional assumptions about the information flow while incorporating into the model the contemporaneous correlation and autocorrelation between volume and volatility.

1.3 Empirical Approach

As in Lamoureux and Lastrapes (1994) I specify the latent process to be

$$F_t = \alpha_0 + \alpha_1 F_{t-1} + \varphi$$

where φ_t is a mean zero random variable. Since an AR(1) process has unconditional mean equal to

$$\alpha_0/(1 - \alpha_1)$$

we can normalize the F_t process to have unit mean by specifying α_0 as

$$\alpha_0 = (1 - \alpha_1)$$

This specification generalizes the model's assumption that the information flow is serially uncorrelated. Explaining and justifying such a process is more difficult.

One possible explanation is that the Tauchen and Pitts' model does allow for non-simultaneous information flows to individuals. Such a specification could be consistent with a model where traders update their beliefs about the private information flows of other traders as they observe price changes. Such an interpretation is similar to recent theoretical models like Anderson (1996) where information asymmetries between market participants exist. Another possible justification for the specification is a market where portfolio reallocations are sticky because of non-trivial trading costs. One appealing aspect of the AR(1) process is that it induces a non-zero auto-covariance for squared return. a stylized fact first noted in the literature by

Mandelbrot (1963). The AR(1) process provides us with three more moments and at the same time only one more parameter, α, the autocorrelation of the information flow. This leaves us with five parameters and seven equations[3]. In particular the moments I use are:

$$\begin{aligned}
E[V_t] &= \mu_2 && (1.10)\\
E[(V_t-\mu_2)^2] &= \sigma_2^2+\mu_2^2\sigma_f^2 && (1.11)\\
E[r_t^2] &= \sigma_1^2 && (1.12)\\
E[(V_t-\mu_2)(r_t^2-\sigma_1^2)] &= \mu_2\sigma_1^2\sigma_f^2 && (1.13)\\
E[(r_t^2-\sigma_1^2)(r_{t-1}^2-\sigma_1^2)] &= \alpha\sigma_f^4\mu_2^2 && (1.14)\\
E[(r_t^2-\sigma_1^2)(V_{t-1}-\mu_2)] &= \alpha\mu_2\sigma_1^2\sigma_f^2 && (1.15)\\
E[(V_t-\mu_2)(V_{t-1}-\mu_2)] &= \alpha\mu_2^2\sigma_f^2 && (1.16)
\end{aligned}$$

Equation 1.13 shows that in the current specification of the MODM there is a positive correlation between volume and volatility, as is observed in the data. In general a more volatile information flow will increase that relation. Equation 1.14 also allows for persistence in volatility, similar to a GARCH specification. Notice that the higher the autocorrelation of the information flow is, the more persistent volatility will be. Equation 1.16 similarly allows for persistence in volume that is positively related to the degree of autocorrelation in the information flow. Equation 1.15 shows that one of the implications of the information flow following an AR(1) process is that future volatility can be more accurately predicted given current volume, a relationship that would certainly be useful in pricing close to expiration options. The moment

[3] I force the sample mean to be zero by de-meaning the data, and hence assure that this particular sample moment is equal to it's hypothesized mean.

conditions are used to estimate the model parameters and form an over-identifying test of the restrictions implied by the model following Hansen (1982).

1.3.1 The GMM Over-Identifying Test

For simplicity we can think of the above moment conditions as an $Rx1$ vector $h(X_t, \theta)$. where

$$X_t = (r_t, V_t)$$

and θ represents the M model parameters. which in our case are α. σ_1^2. σ_2^2. σ_f^2. μ_2. If the series X_t conforms to the MODM. then the expectation of $h(X_t, \theta)$ at any point in time is

$$E[h(X_t, \theta)] = 0$$

which gives us 7 residuals in each time period t. In large samples. given that X_t does in fact conform to the hypothesized model. the sample moments should approximate zero as $T \to \infty$.

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} h(X_t, \theta) \to 0$$

The motivation behind GMM is to choose the model's parameters such that $g_T(\theta)$ is equal to zero. In finite samples and with over-identifying restrictions (when $R > M$). this will not be possible. Some function of the errors instead must be minimized. Similar to Generalized Least Squares. GMM seeks to minimize the quadratic

$$Q = g_T^{'}(\theta)\Sigma^{-1}g_T(\theta) \tag{1.17}$$

where Σ^{-1} is an RxR positive, semi-definite matrix. Hansen (1982) proves that the efficient choice amongst Σ is to use the covariance matrix of the residuals. Note that the first order condition will be

$$G'\Sigma^{-1}g_T(\theta) = 0$$

where G' is the MxR matrix of first derivatives of $g_T(\theta)$ with respect to the the vector θ. This leaves us with M equations in M unknowns, which will be uniquely identified. Hansen also proves the asymptotic distributions of the parameter estimates and the test statistic Q from equation 1.17 to be[1]

$$\begin{aligned} &\sqrt{T}(\widehat{\theta} - \theta) \overset{a}{\sim} N(0, G'\Sigma^{-1}G) \\ &TQ \overset{a}{\sim} \chi^2_{R-M} \end{aligned}$$

In general, since Σ^{-1} will not be known, it will be necessary to estimate the covariance matrix of the residuals. The asymptotic distribution of the parameter estimates and Q do not depend on the efficiency of the covariance matrix estimator, but they do require consistency. Hence one possible and frequently used option is to use the sample estimates. One difficulty with such an approach is that the estimated covariance matrix is not required to be positive semi-definite which could result in a negative realization on Q. Also, as already mentioned, though efficiency is not required, it is desirable. For these two reasons I use Newey and West's (1987) estimator of the covariance matrix.

[1]In the case there are k moments being estimated, the results depend on the ergodicity and stationarity in return and volume along with the existence of the first 2k moments.

Besides being more efficient than the sample estimate. the Newey West estimate of the covariance matrix is guaranteed to be positive semi-definite. The estimator for Σ uses sample information to calculate the contemporaneous covariance matrix $\widehat{\Sigma}_0$ and the first j order auto-covariance matrices $\widehat{\Sigma}_j$, summing them together to get the estimate

$$\begin{aligned} \widehat{\Sigma} &= \widehat{\Sigma}_0 + \sum_{k=1}^{m} w(k,m)(\widehat{\Sigma}_k + \widehat{\Sigma}_k^T) \\ w(k,m) &= 1 - \frac{k}{m+1} \end{aligned}$$

where m is the bandwidth. and w is a weight.

The next obvious problem is how to pick the bandwidth. m. To do this I follow Andrews' (1991) procedure

$$\begin{aligned} m &= 1.1447(T\alpha(1))^{1/3} \\ \alpha(1) &= \frac{\sum_{a=1}^{R} \frac{4\rho_a^2\sigma_a^4}{(1-\rho_a)^6(1+\rho_a)^2}}{\sum_{a=1}^{R} \frac{\sigma_a^4}{(1-\rho_a)^4}} \end{aligned}$$

where ρ is the first-order autocorrelation of the a^{th} moment's residuals. Since m will not. in general. be an integer. I follow convention and round m to the next whole number. In order to check the robustness of the results to the bandwidth estimate. I recalculate the Q-statistic for alternative choices of m.

1.3.2 General Estimation Problems

As Richardson and Smith (1994) and Lamoureux and Lastrapes (1994) note. there is difficulty in estimating variance from the model's moment conditions since it is common for the estimates of the variances (in particular σ_2^2) to be negative in the context of this model. This leads to the question posed by Lamoureux and Lastrapes (1994) as to how much information is truly obtained from the hypothesized moments about σ_2^2. At the same time. even though Hansen proves the Q-statistic will be distributed as a chi-square asymptotically. the same may not be a good approximation in finite samples. Though sample sizes of 35 to 40 in linear models are often thought to mimic asymptotic results. the same is not true in the non-linear case. When the model is highly non-linear. as it is in the case of the MODM. the same may not be true for sample sizes of ten thousand or even more.

For these two reasons I perform simulations to ascertain the behavior of the Q-statistic in a sample size equal to that of the observed data. I calibrate the simulations to the observed data's grand mean parameter estimates for all companies. The simulations are driven by Z_{1t}. Z_{2t}. and ϕ. The first two are taken to be standard normals. consistent with the theory of most MODM since the errors are typically derived as the sum of random variables. ϕ is simulated using a variety of distributions but results are reported using a truncated normal distribution since the value for F_t must not be negative.

For consistency with Richardson and Smith (1994), my sample is the thirty Dow Jones Industrial Average firms[5] from 1982 to 1986. I use daily returns (close to close, dividend adjusted) and volume (adjusted for splits and stock dividends), both taken from the Center for Research in Security Prices (CRSP) tapes. The sample period consists of 1,262 daily observations per firm. I also check the robustness of the sample by looking at the same thirty firms from 1988 to 1991.

1.4 Results

Table 1.1 contains a comparison between Richardson and Smith's Q-statistic and the one calculated under the current specification of the MODM over the 1982-1986 time period. It is important to note that the Richardson and Smith Q-statistic is distributed with three degrees of freedom while the Q-statistic from the current treatment is chi-square with two degrees of freedom.

For the current specification of the MODM twenty-two of the thirty firms reject the MODM at the 5% confidence level compared to only fifteen for Richardson and Smith. All but two companies that reject the MODM at the 5% level in the Richardson and Smith (1994) specification also reject the MODM in the current specification. In fact the data for nine firms reject the MODM even at the 1% level in both specifications. Respecifying the return process and considering the time series moments has not

[5] Richardson and Smith (1994) look at all 30 firms from 1982 to 1986. I have excluded three due to missing information, delisting, or no information on stock splits.

improved the fit to the data at all. It is difficult to know if the improved fit in Richardson and Smith (1994) is attributable to the different specification of mean return or the absence of the time-series moments.

Table 1.2 reports the estimated auto-covariance and variance parameters for the latent process. Though most estimates appear reasonable. six of the thirty firms have an autocorrelation coefficient above one. However only one of those firms. Texaco. has a Q-statistic that can be rejected at the 95% confidence level. This means that only Alcoa. Caterpillar. and Phillip Morris have both Q-statistics that do not reject at the 95% level and implied stationarity in the latent variable. Hence for twenty seven of the thirty firms the MODM is either not supported by the data or gives implausible parameter estimates.

Such a poor fit indicates that the MODM does not do an adequate job at explaining the observed data. To check the robustness of the results I retest the MODM using data from 1988-1991 on the same thirty firms. The results for the Q-statistics from the second sample period can bee seen in Table 1.3. Notice that seventeen of the Q-statistics improve. slightly more than half. While fifteen firms instead of twenty-two reject the MODM in the new sample at the 95% significance level. there are still nine firms that reject the MODM at the 1% level. Checking Table 1.4 reveals that only three firms have estimates for the correlation coefficient greater than one in the 1988-91 sample period and all three have Q-statistics that fail to reject the MODM at the 95% level. That means that only twelve firms both fail to reject the MODM and have sensible estimates for the information flow's parameters.

Given that only seventeen of the thirty Q-statistics improve in the new sample period it is likely that the reduction in rejections of the MODM has occurred by chance. The Q-statistic, nevertheless, is highly non-linear and its distribution in finite samples unknown. To ascertain the likelihood that the rejection of the MODM is due to a non-Chi squared small sample distribution I perform Monte Carlo sampling with data calibrated to the grand mean parameter values. The simulations are carried out to test the ability of GMM to accurately estimate the parameters in finite samples and, most importantly, check the distribution of the Q-statistic. All reported statistics are based on simulating 10,000 data sets and then using GMM to estimate parameters and obtain a Q-statistic. The errors associated with Z_{1t} and Z_{2t} are standard normal in all simulations. This specification is consistent with theoretical models which typically have the errors as the sum of a random process, hypothesizing a close fit to a normal distribution. The specification for ϕ is more problematic. There is no theory to guide the choice of the distribution of the information flow, and given the AR(1) specification of the latent process, the conditional distribution of ϕ will change over time since the latent variable may not be negative. For simplicity I specify the distribution for ϕ to be a truncated normal.

As can be seen from Table 1.5 the GMM over-identifying Q-statistic is not closely approximated by the Chi-squared distribution in the simulations. Despite a relatively large sample size of 1,262 the asymptotic distribution for the statistic tends to over-reject the null hypothesis. For the most part the simulated distribution for the Q-statistic accommodates those found in the actual data tested. This may suggest that the MODM actually does an adequate job of fitting the data but that the Q-statistic behaves poorly in finite samples. Notice, however, that using the simulated Q-statistic distribution nine of the thirty firms reject the MODM at the 5% level of significance and seven of them reject at the 1% level for the 1983-86 sample period. Similarly seven of the thirty firms reject the MODM at the 5% level for the 1988-91 sample period. Though using the simulated distribution for the Q-statistic improves the outlook for the MODM, it is nevertheless difficult to explain the rejection of the MODM by roughly 30% of the firms at the 5% significance level.

1.5 Conclusions

Tauchen and Pitts (1983) present an intuitive explanation for the empirical relation between stock volume and returns by developing a Mixture of Distributions Model, similar to those of Epps and Epps (1976) and Clark (1973). Due to the presence of a latent variable, however, testing the model is not trivial. Though much circumstantial evidence exists supporting the MODM, few direct tests of the MODM have been made. Lamoureux and Lastrapes (1994) test the model using a signal extraction technique and present strong evidence against the MODM as does Watanabe (2000) using a Bayesian approach, while on the other hand Richardson and Smith (1994) and

Anderson (1996) find mixed results when using a Generalized Method of Moments approach. Re-evaluating the model analyzed by Richardson and Smith (1994), it is not completely obvious that they actually fit a version of Tauchen and Pitts' model. By adding in a term to represent non-negative average return, economic interpretation of the latent variable is difficult.

I fit the MODM by leaving out the non-negative average return term and instead I specify the latent variable as following an AR(1) process which incorporates the well known persistence in variance and volume of the data. From this specification I obtain over-identifying moment conditions and use Hansen's (1982) Q-statistic to test the model. Looking at the sample of firms used by Richardson and Smith (1994) from 1983 to 1986, I find strong evidence against the MODM. I check the robustness of the results by changing the sample period from 1983-86 to 1988-91. Though the fit does improve over the new sample period, it is most likely due to chance. Because of the difficulty in estimating the variance of volume, different over-identifying moment conditions are tried to test the robustness of the results with no noticeable improvement in estimation.

Though the model does not appear to fit the data, it does not mean the model can be rejected out of hand. The lack of fit may be due to an improper specification of the latent variable. Though computationally convenient, the AR(1) specification for the information flow may be inappropriate. Such a specification does induce time-series properties in the model's moments that are similar to those observed in the data, but there are numerous alternative specifications to the AR(1) that could accomplish the same task. One other possible reason for the poor fit could be due to parameter instability. Tauchen and Pitts (1983). for instance. posit that trading volume should grow linearly with the number of traders. Since the number of traders is unknown it is not possible to control for this variable. though trading volume over both periods does grow.

