

# ASYMPTOTICS FOR OUT OF SAMPLE TESTS OF CAUSALITY

BY MICHAEL W. MCCRACKEN<sup>1</sup>

NOVEMBER 8, 1999

## HEADNOTE

This paper presents analytical and numerical evidence concerning out of sample tests of causality. The relevant environment is one in which the relative predictive ability of two nested parametric regression models is of interest. Results are provided for three statistics: a regression-based statistic suggested by Granger and Newbold (1977), a t-type statistic comparable to those suggested by Diebold and Mariano (1995) and West (1996), and an F-type statistic akin to Theil's U (1966). Since the limiting distributions under the null are nonstandard, tables of asymptotically valid critical values are provided. The null limiting distributions indicate that overfit models should predict poorly and that the Principle of Parsimony should be applied judiciously. Power calculations under a local alternative provide some guidance on the choice of test statistic and the percentage of the sample withheld for predictive evaluation.

Keywords: causality, forecast evaluation, hypothesis testing, model selection.

JEL categories: C12, C32, C52, C53.

*Department of Economics, Louisiana State University, 2107 CEBA, Baton Rouge LA, 70803;*  
*mmccrac@unix1.sncc.lsu.edu.*

## 1. INTRODUCTION

EVALUATING A TIME SERIES models' ability to forecast is one method of determining its usefulness. Tegene and Kuchler (1994), Swanson and White (1995), Huh (1996), Diebold and Kilian (1997), and Sullivan, Timmermann and White (1998) are a few examples of applications that have determined the appropriateness of a model based on its ability to predict Out-Of-Sample (OOS). When using this methodology a model is determined to be valuable if the resulting forecast errors are deemed small relative to some loss function. Typically this loss function is mean squared error (MSE) though others such as mean absolute error (MAE) and directional accuracy have been used by Leitch and Tanner (1991) and Breen, Glosten and Jagannathan (1989) respectively. This OOS methodology is in contrast to traditional methods (like the classical F-test reported by most statistical software) that determine quality of the predictive model based on its ability to replicate or "fit" the same realizations used to estimate the model.

This paper contributes to recent analytical work on OOS model evaluation, specifically that of West (1996), by providing asymptotic results for OOS tests that compare the predictive ability of two nested models when parameters are estimated. Null limiting distributions are derived for three commonly used tests that compare the OOS predictive ability of two nested models: a regression-based test for equal MSE proposed by Granger and Newbold (1977), a similar t-type test commonly attributed to either Diebold and Mariano (1995) or West (1996), and an F-type test similar in spirit to Theil's U (1966) but perhaps closer to in-sample likelihood ratio tests. Since the limiting distributions of the former two tests are identical they will be referenced simultaneously as "OOS-t" tests; the latter test will be referenced as an "OOS-F" test.

The limiting null distributions of both the OOS-t and OOS-F tests are non-standard. Each can be written as functions of stochastic integrals of quadratics of Brownian Motion. The distributions bear some resemblance to those in Andrews (1993) but are distinct. Tables are provided in order to facilitate the use of these distributions. A limited collection of analytical and numerical results regarding the local power of these tests is also provided. Monte Carlo evidence on the finite sample size and power of these tests and an empirical example can be found in Clark and McCracken (1999).

There are a number of interesting implications of the asymptotics under the null and under the local alternative. First, the null asymptotics provide a simple method of constructing asymptotically valid tests of OOS predictive ability between two nested models. A test can be conducted by simply consulting the provided tables of estimates of asymptotically valid critical values.

In addition, the null asymptotics have implications for the Principle of Parsimony and overfitting when OOS predictive ability is the objective. Assuming for the moment that the predictive model is linear, we know that in-sample predictive ability improves deterministically with the number of extraneous regressors. The results of this paper show that OOS quite the contrary is true. OOS the probability that the unrestricted model has lower predictive ability than the restricted model is increasing in the number of extraneous regressors. This result is particularly intriguing in the context of comparing the predictive ability of the random walk and economic models of asset movements. Meese and Rogoff (1983, 1988), Wolff (1987), Chinn and Meese (1995), and Berkowitz and Giorgianni (1999) are a few examples of such horse races.

Finally, the local alternative results indicate that the choice of sample split and the number of extraneous parameters in the unrestricted model jointly determine whether the OOS-t or OOS-F test is more powerful. The OOS-F has greater local power when the post-sample size is small relative to the in-sample size and when the number of extraneous parameters is small. As more of the sample is used for post-sample evaluation or when the number of extraneous parameters is large, the OOS-t tends to be more powerful. The choice of optimal sample split is less clear and is left to Section four.

The remainder of the paper will proceed as follows. Section two introduces the OOS methodology and provides a brief literature review. The review focuses on uses of the OOS methodology to date and potential applications of the results contained in this paper. Section three and its subsections provide notation, assumptions, theorems and corollaries regarding the null asymptotics. Section four provides a limited set of results regarding the power of both the OOS-t and OOS-F tests under a sequence of local alternatives. Section five concludes and suggests directions for future research. All proofs are presented within the Appendix.

## 2. LITERATURE REVIEW

Recent work by West (1996) has shown how to construct asymptotically valid OOS tests of predictive ability when forecasts are generated using estimated parameters. He provides conditions under which t-type statistics will be asymptotically standard normal. These conditions extend and clarify previous analytical work on OOS hypothesis testing made by Mincer and Zarnowitz (1969), Chong and Hendry (1986), Hoffman and Pagan (1989), Fair and Shiller (1989, 1990), Mizraeh (1992), and Diebold and Mariano (1995).

More recent work on OOS hypothesis testing has also developed. Corradi, Swanson and Olivetti (1999) extend previous work to allow for the comparison of non-nested models when cointegrating relationships exist. McCracken (1999a) provides analytical results for constructing OOS tests when the test involves non-smooth functions such as the indicator or absolute value function. Harvey, Leybourne and Newbold (1998) construct tests of equal predictive ability in the presence of ARCH. Diebold, Gunther and Tay (1997) discuss the evaluation of density forecasts. White (1999) shows how to use the bootstrap to compensate for data-snooping biases when comparing the predictive ability of a large number of models. Sanchez (1998) tests for unit roots using OOS forecast errors.

One test that is considered by several of these authors is whether two models have the same predictive ability with respect to some loss function  $L(\cdot)$ . Diebold and Mariano (1995) suggest a test of the form

$$(2.1) \quad P^{-0.5} \hat{\Omega}^{-0.5} \sum_{t=R}^T [L(\hat{u}_{1,t+1}) - L(\hat{u}_{2,t+1})]$$

where  $\hat{\Omega}$  denotes a consistent estimate of the limiting variance of  $P^{-0.5} \sum_{t=R}^T [L(\hat{u}_{1,t+1}) - L(\hat{u}_{2,t+1})]$ ,  $T + 1 = P + R$ ;  $P$  is the number of OOS observations and  $R$  is the number of observations used to construct the first forecast. In (2.1),  $\hat{u}_{i,t+1}$   $i = 1, 2$  is the forecast error from model  $i$  observed at time  $t+1$  associated with a forecast from time  $t$ .

When each forecast is constructed using  $\hat{\beta}_{i,t}$ , an estimator of the parameters associated with model  $i$ , West (1996)

shows that the test statistic in (2.1) can be asymptotically standard normal. For this to be true however, some conditions must hold.

One condition is that the estimate of the limiting variance,  $\Omega$ , must be appropriately constructed. The estimated limiting variance should not only account for sample variation, heteroskedasticity and serial correlation but also for the fact that forecasts are typically made using parametric models for which the parameters are unknown. If the parameters are estimated using the random data then they too are random and may contribute to the limiting variance.<sup>2</sup> West (1996) provides the correct limiting variance. The correct limiting variance is sometimes complicated but West and McCracken (1998) show that many OOS tests can be conveniently constructed using regression-based tests. These artificial regressions are similar to in-sample diagnostic tests suggested by, for example, Pagan and Hall (1983).

Unfortunately, it is easy to overlook the most crucial condition for limiting normality. For the OOS t-test in (2.1) to be limiting standard normal  $\Omega$  must be positive. If  $\Omega$  is zero then  $P^{-0.5} \sum_{t=R}^T [L(\hat{u}_{1,t+1}) - L(\hat{u}_{2,t+1})] \rightarrow_p 0$ . The problem is more pronounced when we look back at (2.1). This OOS-t statistic involves  $\hat{\Omega}^{-0.5}$  as well. Using results in West (1996), we know that  $\hat{\Omega}$  converges in probability to zero when  $\Omega = 0$ . If we put the two items together it is unclear whether the OOS-t statistic is degenerate, divergent or bounded in probability. What is clear is that the limiting distribution will not be standard normal.

This last problem may seem unlikely but indeed it is quite common. Using results in West (1996) one can easily show that  $\Omega$  equals zero if the two parametric models are nested rather than non-nested. This has serious implications for OOS tests of causality and market efficiency for which the models are inherently nested.

For example, in testing for a causal relationship between aggregate advertising expenditure and aggregate consumption expenditure Ashley, Granger and Schmalensee (1980) construct an OOS-t statistic similar to that in (2.1). Using a method suggested by Granger and Newbold (1977) they test the null that advertising causes consumption using the t-statistic (and standard normal tables) associated with  $\alpha$  from the OLS estimated artificial regression

$$(2.2) \quad \hat{u}_{1,t+1} - \hat{u}_{2,t+1} = \alpha(\hat{u}_{1,t+1} + \hat{u}_{2,t+1}) + \text{error term}.$$

In (2.2)  $\hat{u}_{1,t+1}$  is the one-step ahead forecast error from an autoregressive model for aggregate consumption and  $\hat{u}_{2,t+1}$  is the one-step ahead forecast error from a bivariate autoregressive model for both aggregate consumption and aggregate advertising. Ashley (1981) uses similar methods to test for causality between the consumer price index and its dispersion across different consumption categories. Park (1990) tests for causal relationships in cattle markets using (2.2).

There are also a number of potential applications to tests for the predictability of asset returns and more generally for tests of market efficiency. If the null is that asset returns form a martingale difference sequence then any parametric model for asset returns nests the null model (i.e. a constant zero conditional mean function) within it. For example Mark (1995) constructs OOS-t statistics of the form (2.1) to test the null that changes in exchange rates

are unpredictable. If this is the case then the MSE using the null zero conditional mean model should equal the MSE using a linear model that depends upon certain fundamentals. Kilian (1997) constructs similar tests but under the null that changes in exchange rates form a martingale difference sequence around a nonzero unconditional mean.

It should be mentioned that Mark (1995) and Kilian (1997) each use the bootstrap when conducting their hypothesis tests in these Long-Horizon regressions. They do not reference standard normal tables per se. However, the reason they use the bootstrap is that they are concerned about finite sample size distortions relative to the (claimed) limiting standard normal distribution of (2.1). The results in Section three of this paper indicate that those distortions may also be because the limiting distribution is not standard normal nor is well approximated by a standard normal distribution.

This paper focuses on constructing asymptotically valid OOS sample tests that compare the predictability of two nested parametric models. Three different statistics are considered. The first two, those from (2.1) and (2.2), are OOS-t statistics. The third is an OOS-F statistic of the form

$$(2.3) \quad P \frac{(P^{-1} \sum_{t=R}^T L(\hat{u}_{1,t+1})) - (P^{-1} \sum_{t=R}^T L(\hat{u}_{2,t+1}))}{\hat{c}}.$$

In (2.3),  $\hat{c}$  converges in probability to a certain normalizing constant  $c$ . For the moment it suffices to focus on the most useful case in which  $\hat{c} = P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2$  and  $\hat{u}_{i,t+1} = L(\hat{u}_{i,t+1})$   $i = 1, 2$ . As in the descriptions of (2.1) and (2.2), the restricted and unrestricted models are referenced using the indexes  $i = 1$  and  $i = 2$  respectively.

