Stock Returns and Dividend Yields: Some New Evidence*

David A. Chapman† Timothy Simin‡ Hong Yan§
University of Texas Penn State University of Texas

January 29, 2003

Abstract

We introduce the grid-W bootstrap confidence set, a generalization of the grid-t bootstrap confidence interval from Hansen (1999) to evaluate the joint significance of the market dividend yield as a predictor of the subsequent month’s market return and persistence parameter in the dividend yield autoregression. We find that the asymptotically correct confidence set, constructed using monthly data from 1927 to 2001, is not closed. This implies that the full sample is uninformative about the predictability regression parameters. When the predictability vector autoregression is estimated using post-1947 and post-1963 subsamples, the grid-W confidence set is closed (or nearly closed), but it is always larger than the conventional asymptotic confidence set. Using these asymptotically correct confidence sets in all periods, we fail to reject the null of no predictability at a nominal 10 percent level using post-1947 data, where the conventional confidence set does reject the null of no predictability at this same nominal size.

---

*The most recent version of this paper is available as a pdf file at http://www.bus.utexas.edu/Faculty/David.Chapman/csy.html.
†Finance Department, McCombs School of Business, University of Texas at Austin, Austin, Texas 78712-1179. chapman@eco.utexas.edu.
‡Department of Finance, Smeal School of Business, The Pennsylvania State University, 609 Business Administration Building, University Park, PA 16802, e-mail: tsimin@psu.edu
§Finance Department, McCombs School of Business, University of Texas at Austin, Austin, Texas 78712-1179. Hong.Yan@bus.utexas.edu.
Abstract

We introduce the grid-W bootstrap confidence set, a generalization of the grid-t bootstrap confidence interval from Hansen (1999) to evaluate the joint significance of the market dividend yield as a predictor of the subsequent month’s market return and persistence parameter in the dividend yield autoregression. We find that the asymptotically correct confidence set, constructed using monthly data from 1927 to 2001, is not closed. This implies that the full sample is uninformative about the predictability regression parameters. When the predictability vector autoregression is estimated using post-1947 and post-1963 subsamples, the grid-W confidence set is closed (or nearly closed), but it is always larger than the conventional asymptotic confidence set. Using these asymptotically correct confidence sets in all periods, we fail to reject the null of no predictability at a nominal 10 percent level using post-1947 data, where the conventional confidence set does reject the null of no predictability at this same nominal size.
1 Introduction

The issue of asset return predictability is fundamental to both asset pricing theory and practical portfolio management. Theoretically, the precise form of return predictability provides potentially important information about the dynamics of the trade-off between risk and return. Practically, even though the connection between predictability and equilibrium models of asset returns is an open theoretical issue, investors must still attempt to quantify the potential benefits and costs to engaging in portfolio management strategies that use the information in measured predictability.

The existing literature on measuring return predictability is large. It includes examinations of the effect of pre-sample selection biases, out-of-sample performance of estimates, and the economic significance of predictability, from both classical and Bayesian perspectives. In this paper, we concentrate on the problem of constructing consistent tests using the market dividend yield as the predictor. This particular case is important, in part, because it is one of the most common return predictors examined in the literature but also because the econometric issues involved in tests based on dividend yields generalize to other commonly used predictors.\footnote{See Campbell, Lo, and MacKinlay (1997) and Cochrane (2001) for textbook discussions of dividend yield prediction regressions. Important papers in the return predictability literature include Fama and French (1988), Hodrick (1992), Goetzmann and Jorion (1993), Kim and Nelson (1993). Recent reexaminations of the evidence of predictability include Ang and Bekaert (2001), Torous, Valkanov, and Yan (2001), Campbell and Yogo (2002), Lanne (2002), and Lewellen (2002).}

Inference in predictability regressions based on dividend yields is a challenging econometric problem because the data generating process for monthly dividend yields, in U.S. data over the period from 1927 to 2001 (and subperiods) is highly persistent, and it may well contain an autoregressive unit root. There is a large econometric literature on the estimation problems associated with unit root (or near unit root) dynamics, and these issues turn out to have important implications for inference.
about return predictability.\footnote{See, for example, Chapter 17 in Hamilton (1994) or the survey in Stock (1994).}

Our approach extends results from Hansen (1999) who shows how to construct a consistent confidence interval for an autoregressive unit root, even when the true underlying data generating process is of the local-to-unity form.\footnote{For example, the generic first-order autoregression
\[ x_{t+1} = \mu + \rho x_t + \varepsilon_{t+1} \]
is recast in the local-to-unity framework by the reparameterization
\[ \rho = 1 + c/T \]
where $c < 0$ is fixed as $T \uparrow \infty$.} The innovation in Hansen (1999) is to allow the quantile of a $t$-statistic on the autoregressive unit root to be a function of the autoregressive parameter. In a standard setting, this distribution is independent of the true parameter, and tests and confidence intervals are constructed using the finite-sample $t$-distribution or the asymptotic standard normal. However, when the autoregression is local-to-unity, the $t$-statistic is no longer pivotal. The quantile, as a function of the autoregressive parameter, is constructed at a grid of parameter values using a bootstrap simulation procedure and then the quantile is smoothed over neighboring parameter values using kernel regression. Hansen (1999) demonstrates, in the case of a univariate autoregression, that the intersection of the standard $t$-statistic, as a function of the autoregressive parameter, and the bootstrapped quantile forms a confidence interval that has first-order correct asymptotic coverage \textit{even if the autoregression is local-to-unity}.

We define a straightforward generalization of the grid-$t$ bootstrap confidence interval. Since this confidence set uses the quantiles of the Wald statistic that tests joint hypotheses on the slope coefficient in the predictability regression and the autoregressive parameter in the dividend yield autoregression, we call it the grid-W...
confidence set. By construction, this confidence set is robust to some time-variation in volatility. As a result, it bypasses the problems associated with either ignoring the volatility effects in both excess returns and dividend yields or explicitly attempting to correct for volatility effects in the OLS estimates or their Bayesian counterparts using a parametric model. The grid-W confidence set has the appropriate asymptotic coverage under broader conditions than a confidence set constructed from conventional asymptotic theory, and if the Monte Carlo results in Hansen (1999) extend to the bivariate case, it has excellent finite-sample properties (even in very small samples) for constant variance normal errors. While these are the same conditions used to justify the Bayesian estimator in Stambaugh (1999), an important direction for future research is to extend the results in Appendix B, below, to more thoroughly examine the finite-sample properties of the grid-W set for the types of data-generating processes that more accurately describe financial time series.