One other potential explanation is the distribution of the test statistic. Since only the test statistic's asymptotic distribution is known it is possible that the finite sample distribution is very different. I perform simulations to ascertain the behavior of the Q-statistic in finite samples. The simulations are calibrated to the sample data and carried out for sample sizes of 1.262. the same as the observed data. Surprisingly. GMM is found to perform poorly even when the moment conditions are correctly specified. This sheds some doubt on the outright rejection of the model. One possible reason. as noted by Richardson and Smith (1994). is that the distribution of the Q-statistic will depend on the distribution of the latent variable. For instance. if F_t is generated by a positive. stable. asymmetric distribution with characteristic exponent less than one then only the first two moments will exist for returns which violates one of the conditions Hansen used in his proof for the asymptotic distribution of the over-identifying Q-statistic.

Despite the test statistic's performance in small samples. it is unlikely that the small sample properties alone can account for the poor fit of the model to the data.

Since the model is being tested jointly with an AR(1) specification for the latent variable, it is possible that the data fail to fit the model due to mis-specification of the latent process.

Q-statistics for the MODM using time-series moments
versus those found in Richardson and Smith (1994)

Company	Q_2^2	R&S Q_3^2	Company	Q_2^2	R&S Q_3^2
Alcoa	5.8	11.9$^+$	J.P. Morgan	5.2	9.2*
Allied Signal	7.8*	12.1$^+$	Coca Cola	7.4*	19.8$^+$
American Express	7.1*	5.6	McDonalds	8.4*	5.6
Boeing	11.4$^+$	9.7*	3M	13.8$^+$	6.8
Bethlehem Steel	2.1	7.1	Philip Morris	5.3	4.2
Caterpillar	1.6	8.4*	Merck	8.4*	5.4
Chevron	18.2$^+$	16.5$^+$	Procter & Gamble	8.8*	0.9
Du Pont	19.1$^+$	11.3*	Sears Roebuck	15.7$^+$	20.7$^+$
Disney	1.7	2.5	AT&T	7.1*	11.9$^+$
Eastman Kodak	6.6*	7.2	Texaco	6.9*	8.2*
General Electric	6.7*	14.7$^+$	Union Carbide	1.0	3.2
General Motors	7.8*	8.9*	United Technologies	6.8*	5.1
Goodyear Tire	3.4	6.8	Westinghouse	8.4*	19.2$^+$
IBM	11.5$^+$	1.1	Exxon	16.2$^+$	4.6
International Paper	18.2$^+$	7.3	Woolworth	14.4$^+$	14.9$^+$

+ - Able to reject the MODM at the 1% level of significance
* - Able to reject the MODM at the 5% level of significance

TABLE 1.1. Q-Statistics

The estimated first order autocorrelation and
standard deviation of the latent process

Company	α	σ_F	Company	α	σ_F
Alcoa	0.37	0.40	J.P. Morgan	1.06	0.49
Allied Signal	0.76	0.63	Coca Cola	0.81	0.49
American Express	0.6	0.42	McDonalds	0.55	0.44
Boeing	0.54	0.37	3M	0.36	0.40
Bethlehem Steel	10.1	0.99	Philip Morris	0.66	0.63
Caterpillar	0.37	0.61	Merck	0.39	0.78
Chevron	0.38	0.41	Procter & Gamble	0.49	0.35
Du Pont	0.41	0.28	Sears Roebuck	0.44	0.36
Disney	13.5	1.13	AT&T	0.39	0.76
Eastman Kodak	0.56	0.45	Texaco	2.18	0.44
General Electric	0.46	0.35	Union Carbide	14.6	0.69
General Motors	0.37	0.30	United Technologies	0.76	0.36
Goodyear Tire	1.29	0.31	Westinghouse	0.30	0.54
IBM	0.42	0.29	Exxon	0.33	0.31
International Paper	0.51	0.47	Woolworth	0.91	0.40

TABLE 1.2. Estimated Autocorrelation of Latent Process 1982-86

Q-Statistics for the Two Sample Periods

Company	1983-86	1988-91	Company	1983-86	1988-91
Alcoa	5.8	4.4	J.P. Morgan	5.2	9.7$^+$
Allied Signal	7.8*	4.6	Coca Cola	7.4*	22.4$^+$
American Express	7.1*	6.8*	McDonalds	8.4*	16.5$^+$
Boeing	11.4$^+$	3.2	3M	13.8$^+$	5.2
Bethlehem Steel	2.1	9.5$^+$	Philip Morris	5.3	4.5
Caterpillar	1.6	2.7	Merck	8.4*	15.7$^+$
Chevron	18.2$^+$	7.5*	Procter & Gamble	8.8*	12.4$^+$
Du Pont	19.1$^+$	4.1	Sears Roebuck	15.7$^+$	18.1$^+$
Disney	1.7	5.1	AT&T	7.1*	3.5
Eastman Kodak	6.6*	3.0	Texaco	6.9*	5.0
General Electric	6.7*	7.4*	Union Carbide	1.0	7.6*
General Motors	7.8*	8.2*	United Technologies	6.8*	4.1
Goodyear Tire	3.4	11.4$^+$	Westinghouse	8.4*	5.8
IBM	11.5$^+$	3.9	Exxon	16.2$^+$	4.0
International Paper	18.2$^+$	16.9*	Woolworth	14.4$^+$	3.6

+ Able to reject MODM at the 1% level of significance
* Able to reject MODM at the 5% level of significance

TABLE 1.3. Subsample Q-Statistics

The estimated first order autocorrelation and
standard deviation of the latent process

Company	α	σ_F	Company	α	σ_F
Alcoa	0.43	0.40	J.P. Morgan	0.62	0.39
Allied Signal	0.09	2.37	Coca Cola	0.31	0.29
American Express	0.72	0.21	McDonalds	0.21	0.43
Boeing	0.42	0.39	3M	0.20	0.39
Bethlehem Steel	0.52	0.31	Philip Morris	0.33	0.40
Caterpillar	0.72	0.34	Merck	0.37	0.32
Chevron	0.66	0.45	Procter & Gamble	0.21	0.38
Du Pont	0.32	0.57	Sears Roebuck	0.37	0.44
Disney	0.95	0.17	AT&T	0.19	0.44
Eastman Kodak	0.28	0.65	Texaco	0.56	0.83
General Electric	0.18	0.35	Union Carbide	0.29	1.23
General Motors	0.48	1.0	United Technologies	0.33	0.51
Goodyear Tire	0.25	0.61	Westinghouse	0.83	0.29
IBM	36.9	0.47	Exxon	2.31	0.57
International Paper	0.28	0.34	Woolworth	2.18	0.56

TABLE 1.4. Estimated Autocorrelation of Latent Variable 1988-91

Comparison of the Chi-Squared Distribution
to the Small Sample Monte Carlo Distribution

Distribution of Q^2 from the Monte Carlo Simulations								
1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
12.8	11.1	9.6	8.1	3.7	1.1	0.7	0.5	0.3
Chi-Squared Distribution								
9.2	7.4	6.0	4.6	1.4	0.2	0.1	0.05	0.02

TABLE 1.5. Q-Statistic Distribution

Chapter 2

Forced Divestitures and Stranded Cost Recovery: An Event Study

2.1 Introduction

In October 1992. the United States Congress passed the Energy Policy Act (EPAct) and established the beginning of deregulation in the U.S. wholesale electricity market following the process that had already begun in England. Germany. Chile. and other South American and European countries. The intention of deregulation was to foster competition amongst electricity generators and drive down prices while doing away with the expense of and need for government regulators. The EPAct gave the Federal Energy Regulatory Commission (FERC) the power to open the wholesale power market to a more diverse group of firms. To accomplish this goal of increased competition. FERC subsequently issued Orders 888 and 889 which required utilities that own the transmission lines to provide equal access to all electricity generators. including the utilities' competitors. In many states this has been interpreted to mean that utilities that manage transmission lines must divest their power generating facilities to ensure that all firms have equal access to the market.

Despite Congress' involvement. deregulation of the market for electricity generation is handled by the individual states and has taken two basic forms. In the first type all consumers. including residential. have the ability to choose an electricity producer. That means that homes and businesses alike may specify who they want to purchase their electricity from. In the second form of deregulation residential consumers must continue to buy their electricity from the local utility while businesses are free to chose amongst providers. Typically this is chosen by states in order to test out the results of a deregulated market before forcing it on the voting public. Table 2.1 lists the states that have deregulated markets or that have plans to deregulate the market for electricity generation.

One of the first steps typically taken by states in preparing the market for deregulation is to force the divestiture of the utilities' power generating facilities to ensure that all power generators, not just the utilities, have equal access to the distribution network run by the utilities. In some states, such as California, electric utilities are forced to divest all of their generating capacity to other companies. In other states electric utilities only need to restructure their electric generation into a holding company separate from their electricity distribution unit. In almost all states that have chosen to deregulate the divestitures are forced when any single company holds more than some specified amount of the state's electricity generating capacity.

Any funds raised by the firm's divestiture of assets, in particular their power generating assets, are used to offset the as of yet unrecovered investment costs of the utility. The idea behind the divestitures is that "by law, the proceeds must flow to ... ratepayers as an offset to utility debts they are shouldering" (Smith 1999). Once the divestitures have been realized and sales prices of the assets are known, regulators and utilities then determine the outstanding amount of the utilities' unrecovered investment costs and a method for their recovery. In many cases these reported stranded costs are quite large. Table 2.2 lists the reported stranded costs to the FERC for a sampling of firms. Readily apparent from the table is that the size of the stranded costs are commonly in the billions of dollars. These forced divestitures and stranded cost recovery mechanisms are meant to leave the value of the parent company unaffected by guaranteeing the collection of the unrecovered investment while at the same time preparing the market for deregulation by encouraging the entry of new firms.

In the case that an electric generating plant is sold to another company, a public auction is held. If the plant is sold for less than its book value the remaining stranded costs are to be recovered from consumers in the form of higher prices, fees, or possibly subsidized loans to the utility. If the assets are sold for more than their book value the surplus is to be rebated to consumers through lower prices and fees or to be used to offset stranded costs from other power generation plants owned by the utility. The details of the stranded cost recovery mechanism varies from state to state. In most cases the state regulators allow the company to submit a plan for the stranded cost recovery and the plan is then approved or rejected by the state.

The method of stranded cost recovery varies not only from state to state, but also utility to utility. For instance in Arizona the Tucson Electric Power Company agreed with state regulators to recover their remaining stranded costs through a fixed $0.93 per kilowatt hour competitive transition charge in addition to a floating competitive transition charge that varies inversely with the price of energy and is reevaluated quarterly. Arizona Public Service. on the other hand. chose to recover stranded costs over a five year period through a competitive transition charge that decreases annually. Such differences in recovery mechanisms illustrate the bargaining that takes place between utilities and state regulators as the move to deregulation takes place.

2.2 The Uncertainty of Deregulating

Deregulation of the electricity industry has gained in popularity over the nineties because of the inefficient allocation of resources that occurs under regulation. As is the case in most regulated monopolies. rates charged for electricity are based on the cost of producing the product. Since rates of return are guaranteed to the utilities based on their costs. many electric companies have sought to lay unnecessary lines and build redundant and inefficient power generators. Problems under regulation also arise because there is little incentive for the utilities to improve their service and technology. to cut their costs of production. or to hire competent management teams.

Proponents of deregulation use two arguments to support the idea that a more efficient market will result if power generation is deregulated. First is the belief that a privatized market for electricity generation will do away with the incentive problem of building to much power generating capacity merely to drive up fixed costs. Along the same lines. more efficient generators will now be built since there is no more incentive to drive up variable costs and a new incentive to undertake research and development to lower costs. This more efficient investment, it is argued, will lead to lower prices in the long-run.

Similarly it is argued that in the short run prices will drop (see, for instance, Bessembinder and Lemon (1999)). The incentive to build unnecessary power plants has left the market with too much power generating capacity. This power glut will result in falling electricity rates as the market for electricity is deregulated and the market price for electricity comes closer to its marginal cost of production. Consider for instance the opening up of the Arizona electricity market to generators in other states. Now low cost hydro power generated in Idaho during the summer can be sold to Arizona where summer demand is extremely high, allowing prices in Arizona to fall. Not only will Arizona need less power generating capacity to meet its peak summer demand (and likewise Idaho will need less to meet its peak winter demand since it will then import electricity from Arizona), but the higher marginal cost plants will not be needed for production at all and may be removed from the power grid.

The uncertainty that deregulation induces makes valuing the companies' assets difficult. No one knows with certainty if prices will fall and, if so, by how much. Such uncertainty makes it difficult to forecast what the change in the value of the power generators will be once the market is deregulated. This is problematic for the utilities since much of their investment in power generators has not yet been recovered from their rate payers. If the market value of the assets turns out to be lower than the unrecovered investment costs of the assets themselves. the utilities worry they will never fully recover those costs. These stranded costs include the investments in power plants and power contracts that were made before deregulation. Evidence of the uncertainty in the recovery of the stranded costs is mentioned by the Harvard Electricity Policy Group (1998):

> "The treatment of utility assets and other financial arrangements that might be 'stranded' by the transition to a more competitive market has dominated the attention of utilities. The difficulty of reallocating stranded costs will also affect decisions about the eventual structure of the industry.... Therefore. defining a workable strategy for dealing with these costs is vital to implementing any new market structure." pg 19

Questions also surround the willingness of state governments to allow the firms to recover their stranded costs. For instance many groups argue that the rates of return allowed in the past on the utilities' assets were extremely high when compared to the actual risk being borne by the stockholders of the companies. It is argued that the stranded costs should be absorbed by the stockholders to make up for these past overpayments. Likewise the overinvestment in power generators has led many to argue that the utilities should not be allowed to recover the stranded costs on plants that were unnecessarily built at the ratepayers' expense. The utilities argue. however. that the generating capacity is built with the approval of regulators and. furthermore.

plants are often built because they provide jobs and reliable supplies of electricity to rural areas.

Such uncertainty in the regulatory process and in the market value of the power generators has led utilities to worry about the total recovery of these stranded costs. Failure to recover these costs could significantly affect the value of the firm. Though in theory the firm should recover any and all stranded costs. in reality firms may be exposed to risk associated with their recovery.

Questions about market power in a deregulated market. however. cause concern that deregulation will lead to market concentration and oligopoly pricing. In this scenario the purchasers of the generating capacity will benefit according to their market share and will pay higher prices for the power generators. Questions about market power have been raised in cases like California where Houston Industries has purchased nearly 25% of the state's electric generating capacity while another company. NRG Energy. holds an additional 17%. Likewise NRG holds one third of the generating capacity of New York state while Sithe Energies owns over 40% of Pennsylvania's capacity (see Wagman 1999). In fact half the capacity put on the auction block to date has been purchased by only five companies and between 1997 and 1999 over 75.000 megawatts of electric generating capacity have been sold for at least $24 billion (though not all sales prices have been publicly disclosed). Such a high concentration of market share has left many skeptical that a competitive electric generator industry will in fact exist once the market is deregulated.

If market concentration is indeed expected to result from deregulation then the market value of the utilities' power generators will be higher than the reported stranded costs. This will be evident if acquiring firms pay well over book value. If firms do pay well over book value but the electricity market is in fact expected to be competitive, then the high prices may be due to the Winner's Curse. Overpaying for the assets should result in a lower stock price for the winning bidder while the existence of market power should instead increase the acquiring firms' value.