The OOS-F statistic is not generally used in the form (2.3). For example, Leitch and Tanner (1991) simply report Theil's U without providing any formal test that the unrestricted model has a lower MSE than the random walk. Others, including Mark and Sul (1998) and Ashley (1998), test the null of equal MSE by bootstrapping the ratio of the restricted MSE to the unrestricted MSE. Another group including Urbain (1989), Pesaran and Timmermann (1995), and Swanson and White (1997a) test for equal predictive ability using model selection criteria. In this last case a statistic similar to (2.3) is constructed but includes penalty terms like those associated with well known information criteria (e.g. AIC, SBC, Hannan-Quinn, etc.).

I introduce the OOS-F for a number of reasons. First, it is essentially Theil's U when the null model is a random walk but allows for a wider range of nested parameterizations. Also, given the limiting distribution results in Section three there does not seem to be any need to include penalty terms; the OOS MSE is not decreasing in the number of extraneous parameters as is the case in-sample. Finally, it seems natural to use the OOS-F because it is a direct analog of the in-sample F-test.

### 3. THEORETICAL RESULTS

This section provides the null limiting distributions of both the OOS-t tests in (2.1) and (2.2) and the OOS-F test in (2.3). It does so in five subsections. Section 3.1 presents the basic environment while Section 3.2 presents the assumptions needed for the results in Section 3.3. Section 3.3 presents the limiting distribution of the OOS-t and OOS-F tests first allowing a wide range of likelihood-type loss functions to measure predictive ability. Section 3.4

specializes the results to the leading case in which parameters are estimated by NLLS and MSE is used to measure predictive ability. Since the null limiting distributions are nonstandard, tables of asymptotically valid critical values are provided. Section 3.5 provides a discussion of the asymptotic results and their relation to the Principle of Parsimony.

### 3.1 Environment

Throughout it will be assumed that there is an observed sample  $\{X_s\}_{s=1}^{T+I}$  of length  $T + I$ . Using that sample, the researcher wishes to compare the one-step ahead predictive ability of two nested parametric regression models. This structure allows for many of the relevant applications discussed in Section two. It does eliminate applications like those in Diebold and Nason (1990), Swanson and White (1997b) and McCracken (1999b) who use local-regression, series-based and kernel-based nonparametric methods to estimate the regression function and construct forecasts.

The focus on one-step ahead forecasts rather than  $\tau$ -step ( $\tau > 1$ ) is both substantive and for purposes of clarity. By limiting the discussion to one-step ahead forecasts I am able to derive results for a wide range of potential loss functions used to measure predictive ability. Asymptotics for multi-step forecasts are left to future research.

Given the pair of nested parametric regression models, two sequences of one-step ahead forecasts are constructed using one of three methods. These are referred to as the recursive, rolling and fixed sampling schemes. Within each of these schemes an initial in-sample portion of the data, of length  $R$ , is used to select the two nested models and estimate their respective model parameters. Using the chosen nested models, and the estimated parameters, a sequence of  $P$  one-step ahead forecasts is then generated. See West (1996), West and McCracken (1998), McCracken (1999a), and Pesaran and Timmermann (1999) for more discussion on the use of these three schemes. A brief description is given below.

Pagan and Schwert (1990) use the recursive sampling scheme. Under this scheme a sequence of parametric forecasts is generated with updated parameter estimates. Specifically, at each time  $t = R, \dots, T$  the parameter estimate  $\hat{\beta}_t$  depends explicitly on all information from  $s = 1, \dots, t$ . If OLS is used to estimate the parameters from a linear model with regressors  $Z_s$  and predictand  $y_s$  then  $\hat{\beta}_t = (t^{-1} \sum_{s=1}^t Z_s Z_s')^{-1} (t^{-1} \sum_{s=1}^t Z_s y_s)$ . The first forecast for models  $i = 1, 2$  is then of the form  $\hat{y}_{R+1}(\hat{\beta}_{i,R})$ . The resulting forecast error is constructed as  $\hat{u}_{i,R+1} = y_{R+1} - \hat{y}_{R+1}(\hat{\beta}_{i,R})$ . For some loss function  $L(\cdot)$  the loss associated with the first forecast is constructed as  $L(\hat{u}_{i,R+1})$  and will usually be denoted as  $L_{i,R+1}(\hat{\beta}_{i,R})$ . The second forecast,  $\hat{y}_{R+2}(\hat{\beta}_{i,R+1})$ , is constructed similarly using observations  $s = 1, \dots, R+1$ . The forecast error and loss associated with the second forecast is constructed as for the first forecast. This process is iterated  $P$  times so that for each  $t \in [R, T]$ , the parameter estimates are based upon all data  $s \in [1, t]$ .

Swanson (1998) uses the rolling sampling scheme. Under this scheme a sequence of parametric forecasts, forecast errors and losses are constructed in much the same way as the recursive scheme. What distinguishes the rolling from the recursive is its treatment of observations from the distant past. The rolling scheme uses only a fixed

window of the past  $R$  observations. As  $t$  increases from  $R$  to  $T$ , older observations are not used in estimating the parameters. If OLS is used to estimate the parameters using regressors  $Z_s$  and predictand  $y_s$  then  $\hat{\beta}_t = (R^{-1} \sum_{s=t-R+1}^t Z_s Z_s')^{-1} (R^{-1} \sum_{s=t-R+1}^t Z_s y_s)$ . This implies that the first rolling forecast,  $\hat{y}_{R+1}(\hat{\beta}_{i,R})$ , forecast error, and loss are identical to those for the recursive. The second rolling forecast,  $\hat{y}_{R+2}(\hat{\beta}_{i,R+1})$ , is constructed using only observations  $s = 2, \dots, R+1$  to estimate the model parameters. This implies that the second rolling forecast, forecast error, and loss are distinct from those using the recursive scheme. The process is iterated  $P$  times such that for each  $t \in [R, T]$  the parameter estimates are based upon all data  $s \in [t - R + 1, t]$ .

Ashley, Granger and Schmalensee (1980) use the fixed scheme. This method is distinct from the previous two in that the parameters are not updated with the introduction of new observations. Although this method may seem inefficient it is frequently used when the computational burden is large such as when artificial neural networks are used to form forecasts (Kuan and Liu, 1995). Since the parameter vector is estimated only once each of the  $P$  forecasts,  $\hat{y}_{t+1}(\hat{\beta}_{i,R})$ , uses the same parameter estimate.<sup>3</sup> If OLS is used to estimate the parameters using regressors  $Z_s$  and predictand  $y_s$  then  $\hat{\beta}_t = (R^{-1} \sum_{s=1}^R Z_s Z_s')^{-1} (R^{-1} \sum_{s=1}^R Z_s y_s)$ . Hence for each one-step ahead forecast from time  $t \in [R, T]$ , the parameter estimate is based only upon data  $s \in [1, R]$ .

Using each of the two series of subsequent forecast errors, one from the nesting model and one from the nested model, a test statistic of the form in either (2.1), (2.2) or (2.3) is constructed. Based upon the value of this statistic one either fails to reject or rejects the null of equal predictive ability. The null and alternative can be stated as

$$(3.1.1) \quad H_0: EL_{1,t}(\beta_1^*) \leq EL_{2,t}(\beta_2^*) \text{ vs. } H_A: EL_{1,t}(\beta_1^*) > EL_{2,t}(\beta_2^*).$$

The alternative is one-sided rather than two sided because the two models are nested by construction. Note that if a log-likelihood function is being used to evaluate predictive ability, (3.1.1) implies that  $L(\cdot)$  is defined as the negative of that log-likelihood function.

### 3.2 Assumptions

Before discussing specific assumptions, some notation is required. For the loss function  $L_{i,t}(\beta_i)$  let  $h_{i,t}(\beta_i)$  denote  $\partial L_{i,t}(\beta_i) / \partial \beta_i$  and  $q_{i,t}(\beta_i)$  denote  $\partial^2 L_{i,t}(\beta_i) / \partial \beta_i \partial \beta_i'$ . For any matrix  $A$  with elements  $a_{ij}$  let  $|A|$  denote  $\max_{i,j} |a_{i,j}|$ . For any  $(m \times n)$  matrix  $A$  with column vectors  $a_i$  let  $\text{vec}(A)$  denote the  $(mn \times 1)$  vector  $[a_1', a_2', \dots, a_n']'$ .

Without loss of generality let  $\beta_2^* \equiv (\beta_{2,1}^*, \beta_{2,2}^*)' = (\beta_{2,1}^*, 0)' = (\beta_1^*, 0)'$ . Define a selection matrix  $J \equiv (I_{k_1 \times k_1}, 0_{k_1 \times k_2})$  ( $k_1 \times k_1$ ,  $k > k_1$ ). Since the two models are nested we know that under the null,  $L_{1,t}(\beta_1^*) = L_{2,t}(\beta_2^*)$ ,  $Jh_{2,t} = h_{1,t}$  and  $Jq_{2,t}J' = q_{1,t}$  for all  $t$ .

The following assumptions are not intended to be necessary and sufficient, only sufficient.

ASSUMPTION 1: *The parameter vectors  $\beta_1^*$  and  $\beta_2^*$  are estimated by minimizing the aggregate loss functions,  $A_{1,t}(\beta_1)$  and  $A_{2,t}(\beta_2)$ . For  $i = 1, 2$ , and  $t = R, \dots, T$ ,  $A_{i,t}(\beta_i) = t^{-1} \sum_{j=1}^t L_{i,j}(\beta_i)$ ,  $R^{-1} \sum_{j=t-R+1}^t L_{i,j}(\beta_i)$  and  $R^{-1} \sum_{j=1}^R L_{i,j}(\beta_i)$  for the recursive, rolling, and fixed schemes respectively.*

This first assumption provides two pieces of information. Analytically it states that the parameter estimates are of the form  $\hat{\beta}_{1,t} = \operatorname{argmin} A_{1,t}(\beta_1)$  and  $\hat{\beta}_{2,t} = \operatorname{argmin} A_{2,t}(\beta_2)$ . This allows for both linear and nonlinear models estimated by OLS, NLLS, and maximum likelihood. The substantive part of the first assumption is that it requires that the loss function used to estimate the parameters and the loss function used to measure predictive accuracy are the same. An implication of this assumption is that if MSE is the measure of OOS predictive ability, parameters must be estimated using OLS, NLLS, or maximum likelihood under the additional assumption that the disturbances are normal. One benefit of Assumption 1 is that it otherwise does not place a restriction on the chosen loss function. For example if parameters are estimated by minimizing the negative of a log-likelihood and then the negative of that log-likelihood is used to measure predictive ability, the limiting distribution is essentially the same as if the model had been estimated by OLS and then MSE was used to measure predictive ability.<sup>4</sup>

ASSUMPTION 2: *For  $i = 1, 2$ , (a)  $\beta_i \in \Theta_i$ ,  $\Theta_i$  compact, (b)  $EL_{i,t}(\beta_i)$  is uniquely minimized at  $\beta_i^* \in \Theta_i$  with  $Eq_{i,t}$  nonsingular, (c) In an open neighborhood  $N_i$  around  $\beta_i^*$ , and with probability one  $L_{i,t}(\beta_i)$  is twice continuously differentiable, admitting a mean value expansion*

$$L_{i,t}(\beta_i) = L_{i,t}(\beta_i^*) + h'_{i,t}(\beta_i - \beta_i^*) + (0.5)(\beta_i - \beta_i^*)' q_{i,t}(\tilde{\beta}_i)(\beta_i - \beta_i^*)$$

*for some  $\tilde{\beta}_i$  on the line between  $\beta_i$  and  $\beta_i^*$ , (d) In the open neighborhood  $N_i$ , and for all  $t$  there exists a positive constant  $\varphi$  and a positive random variable  $m_t$  such that  $|q_{i,t}(\beta_i) - q_{i,t}(\beta_i^*)| \leq m_t |\beta_i - \beta_i^*|^\varphi$  with  $Em_t < \infty$  and  $\varphi < \infty$ , (e)  $\sup_{\beta_i \in \Theta_i} |A_i(\beta_i) - EL_{i,t}(\beta_i)| \rightarrow_{a.s.} 0$ .*

Assumption 2 insures that the parameters are identified and are consistently estimated. It is directly comparable to Theorem (2.1) of Newey and McFadden (1994). The substantive component of this assumption is the requirement that the loss function be twice continuously differentiable. This allows for MSE and many log-likelihood type measures of predictive ability but eliminates applications, like that of Weiss and Andersen (1984), that estimate the parameters using LAD and then use MAE as the measure of predictive ability.