When we construct the grid-W confidence set using monthly U.S. data from January 1927 to December 2001, we find that the confidence set is not closed. In particular, in the region of the parameter space that is near an explosive root in the autoregressive parameter and no (or negative) predictability of dividend yields for excess returns, the quantile function of the conventional Wald test depends dramatically on the underlying parameters. This is in stark contrast to conventional asymptotic theory in which the quantile is constant – for all parameters – at the chosen level of the $\chi^2_2$ distribution. This open confidence set suggests, following the analysis in Dufour (1997), that the predictability parameters are locally almost unidentified in the region of the parameter space near the ‘unit root and no predictability’ point, and the data simply do not provide meaningful information about return predictability over the full sample.

This finding is, however, different from to the results that we obtain using data
from January 1947 on or data from 1963 on. Over the post-World War II sample, we find that the grid-W confidence set is larger than the conventional asymptotic confidence set. Although both confidence sets are asymptotically correct when dividend yields are stationary, only the grid-W set has the correct size when dividend yields are close to explosive. This means that the conventional test statistic rejects the null of no predictability too often when that null is correct. These differences between the conventional and the grid-W set can be important. For example, using the January 1963 to December 2001 data set and the conventional confidence set, the null of no predictability is rejected for all levels of dividend yield persistence. However, the grid-W set over this interval fails to reject no predictability if the dividend yields are very close to having a unit root.

In comparing the two subsamples, we find that the confidence sets (both the conventional and the grid-W) are smaller for the longer data set. This is precisely what we would expect to find with data drawn from a stationary Markov process, but it is inconsistent with our full sample results. This suggests that data from the Great Depression, which is included in the full sample, may be drawn from a distinct regime. In any case, it raises serious doubts about what can be learned from the full sample of data under relatively weak restrictions on the data generating process.

The remainder of the paper is organized as follows: Section 2 describes the basic estimation problem, and it contains a review of the relevant recent literature. The motivation and construction of the grid-W bootstrap confidence set is in Section 3, and Section 4 applies the estimator to the U.S. data. The conclusions and implications for future work are contained in Section 5. Appendix A provides some discussion of the technical details of the grid-W confidence set, and Appendix B describes a limited Monte Carlo study of the finite-sample properties of the grid-W confidence set for the predictability problem.
2 The Estimation Problem

The restricted vector autoregression (VAR) used in the return predictability literature is typically motivated by an appeal to an asset price as the discounted value of future dividends.\(^4\) The “approximate” present value identity for the log-dividend yield is:

\[
d_t - p_t \approx k + E_t \left[ \sum_{j=1}^{\infty} \delta^{j-1} (\Delta d_{t+j} - r_{t+j}) \right],
\]

(1)

where \(r_t\) denotes the real, continuously compounded return on the risky asset, \(p_t\) is the real ex-dividend asset price, and \(d_t\) is the real dividend, and prices and dividends are expressed in log-levels and normalized to have unconditional means of zero. \(k\) and \(\delta\) are constants, \(\Delta\) denotes the first difference operator, and \(\lim_{j \to \infty} \delta^j (p_{t+j} - d_{t+j}) = 0\). This equation implies that predictable variation in dividend yields produces predictable variation in either dividend growth, asset returns, or both.

A common choice in the asset return predictability literature is to focus on continuously compounded excess returns and log-dividend yields, expressed as a restricted bivariate vector autoregression:

\[
\begin{align*}
\text{ex}_{t+1} &= \alpha + \beta d_{t+1} + u_{t+1} \\
d_{t+1} &= \omega + \rho d_{t+1} + v_{t+1}
\end{align*}
\]

(2)

where \(\text{ex}_{t+1}\) and \(d_{t+1}\) denote excess returns and log dividend yields, respectively. For convenience, \(\beta\) will be referred to as the “predictability coefficient.” The disturbances \((u_{t+1}, v_{t+1})'\) are assumed to be \(iid\) draws from some fixed bivariate distribution with mean zero and covariance matrix

\[
\Sigma \equiv \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}.
\]

(3)

\(^4\)See Campbell, Lo, and MacKinlay (1997) or Cochrane (2001) for textbook discussions of this derivation.
The OLS estimates $\hat{\beta}_T$ and $\hat{\rho}_T$ (based on a sample of size $T$) of the slope parameters in (2) are consistent, but they are biased in finite samples, because both disturbances are correlated with future values of the log dividend yield.\(^5\) Proposition 1 in Stambaugh (1999) proves that the finite-sample distribution of $\hat{\beta}_T - \beta$ depends on the true values of $\rho$ and $\Sigma$, but that it does not depend on $\beta$ (or the unconditional means). This result is very important for two reasons: (i) it allows for a calibration, based on U.S. data, of the magnitude of the finite-sample bias, and this bias appears to be large (about 1/3 of the OLS estimate using 70 years of monthly data). (ii) Under the assumption that $(u, v)'$ is independent and normally distributed, it also implies that the expected bias in the excess return slope coefficient estimate is

$$E \left[ \hat{\beta}_T - \beta \right] = \frac{\sigma_{uv}}{\sigma_v^2} E [\hat{\rho}_T - \rho]$$  \hspace{1cm} (4)

or

$$E \left[ \hat{\beta}_T - \beta \right] = -\frac{\sigma_{uv}}{\sigma_v^2} \left( \frac{1 + 3\rho}{T} \right) + O \left( T^{-2} \right).$$  \hspace{1cm} (5)

For example, using U.S. data from 1927 to 1996, Stambaugh estimates the regression coefficient of excess return residuals on dividend yield residuals as approximately $-15$, which combined with the downward bias in ordinary least squares (OLS) estimates of the dividend yield autoregressive coefficient, results in a large positive finite-sample bias in the estimate of the predictive ability of lagged dividend yields. He goes on to show that a Bayesian analysis of the predictability regression problem differs substantially from a classical analysis, both in the interpretation of the point estimate and in the economic significance of return predictability.