It should be mentioned that even if market power does not lead to higher electricity prices the utilities may still be able to recover their stranded costs from the sale of their power generators. Because of the political environment of any regulated industry, electric generators have been built and maintained in regions where other valuable resources exist. In such a scenario the market prices of the asset will be well above book values but the winning bidders do not necessarily overpay or expect to enjoy market power. For instance Los Angeles has power generators on beach front property that have not been moved because of the lack of incentives on the utility's side and the difficulty of winning approval for the building of new power generating facilities in the state. In many of the divestitures of the utilities' power generators firms are bidding on much more than a generator. Take the case of Pacific Gas and Electric. When PG&E sold off their hydro generators, included in the deal was 136,000 acres of land and the yearly consumptive rights to 200,000 acre feet of water, an asset whose value will certainly increase as California's population continues to grow. Obviously many of the power generators being sold have other potential uses than just producing electricity. The general uncertainty about the value of the power generating plants. however. has left the electric utilities unclear about what price they can expect for their assets.

2.3 Event Study

As state's prepare to move to a deregulated market for electricity generation the recovery of stranded costs has been guaranteed by regulators. Any amount recovered by the utility in excess of its stranded costs should result in rebates to the utility's rate payers. Conversely. any amount not fully recovered by the sale of the generators should be ultimately recovered from the utility's ratepayers. That means that the value of the utility should be unaffected by the price received at auction for the asset. Since there is supposedly no threat that the utility will not recover all their stranded costs. nor is there any possibility that the utility will receive more than their stranded costs. the utility's stock price should not react to the announced sale of the asset. Looking at the reaction of the stock price to the announced sale can help us ascertain if investors perceive there being no risk to the utility.

If the utilities receive less than their stranded costs in the auctions and their stock price falls. the price move might be indicative of doubt that the remaining stranded costs will actually be recovered. Alternatively if the utility receives more than their stranded costs and the stock price rises. the price increase could be due to the market having previously discounted the possibility the utility would recover their full stranded costs or a suspicion that regulators will not rebate the excess to the ratepayers.

Similarly the firms that buy the divested assets should have no abnormal return associated with the sales announcement if the sales price accurately reflects the assets' value. If there is a negative abnormal return associated with the auction announcement for the winning firm then this may suggest that investors believe the winning firm overpaid for the asset. Similarly if the market perceives that electricity prices will go up due to market concentration. then winning firms may experience a positive abnormal return upon winning the auction. Whether investors believe a firm has underpaid or overpaid will largely be a function of beliefs about the deregulated price of electricity.

To evaluate the impact of the sales announcements on the stock prices I perform an event study. To do so it is necessary to first calculate the sales announcements' impact by choosing a measure of abnormal return. Following the notation of Campbell. Lo. and MacKinlay (1997). define the return for the i^{th} asset at time t as given by

$$R_{it} = \alpha_i + \beta_i R_{Mt} + \epsilon_{it}$$

where α_i is the unconditional expected return. R_M is the return on a market index. and ϵ_i is a normally distributed error term with

$$E[\epsilon_{it}] = 0$$
$$Var[\epsilon_{it}] = \sigma_{\epsilon_i^2}$$

The normal return for the stock would then be $E[R_{it}|R_{Mt}]$, the stock's expected return conditional on the observed market return. This gives the abnormal market return for asset i at time t, ϵ^*_{it}, as the difference between the observed return and the conditional expectation

$$\epsilon_{it} = R_{it} - E[R_{it}|R_{Mt}]$$

Since α and β are not known, we must work with their OLS estimates. To evaluate the effect if the announced sale of the electricity generators we wish to compare the T_1 realized abnormal returns from the event (in this case the three days surrounding the sale announcement) with the distribution of abnormal returns for the stock. Since the distribution of abnormal returns is in general not known, it will need to be estimated using T_0 observations on the stock's returns. Putting things in matrix notation we have

$$\widehat{\epsilon}_i = R_i - \widehat{\alpha}_i\iota - \widehat{\beta}_i\widehat{R}_M = R_i - X_i\widehat{\theta}_i$$

where ι is a (T^*x1) vector of ones, $T^* = T_0 + T_1$, X_i is a (T^*x2) matrix comprised of a column of ones and a column containing the index returns, and $\widehat{\theta}_i$ is $(2x1)$ vector containing $\widehat{\alpha}$ and $\widehat{\beta}$, the parameter estimates. The conditional distribution for the abnormal returns is normal and centered at zero. Denoting the submatrix containing only the T_1 event window observations with a superscript star, the covariance matrix for the abnormal return is given by

$$V_i = E[\widehat{\epsilon_i^*}\widehat{\epsilon_i^{*\prime}}] = \sigma_{\epsilon_i}^2 I^* + \sigma_{\epsilon_i}^2 X_i^*(X_i'X_i)^{-1}X_i^{*\prime} \tag{2.1}$$

where I^* is a $(T_1 x T_1)$ identity matrix. The first term in equation 2.1 is of course the variance caused by the error term. while the trailing term represents the estimation risk. As can be seen from the second term. the sampling error on the parameter estimates will in general induce serial correlation in the estimated abnormal returns even when the true error terms are uncorrelated.

Under the null hypothesis that the announced sale has no effect on the value of the firm the distribution of the event window's abnormal return is given as

$$\widehat{\epsilon_i^*} \sim N(0. V_i^*)$$

To aggregate the abnormal returns over the T_1 day event window we define the cumulative average return as

$$\widehat{CAR}_i = \sum_{i=1.3} \widehat{\epsilon_i^*}$$

The variance V of this sum is given as

$$Var[\widehat{CAR}_i] = V = \frac{1}{T_1^2} \sum_{i=1.T_1} V_i^*$$

To aggregate over the N observed sales announcements we can write the cumulative average abnormal return as the sum of the cumulative abnormal returns for the N events

$$\overline{CAR} = \sum_{i=1.N} \widehat{CAR}_i$$

The cumulative average return for the events will have a variance given by

$$Var[\overline{CAR}] = \overline{\sigma}^2 = \gamma' V \gamma$$

where γ is a $(T_1 x 1)$ vector of ones. To draw inference about the event date returns we form the test statistic J

$$J - Statistic = \frac{\overline{CAR}}{\overline{\sigma}}$$

which is asymptotically distributed as a standard normal.

2.4 Results

Forced divestitures are found using the Wall Street Journal and data provided by the Federal Energy Regulatory Commission. The sales dates and amounts are independently confirmed using data from Bloomberg International. Announcement dates are gathered for twenty four auctions. a list of which is found in Table 2.3 and Table 2.4. Table 2.3 lists the sales dates and the utilities that divested the power generators. Table 2.4 lists the firms that placed the winning bids and the amount they paid for the power generators. Notice that in some cases there was more than one divestiture on the same date. This usually occurred because the same firm divested more than one power generator on the same date but sold the generators to different bidders. When there is more than one event on any given date an equally weighted portfolio of the participating firms is formed and the abnormal return is calculated for that portfolio.

The market portfolio used to calculate the conditional return is hand gathered data on the S&P 500. Stock returns and volume are taken from the Center for Research in Security Prices (CRSP) tapes. The returns are dividend adjusted and the volume is corrected for any stock splits or stock dividends.

Note that a good portion of the divestitures take place in Maine and California. One of the reasons for this is that they were two of the first states to begin deregulation and, hence, two of the first states to force divestitures. Table 2.3 also reveals the large variance in the value of the assets being divested. The prices received on the assets at auction range from a mere $2.5 million to $1.8 billion. Notice that even though Edison International received around $1.2 billion from their divestitures the firm had almost $12 billion in stranded costs as is reported in Table 2.2. Overall the majority of the sales are for amounts more or less on the same order of magnitude as the outstanding stranded costs.

Table 2.5 reports the results for the buyers of the divested power generators. The J-statistic for the buyers' abnormal return is negative, indicating that the stock prices for the winning companies on average went down upon the sales announcement. The negative J-statistic is consistent with the accusation from the financial press that the winning firms had overpaid for the assets but is at odds with the claim that market concentration will lead to higher electricity prices and profit margins. The p-value for the J-statistic, however, is only 76% which suggests the movement in the stock price is likely due to pure chance. One possible test of the results being pure chance is to look at the consistency in the sign of the individual stocks' abnormal return. Though the J-statistic is not significant, it would be odd to find that all of the firms had negative, albeit insignificant, abnormal returns over the event window. The reported sign statistic in Table 2.5 of 0.236 suggests however that the winning firms were nearly as likely to have positive abnormal returns on the sales announcement dates as they were to have negative. Such a finding adds credence to the theory that the negative abnormal return for the auction winners is purely coincidental.

Since there is a large degree of variance in the sales prices I also regress the J-statistic for each individual company's cumulative abnormal return on the announced sales price. If the overall sample's J-statistic is insignificant because of the inclusion of the auctions for the small generators then the slope parameter on the individual J-statistics in the regression should itself be significant. As can be seen in Table 2.5, this does not appear to be the case. The coefficient is estimated to be negative which does suggest that a higher sales price implies a larger negative return. Though this interpretation is consistent with the Winner's Curse theory that the winning firms overbid and at odds with the market power theory, the coefficient is not statistically significant.

One possible explanation for the widely reported belief that the auctions' winning firms had overpaid is found at the bottom of Table 2.5. The average daily share volume from the estimation period is calculated to be 48,012 shares. The average daily share volume over the event window, however, is 61,239 shares. The higher share volume may indicate an increase in the intensity and heterogeneity of investors' beliefs about the reported sale. Several theoretical models including Epps and Epps (1976) and Kandel and Pearson (1995) suggest that trading volume increases when investors receive public information but are split on how the information affects the company. It is possible that the average investor's belief about the winning firm's prospect remains unchanged but that there is now more dispersion in those beliefs. Given the uncertainty over the purchased assets and the future prospects for the power generating industry. such an explanation is at least plausible.

Table 2.6 contains cumulative abnormal return information for the sellers of the power generators. Here the story is quite different. The J-statistic for the event window is 2.62 which is significant at even the 99% level. Such a result is certainly consistent with the reported fears of industry executives that the stranded costs of the power generators would not be fully recovered. If investors believe the promise from regulators that all stranded costs will be recovered and any excess rebated to customers then there should be no change in the stock price at the sales announcement.

The reported abnormal return is also consistent with the possible belief that the utilities would recover in excess of their stranded costs but that they would not be forced to rebate the difference to the rate payers. Such an explanation might be consistent with the existence of a captive regulator. This could in particular be a problem where deregulation is occurring since regulators quite possibly will hope to land jobs with the deregulated companies as the government bureaucracies are disbanded. This scenario, however, would also imply that regulators would allow excessive stranded cost recovery to begin with, making it less likely that firms would ultimately recover more than their reported stranded costs since those numbers would have already been inflated.

Though both theories are possible explanations. the reported reaction of people within and without the industry[1] was generally of surprise at the high sales prices received from the auctions. Recalling the information on stranded costs in Table 2.2. however. suggests that the theory that the rebates to rate payers would not take place is unlikely since the reported stranded costs are larger than the amount received from the divestitures. This is in part due to the fact that the utilities' stranded costs include not only the generating assets being sold but other areas of operation in addition to power generation. In fact many companies like PG&E and Edison were left several billion dollars short of their full stranded cost recovery even after their power generators were sold. A more plausible explanation is that the uncertainty associated with the divestitures is less about if the stranded costs would be rebated to customers than if and when the utilities' stranded costs would be recovered.

To see if the J-statistic is dominated by a few outliers I calculate a sign statistic for the individual utility cumulative returns. Looking at the sign statistic for the sellers

[1] Nancy Loftin from Arizona Public Service noted in a personal conversation that her company along with others in the industry were "very suprised" at how high the winning bids were for the power generators.

of the power generators reveals that a number of the firms actually have a negative stock return over the event window, though a majority do have positive returns. Since industry sentiment is that the utilities received much more for the assets than they had originally anticipated, one might expect a clear majority of firms to have positive returns associated with the sales announcement of their power generators. Given that deregulation is undertaken on a state by state basis, however, the sign statistic could reflect the fact that only some states are believed to potentially disallow the full recovery of the utilities' stranded costs. Alternatively the sign statistic may reveal that the uncertainty associated with the asset sales decreased with time as the firms announced sales prices and fears of high unrecovered stranded costs dissipated.

The number of negative returns over the event windows could also be caused by the insignificance of the sales of the utilities' small power generators. To check this I once again regress the J-statistics form the individual utility cumulative abnormal returns on the sales price of the power generators. If the size of the sale is important, the estimated coefficient should be positive and significant. The estimated coefficient from the regression on the sales price is positive and significant in a one tailed test at the 3.5% level. This suggests that there is a relationship between the sales price received for the asset and the firm's cumulative abnormal return.

Notice, however, that unlike the case for the auctions' winning bidders, the average daily volume falls during the event window for the assets' sellers. This result, however, is not significant. It does add credence, though, to the suspicion that the winners experience abnormally high volume around the sales announcement dates. The fact that the sellers' stocks drop in volume around the announcements suggests that the increase in volume in the stocks of the winning bidders is not due to generally high market volume around the event dates.

It should also be noted that there is a difference in the average size between the firms that buy the divested assets and those that sell them. Table 2.7 lists the market caps for the publicly traded firms in the sample. Note that buyers are 2.5 times larger than the sellers on average. This in part reflects the more diverse operations of the buyers and could be a reason why there is no statistically significant abnormal return to the buyers. Given the larger size of the buyers, it is possible that there is no significant stock price movement even though they overpay for the power generators because the scope of their operations is much larger than that of the sellers. Perhaps the value of the assets being auctioned off are too small to noticeably impact the value of the purchasing firm. Further evidence of this is seen by noting that the median firm that purchases an electricity generator spends roughly 2% of their total equity value on the purchase while the median firm selling an electricity generator receives roughly 10% of their equity value in exchange for the asset.

2.5 Conclusions

The push for deregulation by Federal regulators has caused several states to restructure their markets for electricity. A good portion of states have begun to deregulate their market for electricity generation. In preparation for the deregulated market those same states have required several utilities to divest themselves of all or a portion of their power generators. Since the power generators had been built under regulation the utilities were promised that they would recover their costs and an additional positive return on their investment. Not all of the costs associated with the utilities operations and power generators had been recovered from the rate payers. however. when the utilities were forced to sell off their power generators.

Since the utilities had been regulated when the generators were built. the firms were supposed to recover their costs and the return on their investments even if the market was deregulated. If the utilities received more for their assets than what the rate payers owed. then the utility was to rebate the difference through lower prices. Conversely if they did not receive enough to cover their costs then the rate payers were to make up the difference through additional fees and/or higher rates. That means that the utilities' stock prices should not be effected by the announced sales price of the assets.

There was some concern. however. that the costs of the assets would not be fully recovered. The divestiture of the assets might preclude firms from recovering their investment, essentially stranding the unrecovered costs. Political dealings and the uncertainty about the assets' value in an unregulated market led to uncertainty about the ultimate recovery of the costs.

On the other side of the asset sales, many claimed that the prices ultimately paid for the power generators were too high and a result of the Winner's Curse while others pointed to market power as the explanation for those high prices. If the companies did overpay for the assets, one might expect to see a drop in their stock value, while if purchasing the assets gives them market power the stock price might be expected to rise.

Looking at the cumulative abnormal returns for the utilities selling the power generators reveals a strong, positive market reaction to the announced sale. Such a reaction is consistent with the claim that industry executives were worried that their stranded costs would not be recovered. The abnormal return for the utility is positively and significantly related to the sales price of the asset.