ASSUMPTION 3: *Let  $U_t \equiv [h'_{2,t}, \operatorname{vec}(h_{2,t}h'_{2,t} - Eh_{2,t}h'_{2,t}), \operatorname{vec}(q_{2,t} - Eq_{2,t})]'$ . (a)  $EU_t = 0$ , (b)  $U_t$  is uniformly  $L^8$  bounded, (c) For some  $\delta > d > 2$ ,  $U_t$  is strong mixing with coefficients of size  $-8d/(8-d)$ , (d)  $\lim_{T \rightarrow \infty} T^{-1} E \sum_{j=1}^T U_j U_j' < \infty$ .*

The conditions in Assumption 3 differ from those in, say, West (1996) because the models are nested rather than non-nested. If the models are non-nested then the OOS-t statistics in (2.1) and (2.2) can be asymptotically standard normal and hence one needs to make assumptions sufficient for the application of a central limit theorem. West (1996) and West and McCracken (1998) use a central limit theorem derived by Wooldridge and White (1989). In this paper, the limiting distributions are comprised of functions of stochastic integrals of quadratics of Brownian Motion. Hence we require conditions sufficient for the joint weak convergence of partial sums and the averages of these partial sums to Brownian Motion and stochastic integrals of these Brownian Motion. Hansen (1992) provides sufficient conditions for just such a situation. The details of Assumption 3 above are directly comparable to those for Theorems (2.1) and (3.1) in Hansen (1992).

ASSUMPTION 4: (a)  $Eh_{2,t}h'_{2,t} = cEq_{2,t} \equiv cB_2^{-1}$  for a constant  $c$ , (b)  $E(h_{2,t} | h_{2,t-j}, q_{2,t-j}, j = 1, 2, \dots) = 0$ .

The reasons for imposing Assumption 4 are much the same as Assumption 1. In order to insure that the limiting distribution does not depend upon the underlying data generating process, additional conditions must be imposed on the loss function  $L$ . The first is that the loss function has the property that the expected outer product of the score is proportional to the expected hessian. Moreover that constant of proportionality must be positive and finite. If the loss function  $L(\cdot)$  is the negative of a log-likelihood then that constant is one.

The need for the constant  $c$  arises from the fact that MSE is the most common measure of predictability. If the disturbances from the parametric linear regression model  $y_t = Z'_{2,t}\beta_2^* + u_{2,t}$  are i.i.d. normal and conditionally homoskedastic with variance  $\sigma_u^2$  then the OLS estimates of  $\beta_2^*$  are numerically identical to those estimated using the log-likelihood. That is not the same as saying that  $h_{2,t}$  and  $q_{2,t}$  do not depend on whether you use MSE or the log-likelihood as your measure of predictive ability. For example if we use OLS to estimate the parameters then  $h_{2,t}^{(OLS)} = -2u_{2,t}Z_{2,t}$ ,  $q_{2,t}^{(OLS)} = 2Z_{2,t}Z'_{2,t}$  and hence  $Eh_{2,t}^{(OLS)}h_{2,t}^{(OLS)'} = 4\sigma_u^2EZ_{2,t}Z'_{2,t} \neq 2EZ_{2,t}Z'_{2,t} = Eq_{2,t}^{(OLS)}$ . Similarly if we minimize a negative log-likelihood  $h_{2,t}^{(MLE)} = -\sigma_u^{-2}u_{2,t}Z_{2,t}$ ,  $q_{2,t}^{(MLE)} = \sigma_u^{-2}Z_{2,t}Z'_{2,t}$  and hence  $Eh_{2,t}^{(MLE)}h_{2,t}^{(MLE)'} = \sigma_u^{-2}EZ_{2,t}Z'_{2,t} = Eq_{2,t}^{(MLE)}$ . This difference generates the need for the constant  $c$ . For a more detailed discussion see Section 3.4.

ASSUMPTION 5:  $\lim_{T \rightarrow \infty} P/R = \pi$ ,  $0 < \pi < \infty$ ,  $\lambda \equiv (1 + \pi)^{-1}$ .

This final assumption introduces the means by which the asymptotics are achieved. As in Hoffman and Pagan (1989), West (1996), and White (1999) the limiting distributions are derived by imposing a slightly stronger condition than simply that the sample size  $T$  becomes arbitrarily large. The additional condition is that both the number of in-sample ( $R$ ) and OOS ( $P$ ) observations become arbitrarily large at the same rate. This insures that the parameters estimated in-sample and certain OOS averages are both consistent estimators of their population level analogs.

### 3.3 Asymptotics Under the Null

This section presents the null limiting distributions of the OOS-t statistic in (2.1) and the OOS-F statistic in (2.3). Since the limiting distributions are non-standard, tables of critical values are provided that can be used to test the null in (3.1.1). We will return to the Granger-Newbold statistic, from (2.2), in Section 3.4 where the loss function is specialized to the case of MSE.

There are two main components of the OOS-t and OOS-F statistics in (2.1) and (2.3). Both test statistics depend upon  $\sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]$ . A second component,  $\sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]^2$ , arises in the OOS-t statistic from (2.1). This latter component is a denominator term that was originally designed to estimate the limiting variance of  $P^{-0.5} \sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]$  which, since the models are nested, is equal to zero. To see how these components affect the OOS-t and OOS-F statistics let's rewrite (2.1) and (2.3).

$$(3.3.1) \quad OOS-t = P^{-0.5} \hat{\Omega}^{-0.5} \sum_{t=R}^T [L(\hat{u}_{1,t+1}) - L(\hat{u}_{2,t+1})] = \frac{\sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]}{(\sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]^2)^{0.5}}$$

$$(3.3.2) \quad OOS-F = P \frac{(P^{-1} \sum_{t=R}^T L(\hat{u}_{1,t+1})) - (P^{-1} \sum_{t=R}^T L(\hat{u}_{2,t+1}))}{\hat{c}} = \frac{\sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]}{\hat{c}}.$$

We can see from (3.3.1) and (3.3.2) that the OOS-t and OOS-F are somewhat related. They differ in that the OOS-t has a denominator component that the OOS-F does not have. Notice that since the forecasts are 1-step ahead I am assuming that  $\hat{\Omega} = P^{-1} \sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]^2$  and hence one is not using a serial correlation consistent covariance matrix.<sup>5</sup> I emphasize this case because it is most common. Clark (1999) and Harvey, Leybourne and Newbold (1998) consider OOS tests when serial correlation is of concern.

To gain some intuition as to how these two components contribute to the limiting distributions, consider the following three lemmas. In the following, for  $i = 1, 2$ , define  $H_i(t)$  as  $t^{-1} \sum_{s=1}^t h_{i,s}$ ,  $R^{-1} \sum_{s=t-R+1}^t h_{i,s}$  and  $R^{-1} \sum_{s=1}^R h_{i,s}$  for the recursive, rolling and fixed schemes respectively. Also, define  $B_1$  such that  $Eh_{1,t}h_{1,t}' = cEq_{1,t} \equiv cB_1^{-1}$ . For matrices  $C$  and  $A$  defined in Lemma 3.1 let  $c^{-0.5} A' CB_2^{0.5} h_{2,t} = \tilde{h}_{2,t}$  and  $c^{-0.5} A' CB_2^{0.5} H_2(t) = \tilde{H}_2(t)$ .

LEMMA 3.1: (a) Let  $-J' B_1 J + B_2 = M$  and  $B_2^{-0.5} M B_2^{-0.5} = Q$ , then  $Q$  is idempotent. (b) Let  $A$  be a  $(k \times k_2)$  matrix with  $I_{k_2 \times k_2}$  on the upper  $(k_2 \times k_2)$  block and zeroes elsewhere. There exists a symmetric orthonormal matrix  $C$  such that  $Q = CAA'C$ .

LEMMA 3.2:  $\sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})] = c[\sum_{t=R}^T (T^{0.5} \tilde{H}_2(t))' (T^{-0.5} \tilde{h}_{2,t+1}) - (0.5)T^{-1} \sum_{t=R}^T (T^{0.5} \tilde{H}_2(t))' (T^{0.5} \tilde{H}_2(t))] + o_p(1)$ .

LEMMA 3.3:  $\sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]^2 = c^2 T^{-1} \sum_{t=R}^T (T^{0.5} \tilde{H}_2(t))' (T^{0.5} \tilde{H}_2(t)) + o_p(1)$ .

When deriving the limiting distribution of the in-sample F-statistic one first shows that the statistic can be written as a weighted quadratic of, say, a  $(k \times I)$  limiting standard normal random vector. The second step is to show that the weighting matrix is idempotent of, say, rank  $k_2 \leq k$ . The final step is to apply the continuous mapping theorem and conclude that the limiting distribution is chi-square with  $k_2$  degrees of freedom.

The OOS statistics are roughly the same, at least in spirit. They are comprised of weighted quadratics of standard normal random vectors for which the weighting matrix is idempotent. The OOS statistics differ in that they depend upon weighted averages of an entire sample path of these quadratics. To see this consider  $T^{0.5} \tilde{H}_2(t) \equiv T^{0.5} t^{-1} \sum_{s=1}^t \tilde{h}_{2,s} = (T/t)(T^{-0.5} \sum_{s=1}^t \tilde{h}_{2,s})$  for the recursive scheme and let  $W(s)$  denote a  $(k_2 \times I)$  standard Brownian Motion on  $[\lambda, 1]$  with  $W(0) = 0$  and  $\lambda \equiv (1 + \pi)^{-1}$ . Since the increments  $\tilde{h}_{2,s}$  are conditionally homoskedastic vector martingale differences with unit variance and  $T/t$  is bounded by Assumption 5,  $T^{0.5} \tilde{H}_2(t)$  is well approximated (weakly) by  $s^{-1}W(s)$  for large enough  $T$ . The in-sample result can be thought of as just the endpoint of a similar, but distinct, sample path.

Lemmas 3.2 and 3.3 also clarify the potential need for scaling by the factor  $c$ . When the OOS-F statistic is of interest Lemma 3.2 shows that the data generating process and loss function are irrelevant to the asymptotics but for the factor  $c$ . To eliminate that factor the OOS-F is defined relative to some consistent estimator  $\hat{c}$  of  $c$ . The OOS-t statistics do not require a consistent estimator of  $c$ . The reason for this is that  $c$  arises in both the numerator and denominator of (3.3.1) and hence cancels. Recall that the denominator of (3.3.1) will be akin to the square root of the right hand side of Lemma 3.3.

Lemmas 3.2 and 3.3 provide the building blocks for the following theorems.

THEOREM 3.1: Let  $\hat{c} \rightarrow_p c$ .  $OOS-F = \sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})] / \hat{c} \rightarrow_d F^1$  where  $F^1$  equals  $\int_{\lambda}^1 s^{-1} W'(s) dW(s) - (0.5) \int_{\lambda}^1 s^{-2} W'(s) W(s) ds$ ,  $\lambda^{-1} \{W(1) - W(\lambda)\}' W(\lambda) - (0.5) \pi \lambda^{-1} W'(\lambda) W(\lambda)$ , and  $\lambda^{-1} \int_{\lambda}^1 \{W(s) - W(s - \lambda)\}' dW(s) - (0.5) \lambda^{-2} \int_{\lambda}^1 \{W(s) - W(s - \lambda)\}' \{W(s) - W(s - \lambda)\} ds$  for the recursive, fixed and rolling schemes respectively.