Torous, Valkanov, and Yan (2001) note that the analysis in Stambaugh (1999) assumes that dividend yields are generated by a stationary process; i.e., log dividend

\(^5\)See Stambaugh (1999) (especially page 379) for an intuitive description of the source of the finite sample bias, and see Hamilton (1994) for the asymptotic arguments that demonstrate the consistency of the slope estimators.
yield is integrated of order zero, frequently denoted $I(0)$. They further note that the correct form of the predictive regression depends critically on whether or not the predictor variable (dividend yields) is $I(0)$ or contains a unit root (is integrated of order one, denoted $I(1)$). Torous et al construct asymptotically valid confidence intervals for the predictability coefficient. They use the Bonferroni inequality and a confidence interval for the largest autoregressive root in the predictor variable constructed from an augmented Dickey-Fuller $t$-statistic, following the approach in Cavanagh, Elliott, and Stock (1995) (an extension of Stock (1991)). This approach explicitly allows for local-to-unity asymptotics in the predictor. Torous et al (2001) find: (i) the statistical evidence of return predictability is indeed sensitive to the treatment of the near unit root nature of common predictors; (ii) there is little evidence for long-horizon return predictability – once the near unit-root nature of common predictors is explicitly considered, and (iii) there is reliable evidence of short-horizon (one year or less) predictability using term spreads, the short rate, and dividend yields.

Campbell and Yogo (2002) also argue that explicitly incorporating information about the persistence of the predictor variable can have an important impact on the statistical evidence for or against short-horizon return predictability. They note that test statistics based on local-to-unity asymptotics lead to correct (asymptotic) inference under a broader set of conditions than standard asymptotic theory, but these tests are computationally more demanding than standard tests – sometimes dramatically so. They suggest a two-step testing procedure. First, apply an efficient unit root test to the predictor variable. Given an $\alpha$-level confidence interval constructed from this test, either apply standard or local-to-unity based estimators to construct Bonferroni-type confidence intervals for the predictability coefficient.\(^6\)

\(^6\)Torous et al (2001) – citing evidence from Cavanagh, Elliott, and Stock (1995) – suggest that this type of a two-step approach is likely to suffer substantial size distortion.
Using a test statistic for the predictability parameter that explicitly incorporates information about the autoregressive parameter, Campbell and Yogo argue that there is evidence of significant return predictability using dividend yields at the annual frequency from 1927 to 2001. At the quarterly horizon, the evidence is weaker, but it is still significant for levels of the autoregressive parameter below 0.988.\footnote{Lewellen (2002) uses the same test statistic as Campbell and Yogo (2002) and the assumption of $\rho = 1$ to conclude that monthly returns from 1946 to 2000 (including subperiods) are predictable by dividend yields.}

Ferson, Sarkissian, and Simin (2002) demonstrate that, when both expected returns and the predictor variable are highly persistent, then the evidence for predictability is quite likely to be “spurious”, in the formal econometric sense of Phillips (1986). They go on to demonstrate that the extent of this problem is increasing in both the autocorrelation in the predictor and the relative importance of time variation in the conditional mean. Finally, Ferson et al (2002) show that even a modest amount of data-mining (see Lo and MacKinlay, 1990) can substantially amplify the likelihood of spurious inference, calling into question the validity of specific instruments identified in the literature, such as the term spread, book-to-market ratio, and dividend yield.

There are two contributions of our paper relative to the existing literature. First, the test statistic developed below explicitly constructs a joint confidence interval for $(\beta, \rho)$ that accounts for sample evidence about the correlation of the two estimated parameters, and it does not require direct estimation of the local-to-unity parameter nor does it treat $\rho$ merely as a nuisance parameter. Second, in addition to having first-order optimal asymptotic properties, the limited Monte Carlo evidence in Appendix B and the Monte Carlo analysis of the related univariate test in Hansen (1999) suggests that the confidence set based on the test described in the next section is much better than the coverage provided by a conventional test. However, more work
is needed to assess the finite-sample size (and relative power) of the grid-$W$ approach compared to the results in both Torous et al (2001) and Campbell and Yogo (2002). Finally, although we will focus on the use of dividend yields as the predictor variable, our conclusions generalize readily to a short-horizon study of any highly persistent predictor.

3 The Grid-$W$ Bootstrap Confidence Set for $(\beta, \rho)$

3.1 General Considerations

In order to motivate the construction of the grid-$W$ confidence set, we begin with a simpler problem: constructing a confidence interval for a slope parameter in a simple linear regression. If $\beta$ is the slope, then the standard pivotal statistic for testing hypotheses about the slope is

$$t(\beta) = \frac{\hat{\beta}_T - \beta}{s(\hat{\beta}_T)},$$

where $\hat{\beta}_T$ is an estimate of the slope and $s(\hat{\beta}_T)$ is the standard error of the estimate. Under the assumption that the errors in the regression are normally distributed, then $t(\beta)$ has an exact finite sample distribution that is Student-$t$ with $T - 2$ degrees of freedom. If the errors are only assumed to satisfy certain moment existence conditions, then the asymptotic distribution of $t(\beta)$ will be $\mathcal{N}(0, 1)$.

In either case, given a distribution function for the test statistic, the $\mu$-level confidence interval for $\beta$ is the set

$$C_c = \left\{ \beta \in \mathbb{R} : q(\theta_1) \leq t(\hat{\beta}_T) \leq q(\theta_2) \right\},$$

where $C_c$ is the conventional confidence interval, $q(\theta_1)$ is the $\theta_1$-quantile of the distribution of $t(\beta)$, $q(\theta_2)$ is the $\theta_2$-quantile of the distribution of $t(\beta)$, and

$$\theta_1 = \frac{1 - \mu}{2}; \quad \theta_2 = 1 - \frac{1 - \mu}{2}.$$
So, for example, the 90 percent confidence interval (\(\mu = 0.9\)) for the slope coefficient in a standard regression context is \(\theta_1 = 0.05\) and \(\theta_2 = 0.95\) and

\[
C_c = \left\{ \beta \in \mathbb{R} : -1.645 \leq t \left( \hat{\beta}_T \right) \leq 1.645 \right\}.
\]

It is important to note here that (6) is a constant function of the true underlying slope parameter. The issue here is whether or not this confidence interval is an accurate description of the finite-sample distribution of \(t(\beta)\), and (as a result) whether or not it provides accurate inference regarding the unknown slope parameter.