The results are more mixed for the auction winners. Though the abnormal return for the auction winners is not significant, it is negative and consistent with the Winner's Curse. In addition the volume over the event dates for the auction winners is abnormally high, which may suggest that there are strong differences of opinion on how to interpret the news of the asset sales.

State of Electricity Deregulation by State

States that allow all customers to choose their electricity generator

Arizona	New Hampshire
Connecticut	New Jersey
Delaware	New York
District of Columbia	Ohio
Maine	Pennsylvania
Maryland	Rhode Island
Massachusetts	

States that allow only non-residential consumers to choose their electricity generator

Illinois	Oregon
Montana	

States still in the process of deregulating

Arkansas	New Mexico
California	Oklahoma
Michigan	Texas
Nevada	Virginia

TABLE 2.1. Deregulation by State

Stranded costs* reported by utilities to FERC

Company	Stranded Costs (millions $)
Boston Edison Company	2.083
Dominion Resources Inc.	3.147
Commonwealth Edison Company	4.413
Edison International	11.839
FPL Group Inc.	1.350
GPU Inc.	7.599
Houston Industries Inc.	2.319
Northeast Utilities	6.189
Southern Company	942
Consolidated Edison of NY	3.900
Niagara Mohawk Power	5.029
Northern States Power	765
PG&E Corporation	7.679

* Stranded costs are the outstanding amount of money that the utility is owed by the rate payers.

TABLE 2.2. Stranded Costs

Sales of Power Generators by Utilities

Seller	Sales Date
GPU	07/17/98
GPU	08/03/98
Bangor	09/28/98
Commonwealth Edison	12/30/97
New England Power	08/06/97
Boston Edison	12/10/97
Maine Public Service	07/07/98
Central Maine Power	01/06/98
Eastern Utilities Associates	06/26/98
Eastern Utilities Associates	08/04/98
Eastern Utilities Associates	10/15/98
Eastern Utilities Associates	05/27/98
Commonwealth Energy Systems	05/27/98
United Illumination	10/05/98
Edison International	11/24/97
Edison International	02/04/98
Edison International	03/26/98
Northern States Power	09/16/98
Pacific Gas & Electric	11/18/97

TABLE 2.3. Generator Sales

Purchases of Power Generators

Buyer	Sales Price Millions $
Amergen	100
Edison Mission	1.800
PP&L Global	89
Southern and Dominion	254
US Generating	1.590
Sithe	657
WPS	37.4
FPL	846
Great Bay	3.2
FPL	2.4
NRG	55
Southern	537
Southern	462
Wisvest	272
Houston Power, NRG, and AES	1.105
NRG	29.8
Houston Power	43
Cogen Corporation	2.5
Duke	505

TABLE 2.4. Generator Purchases

Statistics for the buyers of power generators

J-Statistic	-0.294
(P-value)	(76%)
Sign Statistic	0.236
(P-value)	(41%)
Size Regression T-Statistic	-0.06
(P-value)	(49%)

	Estimation Period	Event Window
Average Daily Volume	48012	61239
(P-value)		(11%)

The J-Statistic tests the abnormal return of the stocks around the announced sales dates.

The sign statistic tests if the sign (positive or negative) of the cumulative returns is random.

The size regression looks at the relation between the announced sales price and the abnormal return.

TABLE 2.5. J-Statistics for Buyers

Statistics for the sellers of power generators

J-Statistic	2.62
(P-value)	(0.8%)
Sign Statistic	0.385
(P-value)	(35%)
Size Regression T-Statistic	1.6
(P-value)	(3.5%)

	Estimation Period	Event Window
Average Daily Volume	33717	26606
(P-value)		(46%)

The J-Statistic tests the abnormal return of the stocks around the announced sales dates.
The sign statistic tests if the sign (positive or negative) of the cumulative returns is random.
The size regression looks at the relation between the announced sales price and the abnormal return.

TABLE 2.6. J-Statistics for Sellers

Market Cap for Buyers and Sellers as of January 1. 1998

Buyer	Market Cap Thousands $	Sellers	Market Cap Thousands $
Amergen	5.396.643	Bangor	45.558
Cogen Corp.	135.885	Boston Edison	248.311
Duke	19.926.804	Central Maine Power	494.755
Edison Int'l	10.370.590	CES	715.939
FPL	10.766.502	Commonwealth Edison	6.657.159
Houston Power	7.893.149	Edison Int'l	10.370.590
NRG	4.337.295	EUA	536.445
PP&L	3.970.608	GPU	5.088.363
Southern	17.856.699	Maine Public Service	19.606
Wisvest	3.198.983	New England Power	296.735
		NSP	4.337.295
		PG&E	12.756.803
		United Illumination	647.764
Average	8.385.315	Average	3.247.322

TABLE 2.7. Market Cap of Buyers and Sellers

Chapter 3

Predictability in Asset Returns: Model Size and Inference

3.1 Introduction

Investors and financial economists have expended enormous resources on the study of the predictability of asset returns. In fact most early work in finance, such as Bachelier (1900), was primarily concerned with finding patterns in asset prices. To investors' dismay most early evidence on return predictability, as summarized by Fama (1970), showed little reason to believe that there were any predictable patterns in asset prices that could be consistently exploited by investors to earn abnormal returns. As Roll (1988) points out, "financial economists have even developed a coherent theory ... to explain why (stock returns) *should* be unpredictable."

Though predictability in returns is not a necessary condition for market efficiency (see Lucas(1978)), it does have certain implications about the time-series properties of the investment opportunity set. Asset pricing models, however, are silent as to which variables might be important in explaining the predictability in asset returns in the case of a stochastic investment opportunity set. Such a void in theory complicates the search for predictability since the functional form of predictability and the relevant variables are unknown, if they exist at all. This has left the empiricist with only specification searches for predictability in the quest to disprove the efficient market hypothesis or at least the notion of a static investment opportunity set. In addition, finding predictability through a specification search will complicate the interpretation of statistical significance because of the well known econometric problems associated with running a series of regressions containing everything including the kitchen sink and then only reporting those variables found to be significant as measured by their t-statistics.

In recent years, however, the belief that asset prices are not predictable has suffered some serious setbacks. Fama and French (1988), for instance, find that "predictable price variation due to mean reversion ... (is) about 40% for 3-5 year return variances." Poterba and Summers (1988) find positive serial correlation over short time horizons and strong negative serial correlation over longer horizons. Poterba and Summers (1988) and DeBondt and Thaler (1985) suggest that such patterns could be caused by investors overreacting to news, causing prices to slowly mean revert. McQueen and Thorley (1991) point out that such overreaction stories imply nonlinearities in returns and they use a Markov Chain technique that allows for such nonlinearities, finding "evidence of nonrandom behavior in *postwar* annual returns." Others, such as Richardson (1993), caution against accepting predictability given the performance of univariate statistical tests in Monte Carlo sampling, showing that such tests can not reject the hypothesis of no serial correlation. Lamoureux and Zhou (1996) take a subjectivist approach so as to avoid questions about the sampling properties of the statistics and instead focus on the information content of the observed data, finding that the data strongly support the hypothesis of no predictability.

Somewhat more difficult to explain. however. have been the findings of multivariate tests for predictability. The advantage of multivariate specifications is in the increased power. As Fama (1992) points out. univariate tests "lack power because past realized returns are noisy measures of expected returns." Using a multivariate specification allows for less volatile forecasting proxies for expected return to be used. Keim and Stambaugh (1986) and Campbell (1987). for instance. find that ex post risk premiums can be predicted using ex ante data such as interest rates and credit spreads. Similarly. Campbell (1991) and Campbell and Shiller (1988) find that earnings and dividends are useful in predicting future returns. Such findings. however. are somewhat less than breathtaking. Despite the regression's statistical significance. regression R^2 are typically low. especially for shorter time horizons such as monthly returns. Though Foster. Smith. and Whaley (1997) suggest a "procedure that adjusts the critical R^2 values to account for selecting variables by searching among potential regressors". economic inference remains difficult and the specter of pre-test estimators forces any conclusions drawn from the analysis to be taken with a grain of salt.

In light of the inconclusiveness of traditional techniques (particularly over short time horizons) and to avoid potential pre-test estimator problems. Kandel and Stambaugh (1996) (hereafter referred to simply as KS) choose to gauge predictability in terms of the economic information conveyed by the data. Circumventing the questions of sampling properties of estimators. similar to Lamoureux and Zhou (1996). the focus of KS is on quantifying the information contained in the observed data by looking at a portfolio allocation problem. To measure the economic information in the data. KS focus on how a rational, risk-averse investor would use the data in allocating wealth between a risk-free and a risky asset. Surprisingly they find that, even for cases where there is little statistical evidence of predictability, the investor makes appreciable changes in her portfolio allocation between the risky and risk free asset based upon the predictability found in the data The possible conclusion from the analysis is that, even given pre-test estimators and low R^2, investors themselves would find the degree of predictability at short time horizons (and hence the higher predictability found at longer time horizons) enough to significantly alter their portfolio allocations, providing strong evidence against, if not market efficiency, at least the idea of a static investment opportunity set.

Barberis (2000) extends the KS framework to a multiperiod decision problem and finds that long horizon investors put more of their wealth in stocks, though parameter uncertainty reduces this horizon effect since it creates a hedging demand. Xia (2001) extends the multiperiod framework to continuous time and finds that the horizon effect may induce investors to allocate less wealth to the risky asset. These papers, however, do not address the issue of how the model of predictability itself is chosen by the investor. The investor in KS's paper uses a linear regression model to update prior beliefs about the one month ahead distribution of stock returns, but there is no discussion on how the model is chosen or if, over time, the model can be refined. My paper seeks to ascertain how the search for the model itself changes inference. That is, when the investor observes the data through competing regression models, how does the investor update her priors and, hence, portfolio allocation? By increasing the model size the data will be forced to not only specify parameter values but also the model to be used in forming beliefs about the distribution of returns. This larger space reduces the perceived amount of economic information contained in the data. This brings about a joint-hypothesis problem when testing for predictability in returns with a Bayesian investor: the amount of predictable, economic information contained in the data will be determined by the model (likelihood function) and the loss function (utility function) through which the data are viewed.

3.2 Return Predictability with a Known Model

Following the structure of KS, consider a risk-averse investor at time T with a one month investment horizon. The investor observes data on returns and certain economic variables that are thought to predict the return for month $T+1$. Instead of rejecting or failing to reject a hypothesis about the predictability of returns, the investor must decide how to use the data to allocate wealth between a risky and a risk-free asset. The investor uses the likelihood function to process the data so as to update her prior belief about the return process and, in particular, the conditional mean and conditional variance of the time T+1 return. In order to form beliefs about future returns and to determine the optimal allocation between the two assets, the investor must first have some model that gives rise to a window (likelihood function) through which to view and process the data. When the investor knows the model of the return process. updating the prior beliefs is a simple application of Bayes' Rule since the likelihood function is known. For instance. assume that monthly returns for the risky asset are generated from the following linear process. assumed to be known to the investor:

$$r_{T+1} = x_T'\beta + \epsilon_{T+1} \tag{3.1}$$

where r_{T+1} is the continuously compounded excess return in month $T+1$. X_T is a vector beginning with one and also containing K predictive variables that are observed at the end of month T. β is a vector with an intercept and K slope coefficients. and ϵ_{T+1} is a mean zero. normally distributed random variable. Using Bayes' Rule to combine the likelihood with the prior on the model's parameters (in this case β and σ) gives the posterior distribution on the parameters:

$$p(\overline{\beta}.\overline{\sigma}|\Phi_T) \propto p(\underline{\beta}|\sigma)\cdot p(\underline{\sigma})\cdot p(\Phi_T|\beta.\sigma)$$

where $p(\Phi_T|\beta.\sigma)$ is the likelihood defined by equation 3.1. $p(\underline{\beta}|\sigma)$ and $p(\underline{\sigma})$ are the prior densities. and Φ_T is the data set the investor observes at time T. To get the predictive density for the monthly return. the posterior density on the parameters is combined with the predictive likelihood to get the joint density for the parameters and future returns

$$p(r_{T+1}.\overline{\beta}.\overline{\sigma}|\Phi_T) = p(\overline{\beta}.\overline{\sigma}|\Phi_T)\cdot p(r_{T+1}|\beta.\sigma)$$

Integrating over the unknown parameters gives the predictive density for r_{T+1}, the conditional return for the upcoming month.

$$p(r_{T+1}|\Phi_T) = \int\int p(r_{T+1}, \beta, \sigma|\Phi_T) d\beta d\sigma \qquad (3.2)$$

Combining the utility function, v, with the predictive density gives an objective function of the form

$$\max_{0 \leq \omega < 1} \int v(W_{T+1}) p(r_{T+1}|\Phi_T) dr_{T+1} \qquad (3.3)$$

where the wealth at the end of the month, W_{T+1}, is given by

$$\begin{aligned} W_{T+1} &= W_T[\omega e^{r_{T+1}+i_{T+1}} + (1-\omega)e^{i_{T+1}}] \\ 0 &\leq \omega < 1 \end{aligned}$$

where i_{T+1} is the return on the risk-free asset and ω is the percentage of wealth the investor allocates to the risky asset. For present purposes, assume that the investor's preferences are represented by the power utility function[1], defined as

$$v(W) = \begin{Bmatrix} \frac{1}{1-A} W^{1-A} \; for A > 0, A \neq 1 \\ \ln W \; for A = 1 \end{Bmatrix}$$

The investor must choose the ω that maximizes her expected utility function presented in equation 3.3. Notice that this is different than the portfolio selection problem carried out by Bayesian agents in Klein and Bawa (1976) and others. In these earlier papers the data is used only to estimate the unconditional mean and variance. Their focus is on the role of estimation risk; the uncertainty of what the true unconditional mean and unconditional variance are. Here the investor maximizes her

[1]The power utility function is characterized by constant elasticity so it is not dependent on the choice of W_0, the investor's initial wealth.

expected utility using a vector of predictive variables to form a one step ahead predictive conditional mean and variance. That is, in addition to estimation risk the returns are assumed to be predictable and, hence, the investor also estimates and uses the conditional mean and conditional variance.

The question remains, however, as to how much predictability in the data is needed for there to be a significant difference between the portfolio allocation using the unconditional moments versus the predictive, conditional moments. Gauging predictability in terms of classical statistical tests may result in very different inference than gauging predictability in terms of an investor's portfolio allocation. For instance, KS report that an investor with a non-informative prior and risk-aversion parameter of $A = 2$, given 67 years of data with an average monthly return of 0.49% and a standard deviation of 5.6%, would invest 94% of her wealth in the risky asset. The same investor, after observing the results of a regression of 67 years of return data on 25 regressors resulting in an R^2 of 2.5%, would completely divest herself of the risky-asset given that the estimate for r_{T+1}, conditional on X_T, was 1 fitted regression standard deviation below the sample mean. That is, if

$$\widehat{r}_{T+1} = \bar{r} - \sigma_{\bar{r}} \cdot \sqrt{R^2}$$

where $\widehat{r}_{T+1}$ represents the expected return conditional on X_T, $\bar{r}$ is the average return over the sample, and $\sigma_{\bar{r}}$ is the return's sample standard deviation. This seems to suggest that, despite a p-value of about 75%, which is well within traditional acceptance levels, an investor would nonetheless strongly reject the hypothesis that stock returns are not predictable. Even if the investor had observed 162 years of data with an R^2 of 0 before she observed the 67 years of data with an R^2 of 2.5%, her allocation to the risky asset would be only 60% of her wealth instead of the 94% she would have invested had she not observed the results of the linear regression.