THEOREM 3.2:  $OOS-t = \sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})] / (\sum_{t=R}^T [L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]^2)^{0.5} \rightarrow_d F^2$  where  $F^2$  equals  $F^1 / [\int_{\lambda}^1 s^{-2} W'(s) W(s) ds]^{0.5}$ ,  $F^1 / [\pi \lambda^{-1} W'(\lambda) W(\lambda)]^{0.5}$ , and  $F^1 / [\lambda^{-2} \int_{\lambda}^1 \{W(s) - W(s - \lambda)\}' \{W(s) - W(s - \lambda)\} ds]^{0.5}$  for the recursive, fixed and rolling schemes respectively.

There are a number of things to notice about Theorems 3.1 and 3.2. The first is that both the OOS-t and OOS-F statistics are asymptotically pivotal. This permits the construction of estimates of asymptotically valid critical values without knowledge of the underlying data generating process. Given these critical values, one can conduct

asymptotically valid tests for equal forecast accuracy between two nested parametric models. Tables of these critical values are provided and discussed in Section 3.4.

A second fact to note is that the limiting distributions do not depend upon the choice of loss function  $L(\cdot)$ . So long as the parameters are estimated using the same loss function as is used to measure predictive ability the loss function itself has no effect on the limiting distribution. That does not imply that finite sample size and power performance is invariant to the choice of loss function.

Though the null limiting distributions do not depend upon the loss function itself, the distributions are dependent upon two parameters. The first is the number of excess parameters  $k_2$ . We can see this in the dimension of the vector Brownian Motion  $W(s)$ . It is easier to see if we rewrite  $F^l$ . Consider the recursive sampling scheme. If we let  $W_i(s)$  denote the  $i$ th element of  $W(s)$  then

$$(3.3.3) \quad F^l = \sum_{i=1}^{k_2} \left[ \int_{\lambda}^1 s^{-l} W_i(s) dW_i(s) - (0.5) \int_{\lambda}^1 s^{-2} W_i^2(s) ds \right].$$

This representation is useful for two purposes. First, it provides some insight into the effect that  $k_2$  has on the mean of  $F^l$ . Taking expectations, and noting that each of the  $i = 1, \dots, k_2$  elements is independent and identically distributed, it is straightforward to show that

$$(3.3.4) \quad \begin{aligned} E(F^l) &= -(0.5)k_2 \ln(1 + \pi) && \text{for the recursive scheme} \\ &= -(0.5)k_2\pi && \text{for the rolling and fixed schemes.} \end{aligned}$$

Hence as  $k_2$  increases we expect the distribution of the OOS-F statistic to drift into the negative orthant. This is occurring because the first term in  $F^l$  is mean zero for all  $k_2$  while the second term is increasingly negative. See Section 3.4 for a discussion of this fact and its relevance to the Principle of Parsimony. A less important effect due to  $k_2$  is on the variance of  $F^l$ . Since  $F^l$  can be written as the sum of  $k_2$  i.i.d. terms we know that the variance is monotonically increasing in  $k_2$ .

The effect of  $k_2$  on  $F^2$  is less clear. Since  $F^2$  is nonlinear in its components it is difficult to analytically derive properties concerning its mean and variance. Numerical results suggest that the mean does become increasingly negative in  $k_2$  but that the variance is relatively constant in  $k_2$ .

There is a second reason that the representation in (3.3.3) is useful. One of the assumptions in Section 3.2 was that  $k_2$  was finite. We can heuristically see the need for that assumption by simply taking the limit of  $F^l$  as  $k_2$  goes to infinity. It diverges under the null. Hong and White (1995) show, in the context of series-based nonparametric regressions, that the in-sample F-statistic also diverges under the null as the number of series-terms increases to infinity. They suggest a transformed version of the F-statistic that is asymptotically standard normal as the number of series-terms increases to infinity. It seems that such an argument could be used for the OOS-F statistic as well. Such a proof is beyond the scope of this paper and is left for future research.

A second parameter,  $\pi$ , affects the null limiting distribution of both the OOS-t and OOS-F. It affects the limiting distributions in two ways. It directly affects the weights on each of the components of the statistics (recall

that  $\lambda = (I + \pi)^{-1}$ ). It also affects the range of integration on each of the stochastic integrals through  $\lambda$ . Since the parameter  $\pi$  enters both  $F^1$  and  $F^2$  nonlinearly its affect on their distributions is less clear than it was for  $k_2$ . Looking at (3.3.4) we can say with certainty that the mean of  $F^1$  decreases with  $\pi$  just as it did with  $k_2$ . Numerical results indicate that the variance of  $F^1$  is also monotonically increasing in  $\pi$  for a fixed value of  $k_2$ . For  $F^2$ , numerical results suggest that the mean is decreasing and the variance is increasing in  $\pi$ , but to a lesser extent than for  $F^1$ .

### 3.4 Null Asymptotics When MSE is the Measure of Predictive Ability

If one is interested in using MSE as the measure of predictive ability there are two loose ends remaining from Section 3.3. The first is that the limiting distribution of the OOS-t statistic from (2.2) has not been provided. This has been done to place some added emphasis on the fact that the OOS-t from (2.1) can be applied to a wider range of measures of predictive ability than just MSE. The OOS-t in (2.2) is only applicable when MSE is the measure of predictive ability. The first loose end is alleviated by the following theorem.

**THEOREM 3.3:** Let  $L_{i,t+1}(\hat{\beta}_{i,t}) = u_{i,t+1}^2(\hat{\beta}_{i,t}) \equiv \hat{u}_{i,t+1}^2$  and define  $a_{0,T} \equiv P^{-1} \sum_{t=R}^T [\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]$ ,  $a_{1,T} \equiv P^{-1} \sum_{t=R}^T [\hat{u}_{1,t+1} - \hat{u}_{2,t+1}]^2$ ,  $a_{2,T} \equiv P^{-1} \sum_{t=R}^T [\hat{u}_{1,t+1} + \hat{u}_{2,t+1}]^2$  and  $a_{3,T} \equiv P^{-1} \sum_{t=R}^T [\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]^2$ .  
 $[a_{1,T} a_{2,T} - a_{0,T}^2]^{-0.5} P^{0.5} a_{0,T} - [a_{3,T}]^{-0.5} P^{0.5} a_{0,T} \rightarrow_p 0$ .

Theorem 3.3 states that the two OOS-t statistics are asymptotically equivalent. Hence one can use the same critical values to construct asymptotically valid tests of equal predictive ability when using either of the tests.

Tables I - III relate to the OOS-t statistic. These were generated numerically using the limiting distribution in Theorem 3.2 and hence can be considered estimates of the true asymptotic critical values. The critical values are the 90th, 95th and 99th percentiles of 5000 independent draws from the distribution of  $F^2$  for a given sampling scheme and value of  $k_2$  and  $\pi$ . Generating these draws proceeded as follows. Weights that depend upon  $\pi$  were estimated in the obvious way using  $\hat{\pi} = P/R$ . The necessary  $k_2$  Brownian Motions were simulated as random walks each using an independent sequence of 10,000 i.i.d.  $N(0, T^{0.5})$  increments. The integrals were emulated by summing the relevant weighted quadratics of the random walks from the  $R+1$ st observation to the  $T$ th observation. The random number generator was seeded so that all  $k_2$  and  $\pi$  pairs and all sampling schemes use the same 5000 draws of  $k_2$  sequences of 10,000 i.i.d.  $N(0, T^{0.5})$  increments.

A brief listing of critical values is provided in Tables I - III. Each table corresponds to either the recursive, rolling or fixed scheme. Within each table there are 330 critical values. Each of these correspond to one permutation of three parameters:  $k_2 = \{1, 2, 3, \dots, 9, 10\}$ ,  $\pi = \{0.1, 0.2, 0.4, \dots, 1.0, 1.2, \dots, 2.0\}$  and nominal size of the test takes the values  $\{0.01, 0.05, 0.10\}$ .<sup>6</sup> Tables that allow for larger values of both  $k_2$  and  $\pi$  are available from the author upon request.

The second loose end concerns the OOS-F test when MSE is the measure of predictive ability. Recall the discussion following Assumption 4 in Section 3.2. There we discussed how  $c$  was defined. Specifically it was

shown that if OLS was used to estimate the parameters and MSE was used to measure predictive ability then  $Eh_{2,t}^{(OLS)}h_{2,t}^{(OLS)'} = 2\sigma_u^2 Eq_{2,t}^{(OLS)}$ . This was in contrast to the case where the parameters were estimated by minimizing a negative log-likelihood and then the same negative log-likelihood was used to measure predictive ability. In this latter case we know that  $Eh_{2,t}^{(MLE)}h_{2,t}^{(MLE)'} = Eq_{2,t}^{(MLE)}$ .

The constant  $c$  is intended to soak up any difference between the expected outer product of the score and the expected hessian determined by the choice of loss function. Assumption 4 defines  $c$  as  $2\sigma_u^2$  when MSE is the measure of predictive ability. If so we can consistently estimate  $c$  using  $\hat{c} = 2(P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2)$ .<sup>7</sup>

Unfortunately this is not the most commonly used normalization factor for this type of statistic. When the in-sample F-test is constructed the denominator is the mean square error associated with the unrestricted regression. When Ashley (1998), Mark (1995), Kilian (1997) and others bootstrap versions of this statistic the denominator term is  $P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2$ . When Pesaran and Timmermann (1995) and Swanson and White (1997a) use OOS information criteria (such as AIC, SBC, Hannan-Quinn, etc.) to compare the predictive ability of two nested models they are effectively normalizing by  $P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2$ . In these cases the OOS-F with  $\hat{c} = 2(P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2)$  is not applicable.

By modifying the definition of the OOS-F in accordance with these applications we have

$$(3.4.1) \quad (\text{modified}) \text{ OOS-F} = \frac{\sum_{t=R}^T [\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]}{P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2}.$$

The limiting distribution of this statistic follows immediately from Theorem 3.1.

**COROLLARY 3.1:** *Let  $L_{i,t+1}(\hat{\beta}_{i,t}) = u_{i,t+1}^2(\hat{\beta}_{i,t}) \equiv \hat{u}_{i,t+1}^2$ , then  $[P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2]^{-1} \sum_{t=R}^T [\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2] \rightarrow_d 2F^1$ .*

Since MSE is the most heavily used measure of predictive ability I focus on the limiting distribution in Corollary 3.1 rather than that in Theorem 3.1. Tables IV - VI provide the critical values associated with constructing an asymptotically valid test of the null of equal MSE between two nested models using the modified OOS-F statistic in (3.4.1). Each table corresponds to either the recursive, rolling or fixed sampling scheme. The 330 values reported in Tables IV - VI correspond to the same permutations of  $k_2$ ,  $\pi$  and nominal size of the test that were used in Tables I - III. More detailed tables are available from the author upon request.

It should be emphasized that Tables IV - VI cannot be directly applied to applications where a negative log-likelihood is used to measure predictive ability. It can be used with a simple adjustment. If one is interested in using the OOS-F statistic in the form presented in Theorem 3.1 the critical values presented in Tables IV - VI can be used only after they have been divided by two. Suppose that the recursive scheme is used,  $k_2 = 1$  and  $\pi = 0.4$ . If MSE is used to measure predictive ability, and Corollary 3.1 is applied, then the critical value associated with a 5%

test of the null hypothesis is 1.298. If instead the negative log-likelihood associated with a normal random variate is used to measure predictive ability, and hence Theorem 3.1 is applied, the appropriate critical value is 0.649.

In each of Tables I - VI it is also useful to note that the critical values are not generally monotone in either  $k_2$  or  $\pi$ . In Section 3.3 we discussed the fact that the means of the OOS-F and OOS-t statistics were monotone decreasing in both  $k_2$  and  $\pi$  and hence we expected the distributions to generally drift into the negative orthant as these parameters increased. That does not imply that the upper tails decrease monotonically. Consider the OOS-F statistic for a fixed value of  $\pi$  but allow  $k_2$  to increase. As  $k_2$  increases the mean of the distribution becomes increasingly negative while at the same time the variance increases. Put together, these two forces imply that the upper tail need not be monotonically decreasing in  $k_2$ .