Hansen (1999) notes that the quantiles of the distribution of \(t(\rho)\) in a first-order autoregression are not constant with respect to the unknown parameter, since it has a nonstandard asymptotic distribution that depends on the local-to-unity parameter. The problem is that \(c\), the local-to-unity parameter, cannot be estimated consistently using OLS. He proposes the grid-\(t\) bootstrap confidence interval as an alternative to \(C_c\). This set is defined as

\[
C_{gt} = \{ \rho \in \mathbb{R} : q_{T}^{*}(\theta_1 | \rho) \leq t(\rho) \leq q_{T}^{*}(\theta_2 | \rho) \},
\]

where \(q_{T}^{*}(\bullet | \rho)\) is the empirical quantile function evaluated under the specific value of \(\rho\) for the autoregressive parameter. This function is constructed by simulation and smoothing, and it allows for the quantile function to depend explicitly on the value of the unknown parameter.\(^8\) Note that (7) is different from a conventional bootstrap confidence interval because the conventional bootstrap interval evaluates the \(t\)-statistics and its quantile only at the point estimate \(\hat{\rho}_T\); i.e., the conventional bootstrap retains the assumption that the distribution of the test statistic is constant across the parameter space.

\(^8\)The exact construction of our generalization of the grid-\(t\) confidence interval will be described below.
Hansen (1999) goes on to make three important points regarding the confidence interval in (7): (i) The grid-\( t \) confidence interval has first-order asymptotic consistency even when the process being estimated contains an explosive root;\(^9\) (ii) The conventional confidence interval and a conventional bootstrap confidence interval only attains first-order consistency when the underlying process is stationary; and (iii) A Monte Carlo study of the performance of the grid-\( t \) interval under the assumption of normal shocks with constant variance suggests that it is quite accurate in even relatively small (\( T = 60 \)) samples. As Appendix A notes, results from Elliott and Stock (1994) show that \( t (\beta) \) is also a function of the local-to-unity parameter, and it also has a nonstandard asymptotic distribution.

### 3.2 Constructing the Confidence Set

The extension of the grid-\( t \) bootstrap to the case of the two slope coefficients in the VAR in equation (2) is straightforward. Instead of the \( t \)-statistic on the single slope coefficient, joint confidence sets can be constructed using the Wald statistic designed to test the joint null hypothesis \( H_0 : \beta = \beta_0 \) and \( \rho = \rho_0 \). This statistic is defined as the quadratic form

\[
W_T (\gamma) = (C\hat{\gamma} - r_0)' \left[ CS_T (\hat{\gamma}_T) C' \right]^{-1} (C\hat{\gamma} - r_0),
\]

where \( \gamma \equiv (\alpha, \beta, \omega, \rho)' \), \( \hat{\gamma}_T \) is the OLS estimate of \( \gamma \), \( S_T (\hat{\gamma}_T) \) is a consistent estimator of the asymptotic covariance matrix of the OLS parameters,

\[
C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad r_0 = \begin{bmatrix} \beta_0 \\ \rho_0 \end{bmatrix}.
\]

The grid-\( W \) confidence set is constructed as follows:

\(^9\) A \( \mu \)-confidence interval \( C (\rho) \) is first order consistent if

\[
\Pr (\rho \in C (\rho)) \to \mu
\]

as \( T \uparrow \infty \).
Step #1: Construct $\hat{\gamma}_T$ and $\hat{S}_T(\hat{\gamma}_T)$. The parameter covariance matrix will have the standard form of

$$\hat{S}_T(\hat{\gamma}_T) = T \cdot (X_T'X_T)^{-1} \hat{\Sigma}_T (X_T'X_T)^{-1},$$

where $X_T = (1 \;dy_{t-1})$ is the $(T-1) \times 2$ data matrix and $\hat{\Sigma}_T$ is the estimate of the residual covariance matrix constructed from the fitted residuals $\hat{u}$ and $\hat{v}$. Given the dynamics of the fitted residuals to equation (2), $\hat{\Sigma}_T$ is constructed using the Newey-West heteroskedasticity and autocorrelation consistent covariance matrix estimator. Alternatively, though, the covariance matrix could be estimated under the null of iid shocks.

Step #2: Construct $W_T(\gamma)$ at a grid of values for $\beta_0 \in [\underline{\beta}, \bar{\beta}]$ and $\rho_0 \in [\underline{\rho}, \bar{\rho}]$. In the empirical analysis, below, $\underline{\beta} = \bar{\beta}_T - 4\sigma(\bar{\beta}_T)$ and $\bar{\beta} = \bar{\beta}_T + 4\sigma(\bar{\beta}_T)$, where $\sigma(\bar{\beta}_T)$ is the sample standard deviation of $\bar{\beta}_T$. The grid for $\rho$ is defined analogously.

Step #3: Construct estimates of the quantile function $\hat{q}_T(\mu \mid \beta_0, \rho_0)$ for each $(\beta_0, \rho_0) \in \Gamma_0 \equiv [\underline{\beta}, \bar{\beta}] \times [\underline{\rho}, \bar{\rho}]$ as follows:

Step #3a: For each $(\beta_0, \rho_0) \in \Gamma_0$, construct a set of residuals $u(\beta_0, \rho_0)$ and $v(\beta_0, \rho_0)$ using the actual data on excess returns and dividend yields. These pseudo-residuals make the observed data consistent with the particular parameter pair $(\beta_0, \rho_0)$.

Step #3b: Construct $B$ bootstrap samples from the pseudo-residuals using an overlapping block bootstrap algorithm. For each sample, construct dividend yield and excess return series from $(\beta_0, \rho_0)$ and the bootstrapped residuals.
Step #3c: For each bootstrap sample of excess returns and dividend yields, construct the OLS estimates and the Wald statistic and store the results.

Step #3d: Compute the empirical quantile, $\hat{q}_T (\mu \mid \beta_0, \rho_0)$, of the Wald statistic at over the $B$ bootstrap samples and store for future use.

Step #3e: Repeat Steps #3a through #3d for each point in $\Gamma_0$.