3.3 Return Predictability and Model Choice

One obvious question is why an investor, despite observing 162 years of data with an $R^2 = 0$, wouldn't have discarded or at least updated the model? In KS the investor is implicitly assumed to know the correct data generating process or, at the least, to behave as though the model were the correct process. In frequentist methods there is usually some comparison of the null hypothesis to an alternative nested hypothesis. In this case the alternative hypothesis is that the mean return conditional on the vector X_T is equal to the unconditional mean, or that the slope parameters all equal zero. The Bayesian investor has not been allowed to update her prior beliefs about the model itself since there is no consideration given to the alternative hypothesis. That is, as the problem is currently formed, the agent only uses the data to define the parameter space but not the model space. The result is that the Bayesian agent acts in a very non-Bayesian way: she never update her priors on the possible models. The investor has violated what Poirier (1988) calls Cromwell's Rule: the idea that one should never discount the possibility that their model of the world is wrong.

Take for instance the models of Veronesi (1999), Donaldson and Kamstra (1996), Timmerman (2001), and Cagetti, Hansen, Sargent, and Williams (2202). In these models the economic agents face uncertainty about the current regime and the agents must infer the regime by observing prices over time. In such a setting it is possible to create patterns in returns despite having a static investment opportunity set and an absence of arbitrage. The learning process of economic agents about the regime is the source of the persistence and not a market inefficiency. Timmerman (2001) shows that these structural breaks can cause serious problems for classical inference when ignored. Given the sensitivity of classical tests to ignoring this learning about the regime, it is also possible that the Bayesian agent's optimal portfolio is sensitive to the assumption that the model of asset returns is known.

Introducing uncertainty over the model space makes the investor's maximization problem more complex. Assuming that the investor has several candidate models of the return generating process, then the maximization problem in equation 3.3 takes on the more complicated form of

$$\begin{aligned} & \max_{0\leq\omega<1} \int v(W_{T+1})p(r_{T+1}|\Phi_T)dr_{T+1} \\ = & \max_{0\leq\omega<1} \int \{v(W_{T+1}) \int [p(r_{T+1}|\Phi_T, M_i)p(M_i)dM_i]dr_{T+1}\} \end{aligned} \tag{3.4}$$

where M_i is the i^{th} model, $p(r_{T+1}|\Phi_T, M_i)$ is the predictive density of returns given that we use the likelihood of model M_i in forming the distribution, and $p(M_i)$ is the subjective probability that model i is the actual return generating process. Notice that now the investor not only must integrate over the uncertainty in the parameters of each model, but she must also integrate over the uncertainty of which model to use in forecasting returns. The investor must choose how to optimally combine the information in the models to construct a predictive density of returns. Such a problem is obviously more complex than the original and will force the data to provide us with information about a larger space of uncertainty. Forcing the data to answer more questions simultaneously will likely reduce the perceived precision of the answers.

Though examining the allocation decision with 25 regressors and a known model in KS is meant to shed some light on the importance of pre-test estimators in accepting or rejecting the null hypothesis of no predictability, it is not obvious that the same results of significant portfolio re-allocations are obtained when the agent is allowed to simultaneously determine which model to use in forming the predictive density for returns. The model uncertainty allows the investor to evaluate a number of competing models and incorporate information from each.

3.3.1 Point Mass on the Model of No-Predictability

Given the silence of asset pricing models on the form of predictability in asset prices, it is important that we analyze and understand how uncertainty over something as simple as the inclusion or exclusion of regressors influences model uncertainty and, hence, the portfolio allocation problem. Pastor (2000) and Pastor and Stambaugh

(2000) investigate model uncertainty with regards to asset mispricing. Asset pricing models are nested in a meta model, but their framework allows for model choice only by allowing posterior distributions on parameters to bunch up around zero, thereby effectively turning the parameter off. This is also possible in the KS framework, however it is unclear that the two approaches are identical or even close approximations. While Carlson, Chapman, Kaniel, and Yan (2001) look at the implications of using a partial equilibrium model in a general equilibrium framework, this paper concentrates on model uncertainty in a partial equilibrium.

This paper's approach, and perhaps the closest to frequentist intuition, is to place point mass in the prior that all slope parameters are simultaneously equal to zero. Such a procedure would be akin to using a frequentist F-statistic that all the slopes in the regression are equal to zero as a gauge of the performance of the model. The amount of point mass at zero reflects the uncertainty about the predictive power of the model defined by the candidate regressors in X_T. In the case where the investor is assumed to know the model of returns (no point mass is placed at zero for the slope parameters), there is no accumulation of probability mass possible on the slope parameters being zero since there is no competing model of unpredictability. Including point mass in the prior could allow the agent to formally reject the model (hypothesis) that returns are predictable by allowing probability mass for the parameter to accumulate at zero. This approach is similar to Avramov (2002) and Cremers (2000) who also form posterior probabilities on models by choosing amongst potential regressors to analyze the predictability in stock returns and form optimal combinations of regressors. A similar paper from a frequentist perspective is Lo and MacKinlay (1997), who also make a search over regressors to form portfolios with maximum predictability. Brennan and Xia (2001), on the other hand, documents the sensitivity of the optimal portfolio to model uncertainty in continuous time when the covariance matrix is known and so is the specification of the competing models.

This paper differs from Avramov (2002) and Cremers (2000) in that the question of pre-test estimators is addressed directly. Avramov (2002) and Cremers (2000) focus on choosing amongst 14 regressors found in the existing literature to predict asset returns according to certain statistical measures. For instance, Avramov (2002) uses 14 regressors previously found to be significant in the literature and combines them to form an optimal investment portfolio for a Bayesian agent. The focus is on choosing which of the 14 variables are most important for forecasting returns from an investor's point of view. Cremers (2000) also looks at 14 variables found in the existing literature, but instead of forming optimal investment portfolios for a Bayesian investor the focus is on the posterior distributions and sample evidence in support of the different variables. Cremers (2000) also compares the performance of five model selection criteria and uses out of sample forecasting to check their relative performance, finding that the Bayesian method performs the best. These approaches, however, are potentially misleading since the data was asked first to identify the 14 regressors in the previous studies. Neglecting to incorporate the other variables that were examined and found to be insignificant but never reported will bias the analysis towards finding predictability in returns. Directly addressing the specification search is particularly important in the area of asset predictability since there is very little theory as to the functional form and causes of predictability causing the search for predictability to be rather ad hoc.

To ascertain the importance of model uncertainty along the simplest of dimensions. this paper considers only linear models. In particular the focus is on the choice of regressors to be included in the model. Obviously allowing for other models. such as non-linear. multiple regimes. or structural break models will only increase uncertainty over the predictability of regressors. In addition to the specification of the model. the specification of the loss function is also considered. In the literature the loss function is usually specified as the power utility function. though there is little reason to believe that the power utility function is an accurate way of modeling an investor's portfolio allocation problem. Though use of the power utility function may be convenient and appropriate in constructing equilibrium theory. it may be inappropriate for other tasks such as measuring the information content of asset returns.

To begin. nevertheless. assume the power utility function serves as a relevant loss function and that the investor chooses amongst a linear model of predictable stock returns as in equation 3.1. and a model of unpredictable stock returns. This choice between models is represented by placing probability mass in the prior density that the slope parameters are all equal to zero. thereby allowing the posterior to also contain point mass at zero. Allowing probability mass to accumulate at zero could cause large changes in the allocation between the risky and the risk-free asset if probability mass builds up on zero. One possible reason for this is the known sensitivity of the present portfolio allocation problem to the level of expected return. In the case that point mass is not allowed. even though probability mass may accumulate around the slope parameters being zero. the fact that it is always epsilon away from zero will cause the conditional mean to differ from the unconditional mean. If point mass is included in the prior. however. it is conceivable that the conditional mean becomes equal to the unconditional mean as prior beliefs are updated and the candidate regressors are dropped from the formation of the predictive density of returns.

To see why this might be important. consider the KS investor who uses a predictive regression model with an intercept and one regressor to analyze 708 months of data and obtains an R^2 of 0.6% [2] . Notice in Table 3.1 that an investor with a risk aversion parameter of two would allocate only 93% of her wealth to the risky asset instead of 99% after observing an X_T that gave a forecast for r_{t+1} that was one fitted regression standard deviation below the sample mean. Also note that a similar investor with a risk aversion parameter of 5 would reduce her portfolio weight from 63% to 37%. In fact the portfolio allocation varies by 51 percentage points depending on whether the forecast for r_{t+1} is one fitted regression standard deviation above or below the

[2] This is the R^2 of a regression of excess returns on an intercept and the dividend yield. a variable found in many studies to be significant when regressed with other variables. See. for instance. Campbell (1991).

sample mean. Even if the investor has previously observed 750 months of data with an $R^2 = 0$ before observing the 708 month sample with an $R^2 = 0.6\%$. the variation in the portfolio allocation is still 26 percentage points. Perhaps allowing point mass to accumulate at zero will dampen the effect of observing X_T on the optimal portfolio weights..

More formally. let $\underline{p}$ be the prior probability[3] the investor assigns to the event that all the slope coefficients are equal to zero. Then $1 - \underline{p}$ represents the probability that returns can be forecast using the predictive regressors. In terms of the prior density on the parameters. the investor believes that there is probability $\underline{p}$ that the true value of all the slope parameters is zero (returns are unpredictable) and probability $1 - \underline{p}$ that the slope parameters are distributed multivariate normal. $N(\underline{\beta}, \underline{\sigma}^2\underline{\Sigma})$. where $\underline{\beta}$ is the prior mean for the intercept and k slope parameters and $\underline{\Sigma}$ is the covariance matrix. In the spirit of empirical Bayes. $\underline{\Sigma}$ is assumed to be proportional[4] to the OLS estimate. $(X'X)^{-1}$. The prior distribution for σ^2 is given by $\underline{\nu\sigma}^2/\sigma^2 \sim \chi^2(\underline{\nu})$. where $\underline{\nu}$ is the degrees of freedom and $\underline{\sigma}^2$ is the mean. Increasing $\underline{\nu}$ is equivalent to making the prior distribution more informative since the degrees of freedom are given as the number of observations. T. minus the number of regressors. k. So a prior distribution on σ^2 with $\underline{\nu} = 100$ and $k = 5$ is equivalent to having calculated σ^2 after observing 105 monthly returns. The larger $\underline{\nu}$ is. the more important the prior distribution becomes

[3] Underscores represent priors and overscores represent posteriors.

[4] The OLS estimate is multiplied by $\frac{T}{T_0}$. the ratio of the observed sample size to the degrees of freedom of the prior distribution. This is done so that a more informative prior as given by a larger T_0 will have a smaller covariance matrix.

in relation to the observed data set.

Instead of choosing only one of the two models to use in forming the predictive density of returns, the investor mixes over the two models according to $\overline{p}$, the point-mass in the posterior for $\beta = 0$. That is, if $\overline{p} = 1/2$ then in forming the predictive density, half of the draws will come from the model of no predictability and the other half will come from the model of predictability. It is common in the literature to perceive there to be no "true" economic model, as in Pastor (2000). As such, a search for the true model is seen as a fruitless exercise. Instead economic agents are seen as using the information in competing models to help specify predictive densities, not as making accept or reject decisions amongst hypothesis. Each model is considered potentially useful in understanding the underlying process along some dimension and the agent's problem is determining how much information each model contains and should be incorporated in forming the predictive density. This is similar to the idea of encompassing regressions in a frequentist framework. In general the method in forming the predictive density is as follows:

1. Draw from a uniform (0,1) distribution. If the draw is greater than $\overline{p}$, then draw the full β, otherwise only draw the intercept α.

2. Combine the draw on β (or just α if the model of no predictability was selected) with the vector of regressors X_T to get the draw on r_{T+1}.

3. Continue until convergence in the distribution is achieved.

Forming the posterior on the model of unpredictability is straightforward, but does require the specification of a prior density. Prior densities are usually thought of as having been formed by investors through economic theory or the observation of other data sets not known to the econometrician. Once we have specified a particular prior, the investor is allowed to update her belief in the model of predictable returns versus that of unpredictable returns by observing the data. This is done by re-evaluating the amount of point mass at zero in the posterior density for the slope parameters, forming the posterior probability $\overline{p}$. To update the point mass in the prior we must calculate the Bayes factor, which is given as

$$BF = \frac{\Pr(\beta \neq 0|\Phi_T)}{\Pr(\beta = 0|\Phi_T)} = \frac{p(\underline{\beta}|\sigma) \cdot p(\underline{\sigma}) \cdot p(\Phi_T|\beta,\sigma)}{p(\underline{\alpha}|\sigma) \cdot p(\underline{\sigma}) \cdot p(\Phi_T|\beta,\sigma)} = \frac{\int\int p(r,\beta,\sigma|\Phi_T)d\beta d\sigma}{\int\int p(r,\alpha,\sigma|\Phi_T)d\beta d\sigma}$$

$$= \frac{\Gamma(\overline{\nu}_p/2)/\Gamma(\underline{\nu}/2)}{\Gamma(\overline{\nu}_u/2)/\Gamma(\underline{\nu}/2)} \cdot \frac{(|\underline{\Sigma}_p^{-1}|/|\overline{\Sigma}_p^{-1}|)^{-1/2}}{(\underline{\nu}/\overline{\nu}_u)^{-1/2}} \cdot \frac{(\underline{\sigma}_p^2)^{\underline{\nu}/2}}{(\underline{\sigma}_u^2)^{\underline{\nu}/2}} \cdot$$

$$\frac{[(T-K)s_p^2 + (b-\underline{\beta})'X'X(b-\underline{\beta}) + (\underline{\beta}-\overline{\beta})'\underline{\Sigma}_p^{-1}(\underline{\beta}-\overline{\beta})]^{-\overline{\nu}_p/2}}{[Ts_u^2 + T(\alpha-\underline{\alpha})^2 + \underline{\nu}(\underline{\alpha}-\overline{\alpha})^2]^{-\overline{\nu}_u/2}}$$

where subscripts p and u denote the models of predictable and unpredictable returns, α is the mean return in the observed data set, $\underline{\sigma}^2$ is the prior variance of the error term, s^2 is the variance in the error for the observed data given the model, b is the OLS parameter estimates for the observed data given the model, and T is the number of observations.