### 3.5 Discussion of the Null Distributions:

#### *The Principle of Parsimony and Why Overfit Models Predict Poorly*

The preceding two sections present the null limiting distributions of and the critical values associated with the OOS-t and OOS-F statistics. A quick glance at Tables I - VI indicates that the distributions are nonstandard; they are not well approximated by either the normal or chi-square distributions.

The density plots in Figures 1 - 4 are intended to provide some feel for the behavior of the distributions of  $2F^l$  and  $F^2$  corresponding to the (modified) OOS-F and OOS-t statistics respectively. In order to reduce the number of plots I focus exclusively on the recursive sampling scheme. Plots for the rolling and fixed schemes are qualitatively similar in shape. They do differ in location and scale. When the rolling and fixed schemes are used the statistics have heavier tails and drift into the negative orthant much quicker than when the recursive scheme is used. For example when  $k_2 = 20$  and  $\pi = 50$  the 95th percentiles associated with the (modified) OOS-F statistic are -64.018, -939.127, and -540.728 for the recursive, rolling and fixed schemes respectively.

Figure 1 is comprised of four plots. Each show the effect on the density of  $2F^l$  when  $\pi$  increases from 0.2 to 1.0 to 2.0 holding  $k_2$  constant at 1, 2, 5, and 10. Figure 3 is the same but for  $F^2$ . In each Figure and plot, as  $\pi$  increases the probability that the statistic is negative increases. This is particularly true for the (modified) OOS-F statistic in Figure 1.

Figure 2 is comprised of four plots. Each shows the effect on the density of  $2F^l$  when  $k_2$  increases from 1 to 2 to 5 to 10 to 20 holding  $\pi$  constant at 0.2, 1, 2 and 50.0. Figure 4 is the same but for  $F^2$ . In each Figure and plot, as  $k_2$  increases the probability that the statistic is negative increases. Again, this is especially true for the (modified) OOS-F statistic in Figure 2.

These density plots and the associated percentiles indicate that the probability that both the OOS-F and OOS-t statistics are negative increases in both  $k_2$  and  $\pi$ . For the moment focus on the (modified) OOS-F statistic. Algebraically this implies that under the null

$$(3.5.1) \quad \lim_{T \rightarrow \infty} \text{Prob} \left( \frac{\sum_{t=R}^T [\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]}{P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2} \leq 0 \right) = \lim_{T \rightarrow \infty} \text{Prob} \left( P^{-1} \sum_{t=R}^T \hat{u}_{1,t+1}^2 \leq P^{-1} \sum_{t=R}^T \hat{u}_{2,t+1}^2 \right)$$

increases in both  $k_2$  and  $\pi$ .

What makes (3.5.1) interesting is its affect on model selection based upon OOS predictive performance. In-sample we know that, when parameters are estimated by NLLS, the MSE from a restricted parametric regression model must be numerically greater than the MSE from an unrestricted parametric regression model that nests the restricted model. One repercussion of this numerical ordering of MSE's is on the application of the Principle of Parsimony.

Granger (1995): "If two models appear to fit the data equally well, choose the simpler model (that is the one involving the fewest parameters)."

When in-sample predictive ability is of interest, the fact that the unrestricted MSE must be less than or equal to that for the restricted model places the burden of proof on the unrestricted model. For a researcher to feel confident that the unrestricted model is providing information beyond that contained in the restricted model the unrestricted MSE must be "significantly" lower than the restricted MSE. If it is not significantly lower, the Principle of Parsimony says to choose the less parameterized model.

If OOS predictive ability is of interest that logic no longer holds. OOS the unrestricted MSE can be less than or greater than the restricted MSE. To make matters worse, the plots in Figures 1 - 4 indicate that the probability in (3.5.1) increases the larger is the number of extraneous parameters introduced in the unrestricted regression model. Moreover it seems that this probability increases to one as either  $k_2$  or  $\pi$  become arbitrarily large. This implies that when OOS predictive ability is of interest, the burden of proof that is solely on the unrestricted model in-sample, is also on the restricted model. The MSE from the restricted model must also be "significantly" lower than the MSE from the restricted model.<sup>8</sup>

It is for this reason that it is particularly important to apply tests of significance when comparing the OOS predictive ability of two nested models. Simply reporting the OOS MSE's from two nested models is insufficient. Consider the case in which the value of the OOS-F statistic is zero and hence the restricted and unrestricted models have "equal" predictive ability. The critical values from Tables IV - VI indicate that  $k_2$  and  $\pi$  can always be chosen large enough that zero lies in the rejection region for a given nominal size of the test. If this is the case, we reject the null even though the OOS MSE's are the same.

Another related implication is on the use of OOS information criteria to identify regression models. Swanson and White (1995, 1997a) have applied this methodology. There, and in other applications, the penalty terms used were those commonly associated with in-sample information criteria. In other words, the penalty terms were positive, additive, and increased in  $k_2$ . This form of penalty term is intended to serve as a statistical mechanism for the Principle of Parsimony.

But as mentioned above, it is not clear how the Principle of Parsimony should be applied in an OOS context. As Figures 1 – 4 indicate, for small  $k_2$  and  $\pi$  the unrestricted model has a lower MSE than the restricted model a sizable percentage of the time. This holds even under the null. Hence it may be appropriate to use traditional information criteria for smaller values of  $k_2$  and  $\pi$ . As  $k_2$  and  $\pi$  increase, the restricted model tends to have a lower MSE under the null. In this case it is unclear why penalty terms would be necessary at all. Moreover, including penalty terms could potentially reduce the power of the model selection procedure by artificially inflating the measure of “predictive ability” associated with the unrestricted model, as measured by the information criterion. To eliminate such a problem it may even be the case that negative penalty terms are required in the construction of OOS information criteria particularly for larger values of  $k_2$  and  $\pi$ . In any event, it is not clear that traditional in-sample information criteria are appropriate in an OOS context. Development of a theory of model selection using OOS information criteria is left to future research.

It should be noted that most of the comments made above are based upon (3.5.1), which in turn relies upon Theorem 3.1. In Theorem 3.1 the results are only applicable to correctly specified regression functions. One cannot infer that the same would be true for misspecified models. It may very well be the case that a more heavily parameterized misspecified model has greater predictive ability than a less parameterized misspecified model. Extensions to misspecified models are left to future research.

#### 4. ASYMPTOTICS UNDER A LOCAL ALTERNATIVE

The null asymptotics provide us with a basis for constructing asymptotically valid tests for equal OOS predictive ability between two nested models. If we use the appropriate critical values from Tables I - VI then we know that for large enough  $T$  both the OOS-t and OOS-F statistics will be well sized. However, these null distributions do not provide us with any rationale for choosing between the OOS-t and OOS-F. The null distributions also do not provide us with any information on how to choose the sample split parameter,  $\pi$ , or the sampling scheme in order to maximize power of the test.

This section provides a limited set of evidence concerning the local power of both the OOS-t and OOS-F statistics for each of the recursive, rolling, and fixed sampling schemes and a limited range of values of  $k_2$  and  $\pi$ . The evidence suggests that the recursive scheme is usually the most powerful among the three schemes. The evidence also suggests that which of the OOS-t and OOS-F statistics is more powerful depends jointly upon the values of  $k_2$  and  $\pi$ . Clark and McCracken (1999) provide Monte Carlo evidence on the finite sample size and power performance of these statistics, and comparable tests of encompassing, when the nested model is autoregressive and the nesting model is bivariate vector autoregressive.

Rather than do a complete analysis of all possible local alternatives and parametric regression models I focus on the most relevant application. I presume that one is interested in measuring predictive ability using MSE as the loss function and that the parameters from two nested linear parametric regression model are estimated using OLS. The null and local alternative models can then be specified as

$$(4.1) \quad H_0: y_t = Z'_{1,t} \beta_1^* + u_t \text{ vs.}$$

$$H_A: y_t = Z'_{1,t} \beta_1^* + T^{-0.5} c^{0.5} Z'_{22,t} \beta_{22}^* + u_t = Z'_{2,t} \beta_2^* + (c^{0.5} T^{-0.5} - 1) Z'_{22,t} \beta_{22}^* + u_t$$

where  $Z_{2,t} = (Z'_{1,t}, Z'_{22,t})'$ ,  $\beta_2^* = (\beta_1^{*'}, \beta_{22}^{*'})'$ ,  $\beta_{22}^* \neq 0$  and  $u_t$  is a conditionally homoskedastic martingale difference sequence with unconditional variance  $\sigma_u^2$ .

This alternative specification is chosen for two reasons. The first is that linear models are the parametric regression models of choice in many applications. Examples include Clarida and Taylor (1997) and Meese and Rogoff (1983) where OOS predictive ability is of interest. The second reason is to simplify the algebra involved with deriving the limiting distributions under the local alternative.

Define  $\chi_3$  as  $\int_{\lambda}^1 s^{-1} W(s) ds$ ,  $\lambda^{-1} \int_{\lambda}^1 [W(s) - W(s - \lambda)] ds$  and  $\pi W(\lambda)$  for the recursive, rolling and fixed sampling schemes respectively. Furthermore, define  $\chi_2$  as the square of the denominator term in  $F^2$  from Theorem 3.2. In the following, Assumptions 1 - 5 are maintained.

**THEOREM 4.1:** Let  $L_{i,t+1}(\hat{\beta}_{i,t}) = u_{i,t+1}^2(\hat{\beta}_{i,t}) \equiv \hat{u}_{i,t+1}^2$  and define a selection matrix  $J_2 \equiv (0_{k_2 \times k_1}, I_{k_2 \times k_2})$  ( $k_2 \times k$ ,  $k > k_2$ ). Under the local alternative in (4.1), (modified) OOS-F  $\rightarrow_d F^3$  and OOS-t  $\rightarrow_d F^4$  where  $F^3 = 2F^1 + \pi \lambda \beta_{22}^{*'} J_2' B^{-0.5} Q B_2^{-0.5} J_2 \beta_2^* - 2 \beta_{22}^{*'} J_2' B_2^{-0.5} CA[W(1) - W(\lambda)]$  and  $F^4 = (0.5)F^3 / [\chi_2 + \pi \lambda \beta_{22}^{*'} J_2' B_2^{-0.5} Q B_2^{-0.5} J_2 \beta_{22}^* - 2 \beta_{22}^{*'} J_2' B_2^{-0.5} CA \chi_3]^{0.5}$ .

The first thing to note about Theorem 4.1 is that the statistics are not asymptotically pivotal. The local power of the tests depends upon the data generating process through  $B_2$  and  $\beta_{22}^*$ . Under the null  $\beta_{22}^* = 0$  and  $B_2$  was irrelevant. Both affect the limiting distribution under the local alternative. This is important because it implies that any given set of power calculations using the results of Theorem 4.1 should be interpreted with care. Local power of the test for one data generating process does not imply comparable local power for other data generating processes. Moreover, Nelson and Savin (1990) show that local asymptotics may provide a poor approximation to true finite sample power.

With these caveats in mind, Tables VII - IX provide a brief list of local power characteristics. The calculations apply the same methods used to construct the critical values in Tables I - VI. The random number generator was seeded so that the random walks used to emulate Brownian Motion under the null were also used under the alternative. In this way much of the null numerical calculations used to generate 5000 draws of  $F^1$  and  $F^2$  were directly applied in generating 5000 draws of  $F^3$  and  $F^4$ .