Step #4: Smooth the estimated quantile $\hat{q}_T (\mu \mid \beta_0, \rho_0)$ over neighboring points in $\Gamma_0$ using a generally additive linear model fit with univariate kernel regression estimators.\textsuperscript{10}

Step #5: Find the projection onto $\Gamma_0$ of the intersection of the $W_T (\beta_0, \rho_0)$ and the smoothed quantile function. This set is the bivariate analog of Hansen’s grid-$t$ confidence interval:

\begin{equation}
C_{gW} (\mu) = \{(\beta, \rho) \in \mathbb{R}^2 : W_T (\beta, \rho) \leq \hat{q}_T (\mu \mid \beta, \rho)\}.
\end{equation}

The grid-$W$ confidence set achieves first-order consistency when dividend yields are both covariance stationary and local-to-unity by a generalization of the results in Hansen (1999).\textsuperscript{11} An important direction for future research is to verify through more extensive Monte Carlo experiments than those contained in Appendix B that the performance of the grid-$W$ bootstrap is informative about the joint finite-sample properties of estimates of $\beta$ and $\rho$.

In applying bootstrap methods to financial time series, parametric models – even those that extend to modelling the volatility structure of excess returns and dividend yields – fail to capture all of the dependence in the data. As a result, the fitted residuals to models such as equation (2) will not satisfy an independence assumption. The

\textsuperscript{10}This is a bivariate generalization of the smoothing estimator used in Hansen (1999).

\textsuperscript{11}See the discussion in Appendix A for some details.
standard approach to constructing a bootstrap sample from a sample of dependent data is to use a variant of the “block bootstrap” introduced in Carlstein (1986) or Politis and Romano (1994). Hall, Horowitz, and Jing (1995) discuss the issue of the optimal block size, both asymptotically and in finite sample, and Härdle, Horowitz, and Kreiss (2001) examine issues in the actual performance of time series bootstrap methods.

Despite these challenges, there are important benefits, in the current setting, to examining the full joint confidence set rather than simply correcting the bias in the point estimate of either $\hat{\rho}_T$ or $\hat{\beta}_T$. In particular, standard excess return and dividend yield data sets are clearly inconsistent with the assumption of constant variance that underlies the bias correction in (4) and in its Bayesian counterparts. As a result, the twin issues of how to estimate $\sigma_{uv}/\sigma^2_v$ and the covariance of this estimate with the bias estimate are nontrivial considerations. By implicitly estimating the joint distribution of this covariance estimator with the estimated bias in the autoregressive parameter – while allowing (at least in theory) for time-variation in residual variances – we resolve these issues in a single unified framework.

4 Results Using U.S. Data

4.1 OLS Estimates

The standard data set that we use to motivate the analysis below consists of monthly observations on continuously compounded returns on the CRSP value-weighted market in excess of the one-month Treasury bill yield and the natural logarithm of the annual dividend yield on the value-weighted market. This market dividend yield series is constructed from the with-dividend and ex-dividend returns to the value-weighted market, as described in Fama and French (1988). The full sample consists of $T = 900$
observations from January 1927 to December 2001.

Plots of the raw data are shown in Figure 1, and the accompanying summary statistics are contained in Table 1. The average monthly excess return is slightly in excess of 0.5 percent per month with a standard deviation of 5.52 percent per month. Autocorrelations for the market excess return are weak at the monthly observation frequency, but Figure 1 suggests the standard finding in the literature that excess return volatility is highly persistent. The average annual dividend yield on the market, over the full sample, is 3.91 percent per year. Dividend yields in the last 15 years of the sample have been at the lowest level of the entire period. Dividend yields are highly persistent. Autocorrelations of this series are high initially (starting at 0.95), and they decline slowly with further lags (the sixth-order autocorrelation coefficient is still 0.84).

OLS estimates of the parameters in equation (2) are shown, for different sample periods, in Table 2. Over the full sample period, lagged dividend yields explain roughly 1/4 percent of the variation in excess returns, and the slope coefficient is not significantly different from zero. Dividend yields are highly autocorrelated, possessing a root very close to unity. Results for the sub-periods from January 1947 to December 2001 and January 1963 to December 2001 are shown in Panels B and C of Table 2. Excess returns are positively (and statistically significantly) related to lagged dividend yields only in the January 1947 to December 2001 subperiod, and even here, dividend yields explain less than 1 percent of the variation in excess returns.

4.2 Grid-W Estimates

The estimate of the 10 percent grid-W confidence set for $\beta$ and $\rho$, based on the full sample from January 1927 to December 2001, is shown in Figure 2, along with the conventional asymptotic confidence set (an ellipse) based on the $\chi^2$ distribution for.
The interval is constructed as described in the previous section using $B = 999$ bootstraps on a parameter grid of $60 \times 60$.\footnote{The results were virtually unchanged for computations based on a $100 \times 100$ grid and $B = 1999$. The major computational cost associated with constructing the grid-$W$ confidence set is the grid size, since $\tilde{q}_T(\mu | \beta, \rho)$ must be evaluated at each point in the grid. An increase from $60 \times 60$ to $100 \times 100$ nearly triples the computation time for any given bootstrap number and sample size.} The conventional asymptotic confidence set lies entirely inside the grid-$W$ set, which suggests that a test based on the conventional ellipse will reject a true null hypothesis about $\beta$ and $\rho$ more frequently than the asymptotically correct set.

The more striking result about the grid-$W$ set, however, is that it is not closed over $\Gamma_0$. According to the analysis in Dufour (1997), this implies that over the full sample period the combination of the predictability coefficient and the autoregressive parameter are locally almost unidentified in the region of the parameter space that contains $\Gamma_0$. In this case, the conventional closed confidence set has a true level of zero. As Dufour (1997) notes:

In other words, for the family of tests $[W_T(\gamma)]$ we can always find a hypothesis $[H_0 : \gamma = \gamma_0]$ such that the level of the corresponding test will exceed any nominal level. As a result, the statistics cannot be pivotal functions ... i.e., the distribution of $[W_T(\gamma)]$ depends on $[\gamma]$. [Dufour (1997), pages 1373-1374.

This point is confirmed by examining the fitted $\tilde{q}_T(0.9 | \beta_0, \rho_0)$ function in Figure 3. It demonstrates that the 90% quantile of the Wald statistic depends on the values of $(\beta, \rho)$, and hence the statistic is not pivotal. The grid-$W$ confidence set is a consistent estimate of the true confidence set precisely because it does allow for this dependence.

Since $\Gamma_0$ contains the economically interesting null of ‘no predictability’ and a ‘unit root’, i.e., $\beta = 0$ and $\rho = 1$, in dividend yields, this is problematic for interpreting the results from a predictability VAR. In fact, following the logic in Dufour (1997), the
classical interpretation of this open confidence interval is that the full sample data is simply uninformative about the slope parameters of the standard predictability VAR.