The Bayes factor represents the odds, given the data, that the model of predictability is the return generating process. The Bayes factor itself is defined as the ratio of the marginal densities of returns for the two models, where the marginal densities can be thought of as weighted averages of the likelihoods with the weights given by the prior density on each model's parameters. Calculating the Bayes factor will allow the point mass on $\beta = 0$ in the prior density, $\underline{p}$, to be updated to give the point mass in the posterior as

$$\overline{p} = \frac{\underline{p}}{\underline{p} + (1 - \underline{p})BF}$$

where $\overline{p}$ denotes the posterior probability that all the slope parameters equal zero conditional on the observed data. It should be noted that the Bayes factor is sensitive to the specification of the priors since the priors are used in weighting the likelihoods. The analysis in KS uses an informative prior that is equivalent to the investor having observed 162 years of data with an $R^2 = 0$. The results of the portfolio allocation problem will potentially be different, however, when this informative prior is not only used in forming the posterior on the parameters but also in evaluating which of the two models should be used in forming the predictive density of returns. The other prior used in KS is a diffuse, or flat prior. As Lindley (1957) points out, the Bayes factor is not defined when the prior is not a proper, informative density. That, of course, means that the diffuse prior often used to map problems from a Bayesian to a frequentist framework, like in KS, is not possible. In this paper I will refer to the diffuse prior as one that has 700 fewer degrees of freedom than the informative prior.[5]

[5]In order to maintain the implied prior on R^2, the number of hypothetical observations in the informative prior is a function of the number of regressors. As noted in KS, the addition of one

The analysis when there is point mass in the prior unfortunately does not solely depend on the prior and a few sample statistics from the data, so the solution can not be calculated as a closed form solution of sample statistics[6]. Here the results will depend on the particular data set since the structure of the data will affect the calculation of the Bayes factor and hence the calculation of the predictive density of returns. To initially avoid issues common in the literature of poorly behaved data (unit roots, multicollinearity, endogeneity, etcetera) and to concentrate only on the effect of uncertainty about the model in the portfolio allocation problem. I simulate data to use in calculating the Bayes factor. Simulated data is constructed for a number of variables that have been analyzed for their ability to predict market returns. A list can be found in Appendix A along with regression results. The simulated return, dividend yield, and dividend growth data are calibrated to the observations of the NYSE value weighted index from 1935 to 1993. The other simulated series also are calibrated to the same time period using data obtained from the St. Louis Federal Reserve.

Returning to the simple case of a single regressor and an R^2 of 0.6%. Table 3.2 contains the optimal portfolio weights obtained from the calibrated data using the informative prior found in KS and p = 10%. As can be seen from the table, the

regressor roughly implies that the prior distibution should contain 50 more hypothetical observations so as to maintain the prior over the model's R^2 constant. Here the informative prior is based off of 700 observations, meaning that the informative prior when $k = 1$ is 750.

[6] In Kandel and Stambaugh's paper only the R^2, the number of observations, the number of regressors, the mean, and the standard deviation are needed to solve the investment allocation problem.

difference between the portfolio weights is reduced once point mass that the slope parameters is zero is placed in the prior density. In the case of the informative prior and a risk aversion parameter of 5, the investor now has only a 15 percentage point difference in her optimal portfolio allocation conditional on the regressors predicting a one standard deviation[7] above or below the sample mean for the fitted regression. Though the results do not converge to the case of no predictability, notice that the posterior probability on the model of no predictability is now a full 57%. The investor finds little evidence that returns are predictable and updates her beliefs in favor of unpredictable asset returns by 47 percentage points. In fact, as can be seen from the Bayes factor, had the investor observed the data with a prior belief that returns are as likely to be predictable as they are unpredictable, the investor's portfolio allocation would be almost unchanged by the observation of the predictive regressors at time T since the posterior probability on the model of no predictability would be 92% instead of 57%.

The results from the informative prior, however, may be misleading for two reasons. First, the informative prior is formed after the investor observes a hypothetical sample that has an R^2 of zero. In Table 3.2 the results are based on a prior biased in favor of the model of predictability, represented by $\underline{p} = 10\%$. Had the investor

[7] The deviations from the regression mean must also be simulated. Deviations are also calibrated to the data and reported results are for those deviations that result in the largest movement in portfolio weights. In general, deviations due mainly to a single variable cause less change in portfolio weights since large deviations from the mean of X introduces more variance into the predictive density.

calculated the Bayes factor after observing the hypothetical data used to construct the informative prior, they would have shifted over 99% of the point mass to the model of no predictability, even if they used a prior of $\underline{p} = 1\%$ on the model of no predictability to begin with. That means that the appropriate prior on the model of predictability for the investor with the KS informative prior would be $\underline{p} = 99\%$. Thus, in reality, the investor using the informative prior would not alter her portfolio allocation at all after forming her posterior and then observing the following 708 months of data and the predictive regressors at time T. As noted previously, even the investor with the diffuse prior would have shifted a majority of the probability mass onto the model of no predictability, and had she used a prior of $\underline{p} = 50\%$ she too would leave her portfolio allocation largely unchanged.

The second reason the results may be misleading is the use of the prior in calculating the Bayes factor. Keep in mind that the Bayes factor is similar to a weighted sum of the likelihood ratios for all possible parameter values where the prior distribution serves as the weighting function. This means that the calculation of the Bayes factor will be sensitive to the specification of the prior. Table 3.3 illustrates the sensitivity of the posterior probability to the Bayes factor when different priors are used. Notice that calculating the Bayes factor with an informative prior distribution centered at the data's OLS estimates gives a Bayes factor in favor of the model of predictability. Using $\underline{p} = 10\%$ as the prior on the model of no predictability gives a posterior of $\underline{p} = 1\%$ when combined with this Bayes factor. Furthermore, recalculating the Bayes factor with a diffuse prior distribution centered at the observed data's OLS estimates gives a posterior probability of $\overline{p} = 4\%$ when the prior on the model of predictability is specified as $\underline{p} = 10\%$. Both the diffuse and informative priors centered at the OLS estimates reduce the probability mass on the model of no predictability, whereas the informative prior centered at zero increases the point mass on the model of no predictability from $\underline{p} = 10\%$ to $\overline{p} = 57\%$. This implies that an investor with no previous bias towards a model of no predictability versus one of predictability would, after observing the data, conclude that stock returns were more likely to be predictable than not, despite the low R^2 of 0.6%.

The preceding shows that the use of point mass can avoid possible counter intuitive cases such as large portfolio reallocations in the face of very low R^2, but only when the prior is carefully chosen to bias against the model of predictability. Despite the low R^2 the slope parameter on the independent variable remains significant so that only prior specifications where little probability mass is placed on the observed OLS estimates will result in a Bayes factor favoring the model of unpredictable returns. But what about when the slope variable is significant and the R^2 is similarly larger? By how much does the portfolio allocation change when point mass on the model of predictability is included in the prior and the probability mass remains centered around zero? Repeating the same analysis when the R^2 is 2.7% [8] is presented in Table 3.4. Here the results are very different than their Table 3.2 counterparts. Despite the

[8]This comes from a regression of monthly returns on the one year T-bill moving average and an intercept term.

low R^2, the Bayes factor results in a posterior distribution that places almost no weight on the model of unpredictability even though the prior distribution used to calculate the Bayes factor is centered at zero. The reason is the significant t-statistic on the slope parameter causes the point mass to accumulate on the model of predictability despite the specification of the prior. Here the results are indistinguishable from the base line case of KS. Notice that in spite of the model's inability to predict return with much precision, as evidenced by the low R^2, the investor completely divests herself of the risky asset after observing a conditional mean one regression standard deviation below the mean. This is not surprising given that the F-statistic from the regression is significant at the 99.9% level. Indeed, there is an obvious relation between the posterior odds and the F-statistic since both attempt to gauge the fit of a null hypothesis against some known alternative hypothesis.

In the case of predictability in asset returns, it is known that an extensive search has been carried out to find variables that predict returns. What is the impact on the portfolio allocation problem if we consider the same regressor with other statistically insignificant regressors? Since we know the regressors that appear in the published literature have been selected from dozens and possibly hundreds of other candidates, it is important to know the effects of the well understood pre test estimator problem. How does inference change if we are able to observe the econometrician's path in finding which variables to include and which to exclude from the regression? For instance the reported regression of dividend yield on returns in Appendix A results in an F-statistic of 4.8 despite an R^2 of 0.6%. That same R^2 when the regression contains three regressors results in an F-statistic of 1.4, which is only significant at the 25% level. Nevertheless the KS investor with a diffuse prior rebalances her portfolio by as much as 14 percentage points unless she is allowed to place point mass on the model of no predictability in asset returns. Without point mass in the prior the inference between classical statistics and the Bayesian method can be quite large. The Bayes factor, like the F-statistic, will also be effected by the number of regressors on the right hand side. Increasing the number of regressors while holding the R^2 constant will reduce the Bayes factor in favor of the model of unpredictable returns. Allowing the investor to shift probability mass to the model of no predictability will largely reduce the difference in inference.

As a first pass, notice the results from Table 3.5 which lists the calculated posterior probability of predictable asset returns for different values of K and prior probabilities on the model of unpredictability. The R^2 of the samples remain near the 2.7% of the previous analysis, but now statistically insignificant variables have been included on the right hand side of the regression. The first entry comes from a regression of monthly returns on two variables and an intercept. The portfolio for the investor is determined by the model of predictability since the posterior probability on the model of unpredictable asset returns is close to zero, even if the prior probability on the model of unpredictable returns is 50%. This is not surprising given that the sample's F-statistic is 13.9, significant at even the 99% level. On the other extreme is the case of 10 regressors. which produces a posterior probability of 99% on the model of unpredictability even when the prior probability is less than one percent. In Table 3.5 it is clear that the posterior odds on the model of unpredictability are closely related to the model's statistical significance as measured by the F-statistic. Not surprisingly, the investor tends to favor the model of predictable returns when the F-statistic is significant. In the case of an R^2 of 2.7%. as long as the reported regressor was chosen from among 5 or more regressors it appears that an investor would keep the model of unpredictable returns. In fact. as long as the regression uses 10 or more regressors the R^2 could be over 5% and still the investor would reject the model of predictable returns.

3.3.2 Point Mass on the Nested Models

To get a better idea of how model uncertainty alters the portfolio allocation problem we might also ask the question how one chooses which regressors to include in the model of predictability to begin with? Given that there are K potential regressors to be included in the model. there are in fact 2^K candidate models or combinations of those regressors. But how does an investor choose which model (or which regressors) should be used? The most obvious approach for a Bayesian agent is to allow for the exclusion of candidate regressors as the investor observes the data. Placing point mass in the prior that any given regressor is in fact uncorrelated with the left hand side variable and should be excluded from the regression will allow the investor to reassess the probability that a regressor should be excluded from the regression equation. This allows the investor to refine the model over time instead of making the stark decision to either reject the model of predictability altogether or accept it as is. such as is done in the previous section.

Take. for instance. the behavior of the investor in the KS framework who observed the results of a regression with 25 variables and an R^2 of 2.5%. Perhaps of the 25 regressors included in the model only the first is found to be important as measured by the t-statistic. Assume that the other 24 regressors have t-statistics of 0.1. and that the realization of the first regressor at time T is equal to its sample mean. Yet if the other 24 regressors are one standard deviation below their sample means it is quite conceivable that the predictive density as defined by all 25 regressors could have a conditional mean close to one fitted regression standard deviation below the unconditional mean. As seen in the previous section. the investor would not alter her portfolio allocation if she considers both the models of predictability and unpredictability. unlike the KS investor. But what if the reverse were true? What if the statistically significant variable is one standard deviation below its mean while the other variables are one standard deviation above theirs?

The problem is that in evaluating a model we are primarily concerned with the fit of the model. perhaps as measured by an F-statistic. However when we think about decision making and forecasting. the model needs to be evaluated on a regressor by regressor basis since including spurious regressors can greatly alter the predicted outcome (predictive density) even if the parameter estimates of the true model's regressors are unbiased. That is, the investor should be allowed to pick and choose which regressors should be excluded from forming the conditional mean as the data is observed. Forcing the investor to make a simple decision between the models of unpredictability and predictability may be as misleading as forcing the investor to only consider the model of predictability. Allowing the investor to choose amongst regressors greatly increases the model space by allowing the investor to now choose amongst 2^k models, each with a different prediction for the next month's return. Such a large model space will undoubtedly contain models that predict both that the conditional return will be higher than the unconditional expected return and those that predict it will be lower.

By allowing for the larger model space the data is forced to provide more answers (which regressors should be included and what the parameter values are). As the number of answers required from the data grows, the less perceived precision there is in those answers. At the same time, however, this allows the investor to make decisions while ignoring potentially misleading variables and should produce more accurate inference. Unlike before when the probability of no predictability was chosen directly, here the total point mass on the model of no-predictability is given by the product of the probabilities that each predictive regressor is excluded from the regression model

$$\prod_{i=1,k} \underline{p}_i$$

where $\underline{p}_i$ is the probability that variable i should be excluded from the regression. In the case of 25 regressors, and given that $\underline{p}_i = 0.5$ for all i, the probability of no-predictability is about 0.0000003, meaning that the investor, before observing the data, believes that predictability in returns is a virtual certainty. The real question is how much predictability is perceived to be in the data when the investor is only uncertain about which variables should be included in the regression, not whether or not there is predictability? The answer to this question will help explain if KS results were primarily the result of the investor's (perceived) knowledge of the data generating process or because the data itself strongly reject the notion of no predictability.

In keeping with KS, let the investor's prior distribution for each slope parameter β_i, conditional on $\beta_i \neq 0$, is given by $N(\underline{\beta}_i, \tau_i^2)$. The prior distribution of σ^2 is the standard form:

$$\underline{\nu\sigma}^2/\sigma^2 \sim \chi^2(\underline{\nu})$$

This gives us proper priors and, with normally distributed error terms, conjugate priors. Following the notation in Geweke (1994), define $z_t = y_t - \Sigma_{l \neq j}\beta_l x_{tl}$. The conditional distribution of each slope parameter is given by combining the likelihood function kernel

$$e^{-\Sigma(z_t - \beta_j^2 x_{tj})^2/2\sigma^2} if\ \beta_i \neq 0,\ e^{-\Sigma z_t^2/2\sigma^2} if\ \beta_i = 0$$

with the prior distributions on σ and β_i. The resulting kernel density for β_i, $\beta_i \neq 0$

is

$$e^{-\Sigma(z_i - bx_{ij})^2/2\sigma^2} \cdot e^{-(\beta_i - \overline{\beta})/2\sigma_*^2} \cdot e^{-(b^2/2\varsigma^2 + \underline{\beta}_i^2/2\tau_i^2 - \overline{\beta}_i^2/2\sigma_*^2)} \tag{3.5}$$

where $b = \Sigma_{i=1}^{n} x_{ij} z_i / \Sigma_{i=1}^{n} x_{ij}^2$, $\varsigma^2 = \sigma^2/\Sigma_{i=1}^{n} x_{ij}^2$, $\sigma_*^2 = (\varsigma^{-2} + \tau_i^{-2})$, and $\overline{\beta}_j = \sigma_*^2(\varsigma^{-2}b + \tau_i^{-2}\underline{\beta}_i)$. Notice that if $\beta_i \neq 0$ and σ is known, then β_i is distributed normally. The distribution for σ, conditional on all the betas, is given by the familiar form

$$[\underline{\nu\sigma}^2 + (Y - X\beta)'(y - X\beta)]/\sigma^2 \sim \chi^2(\underline{\nu} + T) \tag{3.6}$$

Sampling from the posterior density is straightforward, though convergence is slow[9] . Gibbs sampling is used to construct the posterior since it is not of standard form. The draws are started using the OLS estimates for the betas and σ. To construct the next draw on the betas, the data is used to calculate the Bayes factor in favor of the model $\beta_j \neq 0$ versus $\beta_j = 0$ conditional on the other betas and σ. The Bayes factor is in turn is used to update the probability that regressor i should be excluded from the regression equation, $\underline{p}_i$. The Bayes factor in favor of the model $\beta_j \neq 0$ versus $\beta_j = 0$ is given by

$$BF = e^{\overline{\beta}_i^2/2\sigma_*^2 - \underline{\beta}_j^2/2\tau_i^2} \cdot \sigma_*/\tau_j$$

To update the probability that regressor j should be excluded from the regression equation the posterior odds ratio $\overline{p}_j$ for the model that $\beta_j = 0$ is calculated using the

[9]Geweke (1994) notes that computational time is "roughly proportional to the cube of the number of regressors." Cremers (2000) and Avramov (2002) use 14 regressors, but Cremers (2000) does not maximize expected utility while Avramov (2002) does not report statistics on convergence. I only do 5 regressors because of problems with convergence in this setting. In this case there are $2^5 = 32$ models being evaluated. With 5 regressors the number of draws on returns to ensure convergence is roughly 5 million, making maximization of the expected utility very difficult and computationally costly.