What distinguishes the two simulations is the need to construct the drift terms in  $F^3$  and  $F^4$ . To do so a particular data generating process needed to be chosen. In order to simplify the presentation, the chosen data generating process was one for which the regressors are i.i.d. orthonormal and are of equal relevance to the

conditional mean function so that  $\beta_2^* = (1, 1, \dots, 1)'$ . After this simplification the limiting distributions under the local alternative can be rewritten as

$$F^3 = 2F^1 + \pi\lambda k_2 - 2 \sum_{i=1}^{k_2} [W_i(1) - W_i(\lambda)]$$

and

$$F^4 = (0.5)F^3 / [\chi_2 + \pi\lambda k_2 - 2 \sum_{i=1}^{k_2} \chi_{i,3}]^{0.5}$$

where  $W_i$  and  $\chi_{i,3}$  represent the  $i$ th components of  $W$  and  $\chi_3$  respectively.

Tables VII - IX report the percentage of 5000 draws of  $F^3$  and  $F^4$  that were greater than the relevant critical values reported in Tables I - VI. In each table the local power of the OOS-t and OOS-F is reported for a range of  $\pi = (0.2, 1.0, 2.0, 50.0)$ ,  $k_2 = (1, 2, 5, 10, 20)$ , and nominal size of the test takes the values (1%, 5%, 10%). For example under the recursive scheme with  $k_2 = 2$  and  $\pi = 1$ , 1400 of the 5000 draws of  $F^3$  were greater than 1.802 and hence at a nominal size of 5% the local power of the (modified) OOS-F statistic is 28%. Similarly, at a nominal size of 5% the local power of the OOS-t statistic is 12.53%. Table VII relates to the recursive scheme, VIII to the rolling, and IX to the fixed.

In each of the three tables and in panels A, B and C it is usually the case that when both  $k_2$  and  $\pi$  are smaller the OOS-F is more powerful than the OOS-t. As either  $k_2$  or  $\pi$  become sufficiently large the OOS-t becomes more powerful. Hence given a particular  $k_2$  or  $\pi$  pair, Tables VII – IX provide some guideline on the choice between the OOS-F and OOS-t statistics.

If we compare local power across Tables VII – IX we can also draw some conclusions on the choice of sampling scheme. Of the 60 possible ( $k_2$ ,  $\pi$ , nominal size) comparisons among the three sampling schemes, when the OOS-t is used the recursive scheme is most powerful 57 times and the rolling 3 times. When the OOS-F is used the recursive is most powerful 39 times and the fixed 21 times. In this latter case, the fixed scheme is most powerful only when both  $k_2$  and  $\pi$  are smaller. It therefore seems that when choosing a sampling scheme the recursive scheme should be the first choice unless both  $k_2$  or  $\pi$  are small in which case perhaps the fixed should be considered.

Deciding upon the optimal sample split parameter  $\pi$  is less clear. The sample split that maximizes the power of the test varies with the statistic, the sampling scheme and sometimes the number of excess parameters  $k_2$ . For the recursive scheme larger values of  $\pi$  ( $\pi = 50.0$ ) are best when the OOS-t is used. For the OOS-F the optimal split is usually small ( $\pi = 0.2$ ) when  $k_2$  is small and moderate ( $\pi = 1.0$ ) when  $k_2$  is larger. For both the fixed and rolling schemes smaller values of  $\pi$  ( $\pi = 0.2$ ) are best when the OOS-F is used. These two schemes differ when using the OOS-t statistic. When using the fixed scheme power is highest using moderate sample splits ( $\pi = 1.0$ ) but when using the rolling scheme power is highest at slightly higher levels ( $\pi = 2.0$ ).

## 5. CONCLUSION

This paper presents the null limiting distributions of three statistics commonly used to test for equal predictive ability between two nested models. The limiting null distributions of these three statistics are non-standard. Numerically calculated critical values are provided so that asymptotically valid tests of equal predictive ability can be constructed.

The limiting distributions of these statistics are also presented under a particular sequence of local alternatives. Though limited, the results indicate that at smaller values of  $k_2$  and  $\pi$  the OOS-F is more powerful but as  $k_2$  and  $\pi$  increase the OOS-t becomes more powerful. The results also indicate that the recursive scheme is generally most powerful though the fixed scheme is most powerful in particular circumstances. The numerical results shed little light on the optimal choice of the sample split parameter  $\pi$ . There are situations where power seems to be monotone in  $\pi$ , but often times it is not. Steckel and VanHonacker (1993) note this type of nonlinear behavior.

Perhaps the most interesting results concern implications for the Principle of Parsimony and overfitting. In Section 3.5 it is shown that the probability of the MSE of an overparameterized model being larger than the MSE of a less parameterized model is increasing in both the number of excess parameters in the overparameterized model and the percentage of the sample used for OOS prediction. This simple result implies that the common in-sample application of the Principle of Parsimony is sometimes inappropriate when applied OOS. This result can also be interpreted as analytical evidence for why overfit models tend to (but do not always) exhibit poor OOS predictive ability (Diebold, 1998, p.47).

A number of questions still remain concerning the OOS predictive ability of nested models. As mentioned in Section 3.3 the assumptions eliminate the potential application of either series-based (Swanson, 1996), local-linear (Diebold and Nason, 1991) or kernel-based (McCracken, 1999b) nonparametric estimation of the regression function. Since these are increasingly prevalent methods of constructing forecasts it would be useful to develop tools that would allow the application of the OOS-t and OOS-F statistics when nonparametric forecasts are used. This may be of particular usefulness when one is interested in testing for market efficiency and hence is concerned with what Fama (1991) refers to as the joint-hypothesis problem.

Another potential extension is the development of OOS model selection criteria. The present paper only considers using OOS predictive ability as a means of choosing between two parametric models. In general there are situations when one wishes to choose from multiple models. Such is the case when the model is known to be autoregressive with unknown lag order. As discussed in Section 3.5 it is not clear that existing in-sample information criteria can be directly extended to the OOS environment.

Other extensions include the application to nondifferentiable loss functions such as MAE, the Linex and the Maximum Score. Furthermore, it would be useful to extend the results so that the predictive models are potentially misspecified. Finally since it is not always the case that the loss function used to estimate the parameters is identical to that used to measure predictive ability, it would be helpful to extend the results in Section three to allow for such a possibility.

## REFERENCES

- AKAIKE, HIROTUGU (1969): "Fitting Autoregressive Models for Prediction", *Annals of the Institute of Statistical Mathematics*, 21, 243-247.
- ANDREWS, DONALD W.K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 61, 821-856.
- ASHLEY, RICHARD (1981): "Inflation and the Distribution of Price Changes Across Markets: A Causal Analysis", *Economic Inquiry*, 19, 650-660.
- (1998): "A New Technique for Postsample Model Selection and Validation", *Journal of Economic Dynamics and Control*, 22, 647-665.
- , CLIVE W.J. GRANGER AND R. SCHMALENSEE (1980): "Advertising and Aggregate Consumption: An Analysis of Causality", *Econometrica*, 48, 1149-1167.
- BERKOWITZ, JEREMY AND LORENZO GIORGIANNI (1999): "Long-Horizon Exchange Rate Predictability", *Review of Economics and Statistics*, forthcoming.
- BREEN, WILLIAM, LAWRENCE R. GLOSTEN AND RAVI JAGANNATHAN (1989): "Economic Significance of Predictable Variations in Stock Index Returns", *The Journal of Finance*, 44, 1177-1189.
- CHINN, MENZIE D. AND RICHARD A. MEESE (1995): "Banking on Currency Forecasts: How Predictable is Change in Money?", *Journal of International Economics*, 38, 161-178.
- CHONG, YOCK Y. AND DAVID F. HENDRY (1986): "Econometric Evaluation of Linear Macro-Economic Models", *Review of Economic Studies*, 53, 671-690.
- CHUNG, Y. PETER AND ZHUNG GUO ZHOU (1996): "The Predictability of Stock Returns--A Nonparametric Approach", *Econometric Reviews*, 15, 299-330.
- CLARIDA, RICHARD H. AND MARK P. TAYLOR (1997): "The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors", *The Review of Economics and Statistics*, 79, 353-361.
- CLARK, TODD E. (1999): "Finite-Sample Properties of Tests for Equal Forecast Accuracy", *Journal of Forecasting*, forthcoming.
- AND MICHAEL W. MCCRACKEN (1999): "Tests of Equal Forecast Accuracy and Encompassing for Nested Models", manuscript, Federal Reserve Bank of Kansas City.
- CORRADI, V., N.R. SWANSON AND C. OLIVETTI (1999): "Predictive Ability with Cointegrated Variables", manuscript, Texas A & M University.
- DAVIDSON, RUSSELL AND JAMES G. MACKINNON (1987): "Implicit Alternatives and the Local Power of Test Statistics", *Econometrica*, 55, 1305-1329.
- DIEBOLD, FRANCIS X. (1998): *Elements of Forecasting*, (Cincinnati, South-Western College Publishing).
- , TODD A. GUNTHER AND ANTHONY S. TAY (1997): "Evaluating Density Forecasts", NBER Technical Working Paper #215.

- AND LUTZ KILIAN (1997): "Measuring Predictability: Theory and Macroeconomic Applications", NBER Technical Working Paper #213.
- AND ROBERT S. MARIANO (1995): "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, 253-263.
- AND JAMES NASON (1990): "Nonparametric Exchange Rate Prediction?", *Journal of International Economics*, 28, 315-322.
- FAIR, RAY C. AND ROBERT SHILLER (1989): "The Informational Content of Ex Ante Forecasts", *Review of Economics and Statistics*, 71, 325-331.
- AND ----- (1990): "Comparing Information in Forecasts from Econometric Models", *American Economic Review*, 80, 375-389.
- FAMA, EUGENE F. (1991): "Efficient Capital Markets: II", *The Journal of Finance*, 46, 1575-1617.
- GRANGER, CLIVE W.J. (1995): "Where are the Controversies in Econometric Methodology?", in C.W.J. Granger, ed., *Modeling Economic Series*, (Oxford University Press, New York).
- AND PAUL NEWBOLD (1977): *Forecasting Economic Time Series*, (London, Academic Press Inc.).
- HALL, PETER (1992): *The Bootstrap and Edgeworth Expansion*, (New York, Springer-Verlag).
- HANSEN, BRUCE E. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes", *Econometric Theory*, 8, 489-500.
- HARVEY, DAVID I., STEPHEN J. LEYBOURNE AND PAUL NEWBOLD (1998): "Forecast Evaluation Tests in the Presence of ARCH", manuscript, Loughborough University and University of Nottingham.
- HOFFMAN, DENNIS AND ADRIAN PAGAN (1989): "Practitioners Corner: Post-Sample Prediction Tests for Generalized Method of Moments Estimators", *Oxford Bulletin of Economics and Statistics*, 51, 333-343.
- HONG, YONGMIAO AND HALBERT WHITE (1995): "Consistent Specification Testing Via Nonparametric Series Regression", *Econometrica*, 63, 1133-1159.
- HUH, CHAN (1996): "Some Evidence on the Efficacy of the UK Inflation Targeting Regime: An Out of Sample Forecast Approach", Federal Reserve Board of Governors, International Finance Discussion Paper #565.
- KILIAN, LUTZ (1999): "Exchange Rates and Monetary Fundamentals: What Do We Learn From Long-Horizon Regressions?", *Journal of Applied Econometrics*, Forthcoming.
- KUAN, CHUNG-MING AND TUNG LIU (1995): "Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks", *Journal of Applied Econometrics*, 10, 347-364.
- LEITCH, GORDON AND J. ERNEST TANNER (1991): "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures", *American Economic Review*, 81, 580-590.
- MAGNUS, J. AND H. NEUDECKER (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, (New York, Wiley).
- MARK, NELSON C. (1995): "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability", *The American Economic Review*, 85, 201-218.
- AND DONGGYU SUL (1998): "Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Seventeen Country Panel", manuscript, Ohio State University.