The grid-\(W\) confidence sets computed from the post-1947 and post-1963 data sets are shown together in Figure 4. Over the longer subperiod, both the conventional and the grid-\(W\) sets are closed. Over the post-1963 sample, the conventional confidence set is still an ellipse, but the grid-\(W\) set is open (again) in the region of the parameter space where \(\rho \geq 1.0\). There are a number of facts about the relationship among these sets that bears discussion.

First, the confidence sets over the latter subperiods are consistent with what we would expect to see with more data drawn from a stationary Markov process. In particular, the sets – both conventional and grid-\(W\) – constructed with the 600 monthly observations post-1947 are smaller than their post-1963 counterparts, constructed using only 468 observations. In other words, inference about the slope parameters in the VAR should be more precise in the post-1947 data. There is some evidence of changes in the mean of the slope parameters, since the confidence sets are not strict subsets of each other, but this effect is not strong.

Second, the conventional asymptotic confidence interval is strictly contained in its larger grid-\(W\) counterpart. Since both sets are consistent when the true data generating process for dividend yields is stationary, the observed differences in these confidence sets suggests that the dividend yield process is very close to a unit root. This relationship between the sets also implies that the conventional test tends to reject a true null hypothesis too often; i.e., the actual size (in any sample) of the conventional asymptotic 10 percent confidence set is larger than 10 percent. The differences in the inference drawn from the conventional versus the grid-\(W\) interval can be quite large. In particular, for the post-1947 data set, the conventional confidence interval rejects the null of no predictability \((\beta = 0)\) for any level of \(\rho\). However,
the asymptotically correct grid-W set shows that for values of \( \rho \) very close to 1 it is impossible to reject that \( \beta = 0 \) at the 10 percent level.

The relationship between the intervals in Figures 2 and 4 merits some discussion. As noted above, in moving from post-1947 to post-1963, we clearly observe the effects of reducing the sample size on our ability to infer the combinations of \( \beta \) and \( \rho \) that are consistent with the data. However, as we move back to the full 900 monthly observations from January 1927 to December 2001, our confidence sets do not become smaller, rather this additional data makes inference about asset return predictability more difficult. This is consistent with the long-standing observation in the empirical literature that asset returns in the 1930s differed fundamentally from returns in other periods.\(^{13}\) This is a serious issue, since the grid-W confidence set assumes that we are observing draws from a Markov process that satisfies some stationarity conditions (although one that is possibly local-to-unity). If a regime shift process (if only in volatility) is a more accurate characterization of the data, then including data generated across regimes may account for the performance of the grid-W confidence set on the full sample versus the later subsamples.

Finally, all of these results are asymptotic in nature, and it is possible that some (or all) of the differences that we are observing can be attributed to small sample errors. The limited Monte Carlo results in Appendix B, constructed using a data-generating process for returns and dividend yields based on the null of no predictability \((\beta = 0)\), \( \rho = 0.99 \), and a GARCH(1,1) specification with constant variance, confirms that the grid-W set is larger than the conventional asymptotic confidence set.

\(^{13}\)See, for example, Officer (1973) and Schwert (1989).
5 Conclusions

In this paper, we have extended the asymptotic confidence interval constructed in Hansen (1999) to the restricted bivariate vector autoregression that examines the extent to which excess returns can be predicted using lagged dividend yields. This extension is important methodologically because it is robust to near unit root dynamics in dividend yields. The results based on the grid-$W$ bootstrap confidence set are dramatically different than those that hold under the assumption that dividend yields are generated by a stationary process. We find that the long time series of monthly data from 1927 to 2001 does not provide reliable information about the predictive power of dividend yields. Formally, we find that the predictability parameter and the dividend yield autoregressive parameter are locally almost unidentified when the volatile data from the 1930s is included in the sample.

When only data from 1947 to 2001 are used in the analysis, we find that the evidence against the null hypothesis of no predictability is much weaker when inference is conducted with the asymptotically more robust grid-$W$ confidence set. In the 1947 to 2001 and 1963 to 2001 data sets, the grid-$W$ set contains the conventional asymptotic confidence set as a strict subset. This is consistent with the conventional test rejecting the null of no predictability more often. Unlike the confidence sets constructed from the longer time series of data, the relationship between the 1947 to 2001 and 1963 to 2001 confidence sets is precisely what we would expect to see with an increase in the number of observations from a stationary process.
Figure 1: Monthly observations on continuously compounded excess returns and log dividend yields from January 1927 to December 2001 (T = 900 observations). Both series are expressed at an annual rate, not in percent.
Figure 2: The grid-W bootstrap confidence set and the conventional asymptotic confidence set based on the $\chi^2$ distribution for (10). The dashed ellipse is the conventional asymptotic set, and the open set defined by the solid irregular bounds is the grid-W set. The estimates are based on monthly observations from January 1927 to December 2001 ($T = 900$ observations). $\times$ marks the parameter combination $\beta = 0$ and $\rho = 0$. 
Figure 3: The $\hat{q}_T(0.9 \mid \beta, \rho)$ function computed over the grid $\Gamma_0$ using monthly observations from January 1927 to December 2001 ($T = 900$ observations).
Figure 4: The grid-W bootstrap confidence sets and the conventional asymptotic confidence sets based on the $\chi^2$ distribution for (10). The dotted ellipse is the conventional asymptotic set, and the irregular oval object in the solid line is the grid-W set for the January 1947 to December 2001 ($T = 600$ observations) data. The ellipse with $\times$ marks and the dashed-dot irregular boundary are the comparable sets based on the January 1963 to December 2001 ($T = 468$ observations) data. $\times$ marks the parameter combination $\beta = 0$ and $\rho = 0$. 
Table 1: Summary Statistics for Excess Returns and Dividend Yields.