Bayes factor (notice that $\overline{p}_j$ will be different each draw since the Bayes factor being calculated is conditional on the values of the other betas. which themselves change with each draw)

$$\overline{p}_j = \frac{\underline{p}_j}{\underline{p}_j + (1 - \underline{p}_j)BF} \tag{3.7}$$

A draw is then made from a uniform distribution on the interval [0.1]. If the draw is less than $\overline{p}_j$. then β_j is set equal to zero and the next slope parameter is drawn. If the draw is greater than $\overline{p}_j$. then the slope parameter β_j is drawn from a conditionally normal distribution. $\beta_j \sim N(\overline{\beta}_j, \sigma_*^2)$. After each of the K slope parameters has been chosen. then σ is drawn conditional on the vector of the betas. Given the new betas and σ. the process is repeated. The process is continued until convergence is achieved.

Since the goal in this analysis is to maximize the expected utility. the predictive density for the month ahead return must be formed. Given the draws on the slope parameters and σ (the posterior density on the parameters). the predictive density for returns is constructed by adding in the predictive likelihood and integrating over the parameter space as in equation 3.4. This is done by taking each of the draws on beta and calculating the expected return given the vector of predictive variables X_T and adding in a normally distributed error term . $\epsilon_t \sim N(0, \sigma^2)$. as in equation 3.1. For each of the draws on the parameters I obtain 5.000 returns since I take 5.000 draws on the error term. So if there are N draws on the parameters there will be $5,000N$ returns forming the empirical predictive distribution for the one month ahead continuously compounded return. The $5,000N$ draws from the predictive density are then used to perform the integration in equation 3, giving the expected utility as a function of the portfolio weights. Maximization of the utility function is then carried out using a simple quadratic interpolation method.

Table 3.6 represents the portfolio weight given to the risky asset by an investor who observes 59 years worth of data with an $R^2 = 3.3\%$ and then forms a predictive density given the current realization of 2 predictive regressors. Results are listed for the investor that only considers the model of predictability as in KS using a diffuse prior, along with results for an investor who places 10% probability on the model of unpredictability and equally distributes the remaining 90% among the 2^k models of predictability, using both a diffuse and an informative prior. Here it is important to note that the simulated standard deviation plays an important role in the analysis since the portfolio weights will be sensitive to which variables are farthest from their mean. One important point to keep in mind is that there are numerous values of the regressors that will result in a one fitted regression standard deviation from the mean. Each set of values for the regressors, however, will result in a different portfolio allocation. Because of the optimal portfolio's sensitivity to the regressor values, all tables contain results for the simulated deviations that result in the largest movement in the optimal portfolio allocation[10].

Table 3.6 reveals that model uncertainty does reduce the portfolio movement

[10] Most of the simulated deviations analyzed result in small portfolio movements, usually around only 5%. Due to the computational costs, however, it is not feasible to report the distribution of the optimal portfolio weights for different simulated deviations.

caused by the observed vector X_T, but there remains significant movement in the portfolio weights. Adding a third variable as in Table 3.7 does little to change the results. Even though only one of the variables is found to be significant, the investor still makes large changes to her portfolio, merely choosing to discard the other regressors. Unlike in the previous section where the investor was forced to either accept all regressors or reject all regressors, it doesn't appear that the search for statistically significant variables is as important for the inference drawn by the investor. Before, the investor would reject the model of predictability despite statistically significant regressors. Now the investor merely tosses out the insignificant variables.

3.4 Choice of the Loss Function

As mentioned previously, the loss function in addition to the likelihood function (model) is important in measuring the economic information in the data. In the case of the portfolio allocation problem it is known that the power utility function is sensitive to the mean but much less sensitive to the variance. Since the Bayesian agent must be a close approximation to the marginal investor for meaningful inference about the time-series properties of the investment opportunity set to be made, it is reasonable to ask if the power utility function is a reasonable choice as the loss function. More generally, what are the properties of the power utility function in the investment allocation problem and how does the choice of the utility function influence the perceived degree of predictability in returns? A related paper is Aït-Sahalia and Brandt (2001) who look at the implications for general portfolio advice from portfolio advisors when assets are thought to be predictable and investors potentially have non-expected utility functions, though the paper does not attempt to address the appropriateness of different utility functions..

As can be seen in Table 3.8, the sensitivity to the mean and the variance will in part be determined by the chosen risk aversion parameter. Take as a starting point an investor with a risk aversion parameter of two that observes a risky asset with a thirty basis point risk premium and a variance of 0.0005. Notice that increasing the variance by a factor of sixteen reduces the portfolio allocation to the risky asset from 99% to 86% when the expected return on the risky asset is held constant. On the other hand, reducing the expected return by 28 basis points while keeping the variance constant reduces the portfolio allocation to the risky asset to 89%. In fact by adding 28.5 basis points to the expected return of 0.3% the variance would have to increase almost 32 times over to induce a portfolio allocation of 86%. Such tolerance of volatility in stock returns is likely inconsistent with the preferences of the average investor, yet possibly consistent with those of the marginal investor, often assumed to be large investors or institutions.

Turning attention to the case of an investor with a risk aversion parameter of five, however, and a risky asset that carries a two basis point risk premium and a variance of 0.0005, notice that the portfolio allocation drops only 8% points when the volatility is larger by a factor of sixteen. Yet adding twentyeight basis points to the risk premium increases the portfolio allocation to the risky asset to 99%. Such sensitivity to the mean and insensitivity to the variance is important in the asset allocation problem because a larger variance will always lead to a larger predicted movement in the expected mean of the return. Recall that in the case of no model uncertainty that a predicted one regression standard deviation from the mean can be written as

$$\widehat{r}_{T+1} = \overline{r} \pm \sigma_{\overline{r}} \cdot \sqrt{R^2}$$

Notice that increasing the variance of the returns $\sigma_{\overline{r}}$ while holding the regression R^2 constant will increase the deviation from the mean. Though classical statistics would not distinguish between the two cases. the Bayesian investor with a power utility function would. An implication is that investors perceive the economic importance of predictability as more significant for high variance stocks than low variance stocks.

Though the power utility function is ubiquitous in the literature and easy to work with. it is not clear that it is an appropriate loss function to use in assessing the economic information content of the predictability of asset returns. The desirable properties of the function such as risk aversion and isoelasticity with respect to wealth have made it theoretically popular and useful for generating equilibrium results. Nevertheless it is potentially inappropriate for making inferences about the decision process for individual investors. More thought and research are needed to ascertain the desirable properties of a utility function when the goal is inference about asset predictability as opposed to equilibrium or general financial theory.

3.5 Conclusions

Predictability in returns is difficult to measure, in part due to the lack of theory about what the sources of predictability are. Classical statistical tests are not overpowering, but in the face of mounting studies that claim to have found predictability in asset returns the sentiment that returns are predictable is rapidly increasing in finance. Understanding the source of predictability is important in developing some type of theory to explain either the inefficiency of the market or the nature of the stochastic investment opportunity set. Kandel and Stambaugh (1996) introduce a Bayesian metric to analyze asset return predictability. Instead of asking the data to make an accept or reject decision about the hypothesis of unpredictable returns. the data is used by an investor to update her beliefs about asset returns and allocate wealth between a risky and a risk free asset. To understand the economic importance of predictability it is important to know how the information in the data is used by the investor in forming her portfolio. Doing so requires the specification of a model and a loss function.

Though many theoretical models use the power utility function it is not clear that such a utility function adequately models the behavior of the marginal investor. This makes specification of the loss function difficult. but due to the wide spread theoretical use of the power utility it is the natural candidate. Choosing a different loss function. however. will lead to different inference about the importance of the predictability in asset returns. It is possible to not only construct a utility function that finds predictability in asset returns important. but possibly to construct one that finds the opposite result. In general the more sensitive the utility function is to the expected return relative to the variance of returns. the more important asset predictability will seem. More thought and research needs to be spent on the appropriateness of the use of the power utility function in this context.

Besides the loss function. a likelihood (model) must also be specified through which to view the data. Inference will not only be sensitive to the chosen utility function but also the model being considered. To understand how investors' would use or interpret the predictability observed in the data it is important to consider how investors choose amongst models in interpreting the predictability. When investors are forced to only consider one model. the result can be large changes to their portfolio allocations. even when they observe an R^2 as low as 0.6% from a regression of three variables on observed returns. Allowing an investor to choose between models can largely solve this problem. greatly reducing the perceived amount of information. In fact an investor may now reject the idea of predictability. making little to no change to her portfolio since she may now place point mass in the posterior on the model of unpredictable returns. For higher R^2 something similar happens as K. the number of regressors on the right hand side. increases. As K increases while holding R^2

constant. the investor becomes increasingly likely to reject the model of predictability when forming her portfolio. Such a result is to be expected since increasing K also increases the likelihood that the obtained R^2 is the result of luck. not because the regressors have any predictive power.

Forcing the agent to reject all predictive variables. however. is also misleading. An investor is not making an accept or reject decision amongst models. but may want to retain certain regressors while discarding others. The question the investor faces is not what the true model is. rather what variables are valuable in understanding the return generating process. Though increasing the number of variables on the right hand side while holding R^2 constant does reduce the perceived predictability in return. portfolio movement is still significant in cases and undeniable.

$K = 1, R^2 = 0.6\%, T = 708$

Diffuse Prior

δ	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	93%	99%	99%	99%	99%	2
	37%	50%	63%	75%	88%	5
δ	-1	-0.5	0	0.5	1	

Informative Prior (T_0 = 750)*

δ	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	99%	99%	99%	99%	99%	2
	52%	58%	64%	71%	78%	5
δ	-1	-0.5	0	0.5	1	

* Equivalent to having observed 750 months of data with an R^2 = 0

δ = Regression Standard Deviation
A = Risk Aversion Parameter
K = Number of Slope Coefficients
T = Number of Months in the Sample
T_0 = Number of Observations in Prior

TABLE 3.1. Portfolio Weights When the Model is Known (K=1,R2=0.006)

$K = 1, R^2 = 0.6\%, T = 708, \underline{p} = 10\%$

Diffuse Prior ($T_0 = 50$)*

	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	97%	99%	99%	99%	99%	2
	47%	56%	64%	73%	81%	5
δ	-1	-0.5	0	0.5	1	

$\overline{p} = 23\%$ Bayes Factor = 0.372

Informative Prior ($T_0 = 750$)*

	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	99%	99%	99%	99%	99%	2
	57%	61%	64%	69%	73%	5
δ	-1	-0.5	0	0.5	1	

$\overline{p} = 57\%$ Bayes Factor = 0.084
F-Statistic = 4.18

* Equivalent to having observed T_0 months of data with an $R^2 = 0$

See Table 3.1 for other definitions
$\overline{p}$ = Posterior probability
$\underline{p}$ = Prior Probability of Unpredictable Returns

TABLE 3.2. Portfolio Weights When the Model is Unknown (K=1,R2=0.006)

$K = 1, R^2 = 0.6\%, T = 708, \underline{p} = 10\%$

Diffuse Prior ($T_0 = 50$)*

Posterior Probabilities $\overline{p}$ on the Model of Unpredictable Returns given Prior $\underline{p}$

$\overline{p}$	0.3%	4%	27%	77%	97%
$\underline{p}$	1%	10%	50%	90%	99%

Informative Prior ($T_0 = 750$)*

Posterior Probabilities on the Model of Unpredictable Returns given the Prior $\underline{p}$

$\overline{p}$	99%	99%	99%	99%	99%
$\underline{p}$	57%	61%	64%	69%	73%

* Equivalent to having observed T_0 months of data with an $R^2 = 0$

TABLE 3.3. Posterior Probabilities for Different Priors

$K = 1, R^2 = 2.7\%, T = 708, \underline{p} = 10\%$

Diffuse Prior ($T_0 = 50$)*

δ	Optimal Portfolio Weights					A
	0%	99%	99%	99%	99%	1
	0%	80%	99%	99%	99%	2
	0%	32%	69%	99%	99%	5
δ	-1	-0.5	0	0.5	1	

Informative Prior ($T_0 = 750$)*

δ	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	82%	99%	99%	99%	99%	2
	33%	50%	69%	86%	99%	5
δ	-1	-0.5	0	0.5	1	

$\overline{p} = 0.25\%$ Bayes Factor = 0.084
F-Statistic = 19.2

* Equivalent to having observed T_0 months of data with an $R^2 = 0$

See Table 3.1 for other definitions
$\overline{p}$ = Posterior probability
$\underline{p}$ = Prior Probability of Unpredictable Returns

TABLE 3.4. Portfolio Weights When the Model is Unknown (K=1,R2=0.027)

$R^2 = 2.7\%. T = 708$

Posterior Probabilities on Model of Predictable Returns

K	2	3	4	5	10	$\underline{p}$
	1%	33%	70%	98%	99%	10%
	7%	82%	95%	99%	99%	50%
	40%	97%	99%	99%	99%	90%

K	2	3	4	5	10
Bayes Factor	13.9	0.22	0.05	0.0018	5.7×10^{-13}
F-Statistic	10.1	6.5	4.8	3.9	1.8

$R^2 = 5\%. T = 708$

Posterior Probabilities on Model of Predictable Returns

K	2	3	5	10	$\underline{p}$
	0%	0%	2%	99%	1
	0%	0%	15%	99%	2
	0.004%	0.01%	62%	99%	5

K	2	3	5	10
Bayes Factor	18.121	6.054	5.6	1.3×10^{-9}
F-Statistic	18.7	12.4	7.3	3.6

$\underline{p}$ = Prior Probability of Unpredictable Returns

TABLE 3.5. Posterior Probabilities for Different Size Models

$K = 2, R^2 = 3.3\%, T = 708, \underline{p} = 10\%$

Diffuse Prior - Kandel & Stambaugh

	Optimal Portfolio Weights					A
	0%	99%	99%	99%	99%	1
	0%	70%	99%	99%	99%	2
	0%	28%	68%	88%	99%	5
δ	-1	-0.5	0	0.5	1	

Diffuse Prior ($T_0 = 50$)

	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	62%	99%	99%	99%	99%	2
	25%	46%	68%	88%	99%	5
δ	-1	-0.5	0	0.5	1	

Informative Prior ($T_0 = 750$)

	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	99%	99%	99%	99%	99%	2
	50%	59%	68%	75%	83%	5
δ	-1	-0.5	0	0.5	1	