- MCCRACKEN, MICHAEL W. (1998): "Data Mining and Out-of-Sample Inference", manuscript, Louisiana State University.
- (1999a): "Robust Out of Sample Inference", manuscript, Louisiana State University.
- (1999b): "An Out-of-Sample, Nonparametric Test of the Martingale Difference Hypothesis", manuscript, Louisiana State University.
- MEESE, RICHARD A. AND KENNETH ROGOFF (1983): "Empirical Exchange Rate Models of the Seventies: Do they Fit Out of Sample?", *Journal of International Economics*, 14, 3-24.
- AND ----- (1988): "Was It Real? The Exchange Rate-Interest Differential Relation Over the Modern Floating-Rate Period", *Journal of Finance*, 43, 933-948.
- MINCER, JACOB AND VICTOR ZARNOWITZ (1969): "The Evaluation of Economic Forecasts," in J. Mincer, ed., *Economic Forecasts and Expectations*, (New York, National Bureau of Economic Research).
- MIZRACH, BRUCE (1992): "The Distribution of the Theil U-Statistic in Bivariate Normal Populations", *Economic Letters*, 38, 163-167.
- MORGAN, W. A. (1939): "A Test for Significance of the Difference Between Two Variances in a Sample from a Normal Bivariate Population", *Biometrika*, 31, 13-19.
- NELSON, FORREST D. AND N.E. SAVIN (1990): "The Danger of Extrapolating Asymptotic Local Power", *Econometrica*, 58, 977-981.
- NEWY, WHITNEY K. AND DANIEL MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing", in R.F. Engle and D.L. McFadden, ed., *Handbook of Econometrics, Volume IV*, (Amsterdam, North-Holland).
- PAGAN, ADRIAN R. AND ANTHONY D. HALL (1983): "Diagnostic Tests as Residual Analysis", *Econometric Reviews*, 2, 159-218.
- AND G. WILLIAM SCHWERT (1990): "Alternative Models for Conditional Stock Volatility", *Journal of Econometrics*, 45, 267-290.
- PARK, TIMOTHY (1990): "Forecast Evaluation for Multivariate Time-Series Models: The U.S. Cattle Market", *Western Journal of Agricultural Economics*, July, 133-143.
- PESARAN, M. HASHEM AND ALLAN TIMMERMANN (1995): "Predictability of Stock Returns: Robustness and Economic Significance", *The Journal of Finance*, 50, 4, 1201-1228.
- AND ----- (1999): "Model Instability and Choice of Observation Window", manuscript, University of California-San Diego Working Paper #99-19.
- RANGLES, RONALD H. (1982): "On the Asymptotic Normality of Statistics with Estimated Parameters", *The Annals of Statistics*, 10, 463-474.
- SANCHEZ, ISMAEL (1998): "Testing for Unit Roots with Prediction Errors", manuscript, University of California San Diego.
- STECKEL, JOEL AND WILFRIED VANHONACKER (1993): "Cross-Validating Regression Models in Market Research", *Marketing Science*, 12 415-427.
- SULLIVAN, RYAN, ALLAN TIMMERMANN AND HALBERT WHITE (1998): "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap", *Journal of Finance*, forthcoming.

- SWANSON, NORMAN R. (1996): "Forecasting Using First-Available versus Fully Revised Economic Time-Series Data", *Studies in Nonlinear Dynamics and Econometrics*, 1, 47-64.
- SWANSON, NORMAN R. (1998): "Money and Output Viewed Through a Rolling Window", *Journal of Monetary Economics*, 41, 455-473.
- , ATAMAN OZYILDIRIM AND MARIA PISU (1996): "A Comparison of Alternative Causality and Predictive Accuracy Tests in the Presence of Integrated and Co-Integrated Economic Variables", manuscript, Pennsylvania State University.
- AND HALBERT WHITE (1995): "A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks", *Journal of Business and Economic Statistics*, 13, 265-275.
- AND HALBERT WHITE (1997a): "A Model-Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks", *The Review of Economics and Statistics*, 79, 265-275.
- AND HALBERT WHITE (1997b): "Forecasting Economic Time Series Using Flexible Versus Fixed Specification and Linear Versus Nonlinear Econometric Models", *International Journal of Forecasting*, 13, 439-461.
- TEGENE, ABEBAYEHU AND FRED KUCHLER (1994): "Evaluating Forecasting Models of Farmland Prices", *International Journal of Forecasting*, 10, 65-80.
- THEIL, H. (1966): *Applied Economic Forecasting*, (Amsterdam, North-Holland).
- URBAIN, J.P. (1989): "Model Selection Criteria and Granger Causality Tests: An Empirical Note", *Economics Letters*, 29, 317-320.
- WEISS, ANDREW A. (1996): "Estimating Time Series Models Using the Relevant Cost Function", *Journal of Applied Econometrics*, 11, 539-560.
- AND A.P. ANDERSEN (1984): "Estimating Time Series Using the Relevant Forecast Evaluation Criterion", *Journal of the Royal Statistical Society, Series A*, 147, 484-487.
- WEST, KENNETH D. (1996): "Asymptotic Inference About Predictive Ability", *Econometrica*, 64, 1067-1084.
- AND MICHAEL W. MCCRACKEN (1998): "Regression-Based Tests of Predictive Ability", *International Economic Review*, 39, 817-840
- WHITE, HALBERT (1999): "A Reality Check for Data Snooping", *Econometrica*, forthcoming.
- WOLFF, CHRISTIAN C.P. (1987): "Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models", *Journal of Business and Economic Statistics*, 5, 87-97.
- WOOLDRIDGE, JEFFREY M. AND HALBERT WHITE (1989): "Central Limit Theorems for Dependent, Heterogeneous Processes with Trending Moments", manuscript, Michigan State University.

## FOOTNOTES

<sup>1</sup> I'd like to thank Todd Clark, Walter Enders, Bruce Hansen, Dek Terrell, Ken West, and seminar participants at LSU and the 1999 Midwest Economic Association meetings for helpful comments.

<sup>2</sup> See Randles (1982) for the in-sample analog.

<sup>3</sup> Notice that the fixed and rolling parameter estimates should be subscripted both by  $t$  and  $R$ . In order to simplify the notation the subscript  $R$  will be suppressed.

<sup>4</sup> See the discussion in Section 3.4.

<sup>5</sup> I also could have estimated  $\Omega$  using squared deviations from the sample mean. Doing so is asymptotically irrelevant and hence is omitted for notational convenience.

<sup>6</sup> These values of  $\pi$  correspond to percentages of in-sample observations  $\lambda = \{0.909, 0.833, 0.714, 0.625, 0.555, 0.500, 0.454, 0.417, 0.385, 0.357, 0.333\}$ .

<sup>7</sup> This estimator is consistent by Theorem 4.1 of West (1996).

<sup>8</sup> This does not mean that the test should be lower tailed. It means that the asymptotic median of the difference in MSE's is now negative.

TABLE I  
PERCENTILES OF THE OOS-t STATISTIC: RECURSIVE SCHEME

$k_2 \setminus \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	1.921	1.784	1.625	1.515	1.462	1.436	1.413	1.343	1.316	1.274	1.238
	(0.95)	1.245	1.111	0.994	0.971	0.863	0.771	0.740	0.705	0.671	0.638	0.610
	(0.90)	0.885	0.780	0.657	0.598	0.512	0.443	0.402	0.370	0.330	0.306	0.281
2	(0.99)	1.986	1.856	1.563	1.436	1.387	1.312	1.276	1.196	1.158	1.127	1.074
	(0.95)	1.274	1.140	0.986	0.868	0.782	0.704	0.623	0.596	0.537	0.507	0.478
	(0.90)	0.932	0.786	0.614	0.541	0.455	0.361	0.295	0.253	0.235	0.194	0.160
3	(0.99)	1.840	1.737	1.542	1.448	1.359	1.252	1.148	1.071	0.976	0.978	0.953
	(0.95)	1.300	1.120	0.968	0.808	0.685	0.610	0.552	0.496	0.438	0.419	0.386
	(0.90)	0.939	0.751	0.551	0.454	0.356	0.279	0.222	0.175	0.108	0.074	0.035
4	(0.99)	1.872	1.731	1.581	1.365	1.195	1.119	1.108	1.041	0.902	0.861	0.854
	(0.95)	1.264	1.101	0.914	0.772	0.609	0.502	0.419	0.345	0.285	0.239	0.221
	(0.90)	0.898	0.742	0.562	0.419	0.263	0.169	0.094	0.052	-0.014	-0.054	-0.106
5	(0.99)	1.849	1.679	1.468	1.242	1.095	0.995	0.979	0.913	0.795	0.732	0.677
	(0.95)	1.222	1.061	0.849	0.689	0.491	0.386	0.308	0.224	0.148	0.107	0.081
	(0.90)	0.866	0.694	0.461	0.315	0.179	0.062	-0.021	-0.083	-0.145	-0.174	-0.228
6	(0.99)	1.836	1.639	1.390	1.200	1.042	0.943	0.859	0.755	0.686	0.610	0.593
	(0.95)	1.192	0.998	0.768	0.615	0.429	0.328	0.259	0.141	0.078	0.055	-0.019
	(0.90)	0.823	0.642	0.394	0.256	0.108	-0.011	-0.101	-0.164	-0.218	-0.266	-0.319
7	(0.99)	1.836	1.649	1.341	1.154	0.994	0.872	0.810	0.637	0.549	0.476	0.438
	(0.95)	1.199	0.976	0.742	0.546	0.372	0.279	0.191	0.072	-0.002	-0.034	-0.105
	(0.90)	0.811	0.615	0.359	0.213	0.062	-0.088	-0.152	-0.230	-0.305	-0.363	-0.449
8	(0.99)	1.789	1.659	1.298	1.090	0.879	0.788	0.728	0.503	0.444	0.401	0.359
	(0.95)	1.193	0.928	0.677	0.462	0.302	0.198	0.105	0.020	-0.058	-0.101	-0.176
	(0.90)	0.773	0.574	0.329	0.139	0.003	-0.131	-0.203	-0.293	-0.383	-0.452	-0.516
9	(0.99)	1.813	1.607	1.268	1.112	0.804	0.724	0.634	0.523	0.427	0.391	0.305
	(0.95)	1.112	0.912	0.617	0.397	0.276	0.121	0.030	-0.055	-0.122	-0.193	-0.257
	(0.90)	0.733	0.561	0.273	0.096	-0.068	-0.187	-0.286	-0.377	-0.437	-0.518	-0.579
10	(0.99)	1.743	1.534	1.193	1.035	0.758	0.621	0.506	0.419	0.347	0.285	0.185
	(0.95)	1.082	0.890	0.566	0.358	0.205	0.043	-0.072	-0.162	-0.222	-0.296	-0.339
	(0.90)	0.749	0.529	0.226	0.032	-0.130	-0.248	-0.355	-0.454	-0.524	-0.591	-0.651