<table>
<thead>
<tr>
<th></th>
<th>Excess Returns</th>
<th>log (Dividend Yields)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.00514</td>
<td>0.03912</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.05522</td>
<td>0.01458</td>
</tr>
<tr>
<td>Autocorrelations:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>−0.015</td>
<td>0.951</td>
</tr>
<tr>
<td>2</td>
<td>−0.104</td>
<td>0.926</td>
</tr>
<tr>
<td>3</td>
<td>0.017</td>
<td>0.908</td>
</tr>
<tr>
<td>4</td>
<td>0.079</td>
<td>0.889</td>
</tr>
<tr>
<td>5</td>
<td>−0.003</td>
<td>0.865</td>
</tr>
<tr>
<td>6</td>
<td>0.023</td>
<td>0.840</td>
</tr>
</tbody>
</table>

Excess returns are the continuously compounded monthly returns to the CRSP value weighted portfolio in excess of the continuously compounded yield on a one month Treasury bill. Dividend yield is the annual dividend yield on the value-weighted market, computed from the monthly with-dividend and ex-dividend portfolio returns, as in Fama and French (1988). The data are from January 1927 to December 2001, for a total of $T = 900$ observations.
Table 2: OLS Regression Statistics.


<table>
<thead>
<tr>
<th>Dependent Var.</th>
<th>log (Div. Yld.)</th>
<th>R-squared (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Return</td>
<td>0.1885</td>
<td>(0.2066)</td>
</tr>
<tr>
<td>log (Div. Yld.)</td>
<td>0.9807</td>
<td>(0.0171)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>log (Div. Yld.)</th>
<th>R-squared (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Return</td>
<td>0.3380</td>
</tr>
<tr>
<td>log (Div. Yld.)</td>
<td>0.9931</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>log (Div. Yld.)</th>
<th>R-squared (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Return</td>
<td>0.2830</td>
</tr>
<tr>
<td>log (Div. Yld.)</td>
<td>0.9911</td>
</tr>
</tbody>
</table>

Excess returns are the continuously compounded monthly returns to the CRSP value weighted portfolio in excess of the continuously compounded yield on a one month Treasury bill. Dividend yield is the annual dividend yield on the value-weighted market, computed from the monthly with-dividend and ex-dividend portfolio returns, as in Fama and French (1988). The data are from January 1927 to December 2001, for a total of \( T = 900 \) observations. Robust standard errors, based on the Newey-West estimator with 6 lags are reported in parentheses.
A Some Technical Details

In this appendix, we discuss the extension of the results in Hansen (1999) to the restricted bivariate vector autoregression in equation (2). None of these results are original to our analysis, and our discussion follows Hansen (1999) quite closely in all regards with the obvious minor extensions.

We begin with the bivariate predictive system in equation (2), restated here (in slightly different notation) for convenience:

\[ y_t = \mu_y + \gamma x_{t-1} + \eta_{1t} \]

\[ x_t = \mu_x + \rho x_{t-1} + \eta_{2t}, \]

where \( \eta_t = (\eta_{1t}, \eta_{2t})' \) is a martingale difference sequence satisfying

\[ E(\eta_t \eta_0' | \eta_{t-1}, \eta_{t-2}, \ldots) = \Sigma. \]

For notational convenience, define \( \delta = corr(\eta_{1t}, \eta_{2t}) \).

The \( t \)-statistics for the OLS estimates of \( \rho \) and \( \gamma \) satisfy the following conditions

\[ t(\rho) \equiv \frac{\hat{\rho} - \rho}{s(\hat{\rho})} \Rightarrow \frac{\int_0^1 W_c(s) dW(s)}{\left( \int_0^1 W_c^2(s) ds \right)^{1/2}} \]

and

\[ t(\gamma) \equiv \frac{\hat{\gamma} - \gamma}{s(\hat{\gamma})} \Rightarrow \delta \left\{ \int_0^1 W_c^\mu(s) dW(s) \right\}^{1/2} + \sqrt{(1 - \delta^2)}z, \]

where \( s(\hat{\rho}) \) and \( s(\hat{\gamma}) \) are the standard errors for the OLS estimates \( \hat{\rho} \) and \( \hat{\gamma} \), \( c \) is the local-to-unity parameter (as in \( \rho = 1 + c/T \), where \( c < 0 \)), and \( \Rightarrow \) denotes weak convergence of the associated probability measure. \( W(s) \) is a standard Brownian motion. \( W_c \) is the detrended Brownian motion that satisfies

\[ dW_c(s) = cW_c(s) ds + dW(s) \]
and $W_c(0) = 0$, and

$$W'_c(s) = W_c(s) - \int_0^1 W_c(r) \, dr.$$  

Finally, $z$ is a standard normal random variable that is asymptotically independent of the functionals of $W_c(s)$. (14) is directly from Hansen (1999), and (15) is from Elliott and Stock (1994). As Elliott and Stock (1994) note, if $\delta = 0$ then the $t$-statistic on $\gamma$ is asymptotically normal, and it is also independent of the demeaned augmented Dickey-Fuller $t$-test for $\rho = 1$. However, as the results in Stambaugh (1999) demonstrate, $\delta \neq 0$ in the excess return and dividend yield data.

The relevant point here is that both of these distributions are dependent on the value of $\rho$ through $c$, which is not consistently estimated by OLS. Equations (14) and (15) together are the bivariate generalization of the observation in Hansen (1999) that the standard $t$ test for $\rho$ (now $\rho$ and $\gamma$) is not pivotal. Following Hansen (1999), the sample $Z_T = (y_T, x_T)$ is generated from a distribution

$$G_T(z | \gamma, \rho, \eta) = P(Z_T \leq z | \gamma, \rho, \eta)$$  

where $(\gamma, \rho) \in \mathbb{R}^2$ and $\eta \in \Xi_1 \times \Xi_2$. The product space $\Xi_1 \times \Xi_2$ is endowed with some metric $d(\eta, \eta')$. Corresponding to each parameter pair $(\gamma, \rho)$ there exists some estimator $\hat{\eta}(\gamma, \rho) \in \Xi_1 \times \Xi_2$.