Posterior Probability	Regressor 1	Regressor 2
Regressor is Excluded	56%	0.001%
T-Statistic	0.73	-4.6

TABLE 3.6. Portfolio Weights When the Regressors are Chosen (K=2)

$K = 3, R^2 = 3.3\%, T = 708, \underline{p} = 10\%$

Diffuse Prior ($T_0 = 50$)

	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	69%	99%	99%	99%	99%	2
	27%	46%	68%	84%	99%	5
δ	-1	-0.5	0	0.5	1	

Informative Prior ($T_0 = 850$)

	Optimal Portfolio Weights					A
	99%	99%	99%	99%	99%	1
	99%	99%	99%	99%	99%	2
	50%	57%	68%	73%	80%	5
δ	-1	-0.5	0	0.5	1	

Posterior Probability	Regressor 1	Regressor 2	Regressor 3
Regressor is Excluded	75%	0.01%	86%
T-Statistic	0.81	-4.9	0.71

TABLE 3.7. Portfolio Weights When the Regressors are Chosen (K=3)

$A = 2, T = 708, T_0 = 0$

	Optimal Portfolio Weights			Variance
	52%	86%	99%	0.0084
	55%	99%	99%	0.0042
	60%	99%	99%	0.0021
	69%	99%	99%	0.00105
	89%	99%	99%	0.0005
Mean	0.02%	0.3%	0.585%	
δ	-0.5	-0.25	0	

$A = 5, T = 708, T_0 = 0$

	Optimal Portfolio Weights					Variance
	10%	17%	24%	29%	33%	0.0084
	11%	24%	38%	47%	56%	0.0042
	12%	39%	68%	83%	99%	0.0021
	14%	67%	99%	99%	99%	0.00105
	18%	99%	99%	99%	99%	0.0005
Mean	0.02%	0.3%	0.585%	0.77%	0.96%	
δ	-0.5	-0.25	0	0.25	0.5	

δ is the number of standard deviations from the mean

TABLE 3.8. Optimal Portfolio Weights for the Power Utility Function

Appendix A

Descriptive Statistics of Historical Data Used in Simulating New Data

	Averages	Standard Error	Auto-Correlation	Correlation
monthly returns*	0.5%	0.0460	3.3%	100%
dividend yield*	3.9%	0.0135	71.0%	8.2%
dividend growth*	-0.2%	0.1123	-44.0%	9.4%
spread^{+} Aaa - Tbill$^{\wedge}$	2.0%	1.2608	96.0%	11.2%
spread^{+} Baa - Tbill$^{\wedge}$	3.1%	1.5635	97.0%	10.7%
Tbill - MA(1yr)	-0.30	0.9346	91.0%	-16.4%

monthly returns: Continuously compounded monthly return for NYSE stocks
dividend yield: Sum of all dividends for NYSE stocks paid over the last year divided by the current sum of the prices
dividend growth: The growth in the dividend yield over the previous month
spread Aaa: The spread between Aaa bonds and a one month T-bill
spread Aaa-Tbill: The spread between Baa bonds and a one month T-bill
Tbill-MA(1yr): The difference between the 1 month T-bill and its one year moving average

These variables have appeared in several studies. including Campbell (1991) and (1986). Harvey(1987). Chen. Roll. and Ross(1986) just to name a few.
* : Data obtained from CRSP
+ : Data obtained from the St.Louis Federal Reserve Bank
^: Data obtained from Ibbotson Associates 1997 yearbook

Table A.1. Descriptive Statistics of Historical Data

K = 1, R^2 = .006

R Square	0.0067			
Adjusted R Square	0.0054			
Standard Error	0.0463			
F-Statistic	4.81	Significance	0.03	
	Coefficients	Standard Error	T-Stat	P-value
Intercept	-0.0053	0.0054	-0.9832	0.3259
Dividend Yield	0.2819	0.1285	2.1943	0.0285

K = 1, R^2 = .027

R Square	0.027			
Adjusted R Square	0.026			
Standard Error	0.046			
F-Statistic	19.541	Significance	0.00001	
	Coefficients	Standard Error	T-Stat	P-value
Intercept	0.0034	0.0018	1.8648	0.0626
Moving Average	-0.0081	0.0018	-4.4205	0.0000

TABLE A.2. Regression Results for Single Variable Regressions

$K = 3.\ R^2 = .034$

R Square	0.034
Adjusted R Square	0.030
Standard Error	0.046

F-Statistic	8.178	Significance	0.00002	
	Coefficients	Standard Error	T-Stat	P-value
Intercept	-0.007	0.0054	-0.9832	0.3259
Dividend Yield	0.772	0.352	2.193	0.029
Moving Average	-0.009	0.002	-4.607	0.000
Dividend Growth	-0.003	0.006	-0.486	0.627

$K = 5.\ R^2 = .049$

R Square	0.049
Adjusted R Square	0.042
Standard Error	0.045

F-Statistic	7.178	Significance	0.000001	
	Coefficients	Standard Error	T-Stat	P-value
Intercept	-0.0100	0.0061	-1.6463	0.1002
Moving Average	-0.0086	0.0023	-3.7956	0.0002
Dividend Yield	0.3975	0.1358	2.9276	0.0035
Dividend Growth	0.0404	0.0152	2.6535	0.0081
Aaa Spread	0.0062	0.0047	1.3320	0.1833
Baa Spread	-0.0049	0.0038	-1.2877	0.1983

TABLE A.3. Regression Results for Multiple Variable Regressions

Appendix B

Posterior Means for the Model of Predictability

Underscores represent prior distributions and overscores represent posterior distributions.

β is the vector of parameters

Σ is the covariance matrix for β

X is the $Tx(K+1)$ matrix of observations on the regressors and intercept

Y is the T length vector containing the observed returns

v is the degrees of freedom

I_T is the TxT identity matrix

σ^2 is the variance of the error term

$$\overline{\beta} = \overline{\Sigma}(\underline{\Sigma}^{-1}\underline{\beta} + X^T X b)$$

$$\overline{\sigma}^2(v_0 + T) = v_0\underline{\sigma}^2 + (Y - Xb)^T(I_T + X\overline{\Sigma}X^T)(Y - Xb)$$

$$\overline{\Sigma} = (\underline{\Sigma}^{-1} + X^T X)^{-1}$$

REFERENCES

[1] Admati. A.. and P. Pfleiderer, 1988. "A Theory of Intraday Patterns: Volume and Price Variability." The Review of Financial Studies. 1:3-40.

[2] Aït-Sahalia. Y., M. Brandt, 2001. "Variable Selection for Portfolio Choice." Journal of Finance, 56:1165-1628.

[3] Anderson, T.. 1996, "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," Journal of Finance 51:169-204.

[4] Andrews. D., 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." Econometrica. 59:817-858.

[5] Avramov. D.. 2002. "Stock Return Predictability and Model Uncertainty." Journal of Financial Economics. Forthcoming.

[6] Bachelier. L.. 1900. "Theory of Speculation." in Cootner A.P.. (ed.). The Random Character of Stock Market Prices. MIT Press. Cambridge. MA. 1964: Reprint.

[7] Barberis. N.. 2000. "Investing in the Long-Run when Returns are Predictable." Journal of Finance. 55:225-264.

[8] Bessembinder. H. and M. Lemmon. 1999. "Pricing and Gains from Trade in Competitive Electric Power Markets." Working Paper.

[9] Bollerslev. T.. R. Chou and K. Kroner. 1992. "ARCH Modelling in Finance." Journal of Econometrics. 52:5-59.

[10] Brennan. M.. and Y. Xia. 2001. "Asset Pricing Anomalies." The Review of Financial Studies. 19:905-942.

[11] Cagetti, M.. L. Hansen. T. Sargent. and N. Williams. 2002. "Robustness and Pricing with Uncertain Growth." Review of Financial Studies. 15:363-404.

[12] Campbell. J.. 1987. "Stock Returns and the Term Structure." Journal of Financial Economics. 18: 373-399.

[13] Campbell. J.. 1991. "A Variance Decomposition for Stock Returns." The Economic Journal. 101: 157-179.

[14] Campbell. J. and A. Lo and A. MacKinlay. 1997. The Econometrics of Financial Markets. Princeton University Press.

[15] Campbell, J., and R. Shiller, 1988. "Stock Prices, Earnings, and Expected Dividends," Journal of Finance, 43: 661-676.

[16] Carlson, M., and D. Chapman and R. Kaniel and H. Yan, 2001. "The Utility Costs of Consumption/Portfolio Rules in Partial and General Equilibrium," Working Paper, University of Texas.

[17] Clark, P., 1973. "A Subordinate Stochastic Process Model with Finite Variance for Speculative Prices," Econometrica, 41:135-155.

[18] Copeland, T., 1976. "A Model of Asset Trading under the Assumption of Sequential Information Arrival," Journal of Finance, 31:1149-1168.

[19] Cremers, M., 2000. "Stock Return Predictability," Working Paper, New York University.

[20] Crouch, R., 1970. "The Volume of Transactions and Price Changes on the New York Stock Exchange," Financial Analysts Journal, 60:199-202.

[21] DeBondt, W. and R. Thaler, 1985. "Does the Stock Market Overreact," Journal of Finance, 40:793-805.

[22] Donaldson, R.G. and M. Kamstra, 1996. "A New Dividend Forecasting Procedure That Rejects Bubbles in Asset Prices: The Case of 1929's Stock Crash," Review of Financial Studies, 9:333-383.

[23] Epps, T. and M. Epps, 1976. "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture of Distributions Hypothesis," Econometrica, 44:305-321.

[24] Fama, E., 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work," Journal of Finance, 25:383-417.

[25] Fama, E., 1991. "Efficient Capital Markets: II," Journal of Finance, 46:1575-1617.

[26] Fama, E., and K. French, 1988. "Permanent and Temporary Components of Stock Prices," Journal of Political Economy, 96: 246-273.

[27] Fama, E., and K. French, 1992. "The Cross Section of Expected Stock Returns," Journal of Finance, 47: 427-465.

[28] Foster, F., T. Smith, and R. Whaley, 1997. "Assessing Goodness -of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2," Journal of Finance, 52:591-607.

[29] Foster, F., and S. Viswanathan, 1990, "A Theory of the Intraday Variations in Volume, Variance, and Trading Costs in Securities Markets," The Review of Financial Studies, 3:593-624.

[30] Geweke, J., 1994, "Variable Selection and Model Comparison in Regression," Bayesian Statistics 5, Oxford University Press, 609-620.

[31] Granger, C.W. and O. Morganstern, 1970, Predictability of Stock Market Prices, Heath Lexington Books, Lexington, MA.

[32] Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moment Estimators," Econometrica, 50:1029-1054.

[33] Harris, L., 1986, "Cross-Security Tests of the Mixture of Distributions Hypothesis," Journal of Financial and Quantitative Analysis, 21:39-46.

[34] Harvard Electricity Policy Group, 1998, "Reshaping the Electricity Industry: A Public Policy Debate," Center for Business and Government, Harvard University.

[35] Jacquier, E. and N. Polson and P. Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models," Journal of Business and Economic Statistics, 12: 371-388.

[36] Kandel, E. and N. Pearson, 1995, "Differential Interpretation of Public Signals and Trade in Speculative Markets," Journal of Political Economy, 103:831-872.

[37] Kandel, S. and R. Stambaugh, 1996, "On the Predictability of Stock Returns: An Asset Allocation Perspective," Journal of Finance, 51:385-424.

[38] Karpoff, J., 1987, "The Relation between Price Changes and Trading Volume: A Survey," Journal of Financial and Quantitative Analysis, 22:109-126.

[39] Keim, D., and R. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," Journal of Financial Economics, 17: 357-390.

[40] Klein, R., and V. Bawa, 1976, "The Effect of Estimation Risk on Optimal Portfolio Choices," Journal of Financial Economics, 3:215-231.

[41] Lamoureux, C. and W. Lastrapes, 1990, "Heteroskedasticity in Stock Return Data: Volume versus GARCH effects," Journal of Finance, 45:221-229.

[42] Lamoureux, C. and W. Lastrapes, 1994, "Endogenous Trading Volume and Momentum in Stock-Return Volatility," Journal of Business and Economic Statistics, 12:253-260.

[43] Lamoureux, C., and G. Zhou, 1996, "Temporary Components of Stock Returns: What Do the Data Tell Us," Review of Financial Studies, 9:1033-1059.

[44] Liesenfeld, R., 1998, "Dynamic Bivariate Mixture Models: Modeling the Behavior of Prices and Trading Volume," Journal of Business and Economic Statistics, 16:101-109.

[45] Lindley, D., 1957, "A Statistical Paradox," Biometrica, 44:187-192.

[46] Lo, A., A. MacKinlay, 1997, "Maximizing Predictability in the Stock and Bond Markets," Macroeconomic Dynamics, 1:131-170.

[47] Lucas, R. Jr., 1978, "Asset Prices in an Exchange Economy," Econometrica, 46:1429-1445.

[48] Mandelbrot, B., 1963, "The Variation of Certain Speculative Prices," The Journal of Business, 36:394-419.

[49] McQueen, G., and S. Thorley, 1991, "Are Stock Returns Predictable? A Test Using Markov Chains," Journal of Finance, 46:239-263.

[50] Nelson, D., 1990, "ARCH models as Diffusion Approximations," Journal of Econometrics, 45:7-38.

[51] Newey, W., and K. West, 1987, "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, 55:703-708.

[52] Pástor, L., 2000, "Portfolio Selection and Asset Pricing Models," Journal of Finance, 55:179-223.

[53] Pástor, L., and R. Stambaugh, 2000, "Comparing Asset Pricing Models," Journal of Financial Economics, 56:335-381.

[54] Poirier, D., 1988, "Frequentist and Subjectivist Perspectives on the Problems of Model Building in Economics," Journal of Economic Perspectives, 2:121-144

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

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

[57] Richardson, M., and T. Smith, 1994, "A Direct Test of the Mixture of Distributions Hypothesis: Measuring the Daily Flow of Information," Journal of Financial and Quantitative Analysis, 29:101-116.

[58] Roll, R., 1988, "R^2," Journal of Finance, 43: 541-566.

[59] Smith, R., 1999, "Many PG&E Rivals are Expected to Bid for Utility's Huge Hydropower System," The Wall Street Journal.

[60] Tauchen, G., and M. Pitts, 1983, "The Price Variability-Volume Relationship on Speculative Markets," Econometrica, 51:485-505.

[61] Timmerman, A., 2001, "Structural Breaks, Incomplete Information, and Stock Prices," Journal of Economic and Business Statistics, 19:299-315.

[62] Veronesi, P., 1999, "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model," Review of Financial Studies, 5:975-1007.

[63] Wagman, D., September 23, 1999, "The New Market Players: Who's Buying Where," Resource Data International, Financial Times Energy.

[64] Watanabe, T., 2000, "Bayesian Analysis of Dynamic Bivariate Mixture Models: Can They Explain the Behaviors of Returns and Trading Volume?" Journal of Business and Economic Statistics, 18:199-210.

[65] Wessel, D., January 8, 1999, "Surge of Interest in Electricity Mirrors Today's Internet Hype," The Wall Street Journal.

[66] Xia, Y., 2001, "Learning about Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation," Journal of Finance, 56:205-246.