The Wald statistic

$$W_T(\gamma, \rho) = \left[ \hat{\gamma} - \gamma \right] \left( \text{cov}(\hat{\gamma}, \hat{\rho}) \right)^{-1} \left[ \hat{\rho} - \rho \right]$$

testing the hypothesis $H_0 : \gamma_0 = \gamma$ and $\rho_0 = \rho$ has a sampling distribution that might depend on $(\gamma, \rho, \eta)$. This distribution function is denoted

$$F_T(z | \gamma, \rho, \eta) = P(W_T(\gamma, \rho) \leq z | \gamma, \rho, \eta),$$

\footnote{See also the results in Stambaugh (1999).}
and the inverse of this distribution function is \( q_T(\bullet \mid \gamma, \rho, \eta) \); i.e., the \( \theta \) quantile of \( W_T(\gamma, \rho) \) is \( q_T(\theta \mid \gamma, \rho, \eta) \). Finally, the \( \theta \)-level grid bootstrap confidence set for \((\gamma, \rho)\) is

\[
C_g = \left\{ (\gamma, \rho) \in \mathbb{R}^2 : W_T(\gamma, \rho) \leq q^*_T(\theta \mid \gamma, \rho) \right\},
\]

where \( q^*_T(\theta \mid \gamma, \rho) = q^*_T(\theta \mid \gamma, \rho, \hat{\eta}(\gamma, \rho)) \) is the bootstrap estimate of the \( \theta \) quantile function for the Wald statistic, as a function of \((\gamma, \rho)\).

The extension of Proposition 2 in Hansen (1999) states:

**Proposition:** Suppose that for a sequence of estimators \( \hat{\eta}(\gamma, \rho) \in \Xi_1 \times \Xi_2 \) and

\[
d(\hat{\eta}(\gamma, \rho), \eta) \to_p 0.
\]

Suppose that for all sequences \( \eta_T \) such that

\[
d(\eta_T, \eta) \to 0
\]

that

\[
F_T(z \mid \gamma, \rho, \eta_T) \Rightarrow F(z \mid \gamma, \rho, \eta),
\]

where \( F(z \mid \gamma, \rho, \eta) \) is a continuous distribution function, then

\[
P((\gamma, \rho) \in C_g) \to \theta
\]

as \( T \uparrow \infty \).

The proof in Hansen (1999) carries over directly to this bivariate confidence set. The basic steps of the proof involve the definition of the quantile function and the uniform convergence of the empirical distribution of the Wald statistic to its asymptotic limit.

As Hansen (1999) notes: “The requirement (in the proposition) is that the nuisance parameters \((\eta(\gamma, \rho))\) are consistently estimated, while no restriction is made concerning the estimate(s) of the parameter(s) of interest.” [Hansen (1999), page 596].
This is important – and provides the grid-$W$ confidence set with broader consistency properties relative to a conventional asymptotic confidence set precisely because it does not require consistent estimation of the local to unity parameter, $c$, when the order of integration of dividend yields is unknown.

The second major result that we need is the extension of Theorem 2 in Hansen (1999), which shows that the proposition above applies to the predictability regression case.

**Theorem:** In model (11) through (13), if $\rho = 1 + c/T$ and $E(\eta_{2t}^{2r}) < \infty$ for some $r > 1$, then

$$P((\gamma, \rho) \in C_\theta) \rightarrow \theta$$

as $T \uparrow \infty$.

Again, the proof in Hansen (1999) carries over to this setting, using a multivariate version of Donsker’s Theorem (Davidson (1994), Theorem 27.17) and a generalization of Hansen (1992) Theorem 3.1 to the bivariate case.
B A Small Monte Carlo Analysis

In this appendix, we provide some limited information on the size of the grid-W confidence set for a known data-generating process. The excess returns and dividend yields are generated under the restricted first-order VAR from equation (2) in the body of the paper. The innovations to these series are GARCH(1,1) driven by Student-t errors with constant unconditional covariance between the shocks.

The data is generated under the null hypothesis that returns are not predictable ($\beta = 0$) and that dividend yields are very persistent but stationary ($\rho = 0.99$). The remainder of the model parameters are set to match the moments in the actual data. Under these assumptions, the standard confidence set is asymptotically correct. The issue here is which confidence set is more accurate in finite sample.

The design of this Monte Carlo is straightforward:

1. Generate 10,000 time series of length $T = 400$ (months) using the VAR-GARCH(1,1) data-generating process.

2. For each simulated sample path, estimate VAR in equation (2) and construct the 90% confidence set for:

   (a) the conventional asymptotic confidence set based on the $\chi^2$ distribution for the Wald test,

   (b) the grid-W confidence set using a grid size of size $13 \times 13$, 1000 bootstrap samples, with a pre-whitened block bootstrap using overlapping blocks of size 60,\(^{15}\) and

   (c) a grid-W estimator where the bootstrapping is done directly from the true parametric model without blocking.

\(^{15}\)Since the grid is actually coarser than the grid applied in the analysis in the body of the paper, the results reported below should be conservative.
3. For each of the 10,000 confidence sets, we count the number of time the \((\beta, \rho)\) pair \((0.0, 0.99)\) actually falls into the 10% rejection region.

The actual rejections rate provide information about the actual size of the test, as opposed to the nominal size (based on the asymptotic approximation). The results of this experiment are:

<table>
<thead>
<tr>
<th></th>
<th>Conventional Asymptotic</th>
<th>grid-W</th>
<th>grid-W (Exact)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Rejection Rates</td>
<td>35.14%</td>
<td>5.88%</td>
<td>9.98%</td>
</tr>
</tbody>
</table>

The conventional confidence set rejects the true null of 'no predictability' at a rate that is far in excess of the nominal 10% level. Consistent with the results in Figures 2 and 4, the larger confidence set of the grid-W test should reject \(H_0\) less frequently, and we find that this is indeed the case. In this particular simulation, we find that the actual rejection rate is too low (compared to the nominal size), but actual and nominal size are much closer than for the conventional test statistic. The grid-W with an exact bootstrap does better than the grid-W test based on the bootstrap algorithm that uses a block design. The actual rejection probability corresponds almost exactly to the nominal size of the test (9.98% instead of 10%). This result is interesting because it provides some evidence about how the block algorithm can affect the actual size of a test based on the confidence set, but this superior performance also reflects the unrealistic situation of being able to precisely specify an accurate parametric model for the residuals.

These results suggest that the grid-W confidence set may possess substantial finite-sample advantages over a test based on conventional asymptotic theory – *even in a situation where that theory is theoretically accurate*. Important directions for future work include: (*i*) examining the power properties of tests based on the grid-W confidence set for reasonable alternative models to no predictability, and (*ii*) examining the performance of grid-W based tests for different parameterizations of GARCH.
models and for stochastic volatility data-generating processes. The current version of
the grid-W computations are constructed in Matlab, and these future Monte Carlo
studies will require that the code be rewritten in Fortran or C++ in order to speed
up the substantial amount of time spent in loops while recursively building the time
series that are required at each point in the parameter grid.
References