TABLE II  
PERCENTILES OF THE OOS-t STATISTIC: ROLLING SCHEME

$k_2 \setminus \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	1.875	1.799	1.604	1.447	1.340	1.221	1.179	1.098	1.021	0.969	0.882
	(0.95)	1.251	1.117	0.970	0.859	0.722	0.651	0.575	0.510	0.455	0.382	0.334
	(0.90)	0.903	0.776	0.637	0.530	0.401	0.317	0.246	0.180	0.136	0.116	0.078
2	(0.99)	1.959	1.757	1.504	1.325	1.180	1.165	0.996	0.953	0.883	0.744	0.640
	(0.95)	1.280	1.105	0.884	0.753	0.631	0.484	0.401	0.304	0.235	0.166	0.103
	(0.90)	0.915	0.755	0.569	0.425	0.280	0.155	0.111	0.026	-0.050	-0.094	-0.140
3	(0.99)	1.860	1.669	1.473	1.271	1.076	0.984	0.896	0.773	0.614	0.504	0.431
	(0.95)	1.274	1.088	0.842	0.667	0.490	0.381	0.251	0.146	0.066	-0.016	-0.084
	(0.90)	0.938	0.718	0.521	0.346	0.201	0.064	-0.042	-0.137	-0.224	-0.302	-0.346
4	(0.99)	1.905	1.700	1.503	1.183	1.003	0.903	0.755	0.656	0.455	0.342	0.234
	(0.95)	1.267	1.087	0.852	0.585	0.376	0.274	0.136	0.024	-0.080	-0.173	-0.222
	(0.90)	0.866	0.731	0.494	0.248	0.098	-0.047	-0.164	-0.262	-0.362	-0.434	-0.505
5	(0.99)	1.881	1.627	1.347	1.112	0.927	0.790	0.657	0.504	0.307	0.193	0.123
	(0.95)	1.229	1.034	0.716	0.479	0.280	0.155	-0.019	-0.090	-0.219	-0.329	-0.385
	(0.90)	0.825	0.694	0.402	0.154	-0.025	-0.168	-0.305	-0.399	-0.508	-0.589	-0.674
6	(0.99)	1.826	1.680	1.312	1.007	0.850	0.641	0.558	0.336	0.195	0.069	0.017
	(0.95)	1.176	0.966	0.621	0.407	0.225	0.058	-0.119	-0.218	-0.336	-0.428	-0.535
	(0.90)	0.811	0.602	0.319	0.088	-0.095	-0.262	-0.423	-0.523	-0.638	-0.732	-0.821
7	(0.99)	1.842	1.620	1.233	0.989	0.751	0.526	0.485	0.227	0.055	-0.039	-0.127
	(0.95)	1.154	0.936	0.628	0.346	0.171	-0.011	-0.182	-0.320	-0.433	-0.531	-0.663
	(0.90)	0.791	0.573	0.279	0.038	-0.157	-0.326	-0.497	-0.611	-0.750	-0.841	-0.933
8	(0.99)	1.819	1.582	1.178	0.918	0.702	0.466	0.349	0.132	-0.018	-0.176	-0.302
	(0.95)	1.157	0.924	0.562	0.258	0.081	-0.099	-0.281	-0.432	-0.552	-0.672	-0.785
	(0.90)	0.758	0.541	0.244	-0.042	-0.244	-0.408	-0.576	-0.727	-0.838	-0.957	-1.040
9	(0.99)	1.768	1.510	1.110	0.845	0.600	0.408	0.235	0.036	-0.099	-0.277	-0.407
	(0.95)	1.117	0.892	0.504	0.213	0.021	-0.156	-0.374	-0.529	-0.623	-0.785	-0.885
	(0.90)	0.742	0.520	0.193	-0.105	-0.322	-0.491	-0.657	-0.803	-0.951	-1.049	-1.153
10	(0.99)	1.713	1.428	1.075	0.808	0.536	0.298	0.122	-0.064	-0.248	-0.381	-0.482
	(0.95)	1.068	0.872	0.443	0.133	-0.038	-0.258	-0.466	-0.605	-0.765	-0.909	-1.011
	(0.90)	0.727	0.500	0.138	-0.144	-0.374	-0.568	-0.757	-0.902	-1.045	-1.167	-1.288

TABLE III  
PERCENTILES OF THE OOS-t STATISTIC: FIXED SCHEME

$k_2 \setminus \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	2.201	2.051	1.974	2.061	2.037	2.024	1.992	2.018	1.996	2.016	1.993
	(0.95)	1.506	1.416	1.364	1.428	1.346	1.252	1.301	1.293	1.249	1.235	1.218
	(0.90)	1.149	1.079	1.042	1.040	0.976	0.917	0.896	0.893	0.908	0.834	0.862
2	(0.99)	2.145	2.089	1.923	1.947	1.964	1.749	1.751	1.665	1.725	1.646	1.613
	(0.95)	1.468	1.342	1.301	1.265	1.164	1.072	1.034	1.046	0.977	0.982	0.955
	(0.90)	1.096	0.999	0.901	0.873	0.798	0.711	0.680	0.639	0.578	0.556	0.520
3	(0.99)	2.045	1.977	1.957	1.805	1.739	1.602	1.520	1.597	1.463	1.513	1.407
	(0.95)	1.432	1.277	1.195	1.095	1.014	0.909	0.893	0.851	0.761	0.735	0.733
	(0.90)	1.063	0.922	0.793	0.705	0.621	0.540	0.511	0.455	0.386	0.373	0.306
4	(0.99)	2.013	1.883	1.829	1.687	1.528	1.467	1.475	1.422	1.318	1.255	1.277
	(0.95)	1.369	1.281	1.110	0.997	0.883	0.755	0.689	0.650	0.607	0.566	0.509
	(0.90)	1.004	0.895	0.764	0.575	0.476	0.367	0.340	0.273	0.204	0.171	0.081
5	(0.99)	1.930	1.878	1.716	1.596	1.405	1.254	1.301	1.230	1.171	1.115	1.034
	(0.95)	1.333	1.193	1.009	0.863	0.725	0.646	0.570	0.486	0.410	0.365	0.291
	(0.90)	0.945	0.838	0.636	0.487	0.374	0.258	0.193	0.115	0.020	-0.022	-0.085
6	(0.99)	1.933	1.874	1.628	1.481	1.382	1.146	1.188	1.091	1.016	1.007	0.878
	(0.95)	1.269	1.122	0.936	0.771	0.652	0.538	0.487	0.367	0.314	0.222	0.152
	(0.90)	0.912	0.764	0.552	0.400	0.299	0.169	0.103	0.003	-0.106	-0.146	-0.235
7	(0.99)	1.925	1.859	1.556	1.377	1.257	1.105	1.103	0.987	0.896	0.828	0.765
	(0.95)	1.263	1.086	0.878	0.692	0.557	0.446	0.346	0.254	0.191	0.074	0.014
	(0.90)	0.895	0.731	0.513	0.332	0.215	0.060	-0.003	-0.147	-0.252	-0.308	-0.386
8	(0.99)	1.856	1.827	1.467	1.245	1.146	1.029	0.980	0.860	0.786	0.762	0.666
	(0.95)	1.249	1.064	0.807	0.623	0.481	0.363	0.268	0.151	0.054	-0.042	-0.120
	(0.90)	0.868	0.663	0.467	0.247	0.153	-0.029	-0.115	-0.227	-0.343	-0.440	-0.502
9	(0.99)	1.878	1.697	1.440	1.198	1.124	0.902	0.791	0.683	0.644	0.595	0.507
	(0.95)	1.197	1.031	0.754	0.537	0.416	0.305	0.162	0.050	-0.067	-0.171	-0.242
	(0.90)	0.844	0.655	0.396	0.182	0.034	-0.111	-0.224	-0.303	-0.437	-0.543	-0.625
10	(0.99)	1.824	1.604	1.354	1.126	0.998	0.797	0.659	0.557	0.550	0.505	0.415
	(0.95)	1.143	1.007	0.688	0.455	0.337	0.167	0.040	-0.057	-0.174	-0.246	-0.358
	(0.90)	0.797	0.616	0.348	0.125	-0.055	-0.210	-0.305	-0.398	-0.559	-0.645	-0.729

TABLE IV  
PERCENTILES OF THE (MODIFIED) OOS-F STATISTIC: RECURSIVE SCHEME

$k_2 \setminus \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	1.608	2.129	2.768	3.179	3.459	3.584	3.771	3.589	3.838	3.882	3.951
	(0.95)	0.850	1.038	1.298	1.554	1.567	1.548	1.583	1.623	1.599	1.553	1.518
	(0.90)	0.530	0.659	0.814	0.796	0.798	0.751	0.759	0.698	0.685	0.687	0.616
2	(0.99)	1.996	2.691	3.426	3.907	4.129	4.200	4.362	4.304	4.309	4.278	4.250
	(0.95)	1.184	1.453	1.733	1.891	1.820	1.802	1.819	1.752	1.734	1.692	1.706
	(0.90)	0.794	0.912	1.029	1.077	1.008	0.880	0.785	0.697	0.666	0.587	0.506
3	(0.99)	2.418	3.092	4.080	4.136	4.322	4.341	4.337	4.192	4.089	4.365	4.184
	(0.95)	1.434	1.710	2.062	2.073	1.978	1.909	1.930	1.795	1.715	1.710	1.612
	(0.90)	0.970	1.064	1.117	1.121	0.960	0.857	0.691	0.599	0.386	0.276	0.127
4	(0.99)	2.714	3.440	4.541	4.609	4.378	4.202	4.586	4.477	4.337	4.247	4.096
	(0.95)	1.566	1.964	2.246	2.194	1.900	1.809	1.578	1.376	1.256	1.122	1.029
	(0.90)	1.060	1.225	1.313	1.184	0.829	0.545	0.354	0.197	-0.058	-0.234	-0.456
5	(0.99)	2.902	3.673	4.466	4.434	4.249	4.351	4.349	4.187	3.945	3.783	3.783
	(0.95)	1.688	2.082	2.235	2.242	1.773	1.449	1.316	1.045	0.718	0.502	0.459
	(0.90)	1.130	1.277	1.228	0.958	0.614	0.241	-0.099	-0.361	-0.656	-0.820	-1.072
6	(0.99)	3.212	3.846	4.545	4.676	4.637	4.703	4.286	4.144	3.981	3.525	3.321
	(0.95)	1.828	2.124	2.217	2.121	1.660	1.360	1.181	0.761	0.413	0.299	-0.109
	(0.90)	1.220	1.313	1.164	0.890	0.419	-0.044	-0.405	-0.776	-1.072	-1.395	-1.664
7	(0.99)	3.450	4.098	4.508	4.419	4.271	4.312	4.150	3.677	3.155	3.090	2.880
	(0.95)	2.000	2.239	2.424	2.057	1.604	1.282	0.928	0.378	-0.008	-0.199	-0.591
	(0.90)	1.272	1.333	1.118	0.799	0.242	-0.363	-0.728	-1.194	-1.657	-2.033	-2.507
8	(0.99)	3.408	4.130	4.645	4.625	4.202	4.147	3.912	3.185	2.933	2.952	2.484
	(0.95)	2.136	2.312	2.373	1.895	1.390	0.943	0.587	0.131	-0.372	-0.680	-1.14
	(0.90)	1.338	1.369	1.058	0.552	0.014	-0.632	-1.076	-1.633	-2.174	-2.731	-3.16
9	(0.99)	3.540	4.388	4.703	4.873	4.122	4.066	3.753	3.027	2.925	2.802	2.186
	(0.95)	2.168	2.440	2.219	1.714	1.286	0.631	0.198	-0.356	-0.851	-1.241	-1.696
	(0.90)	1.354	1.432	0.920	0.393	-0.327	-1.007	-1.595	-2.229	-2.666	-3.250	-3.794
10	(0.99)	3.646	4.433	4.813	4.718	3.944	3.645	3.194	2.578	2.282	2.152	1.436
	(0.95)	2.202	2.489	2.157	1.536	1.055	0.205	-0.431	-1.071	-1.459	-1.988	-2.378
	(0.90)	1.458	1.401	0.884	0.155	-0.600	-1.341	-2.008	-2.782	-3.348	-3.839	-4.437

TABLE V  
 PERCENTILES OF THE (MODIFIED) OOS-F STATISTIC: ROLLING SCHEME

$k_2 \setminus \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	1.638	2.230	2.812	3.300	3.634	3.811	3.688	3.721	3.924	3.612	3.765
	(0.95)	0.854	1.112	1.394	1.644	1.627	1.583	1.574	1.469	1.488	1.378	1.215
	(0.90)	0.536	0.667	0.838	0.865	0.773	0.693	0.602	0.482	0.390	0.355	0.276
2	(0.99)	2.036	2.819	3.544	3.988	4.066	4.398	4.403	4.109	4.293	4.046	3.566
	(0.95)	1.232	1.481	1.802	1.889	1.841	1.695	1.495	1.264	1.015	0.783	0.504
	(0.90)	0.812	0.920	1.028	1.004	0.806	0.468	0.399	0.095	-0.198	-0.394	-0.623
3	(0.99)	2.476	3.128	4.135	4.120	4.264	4.519	4.386	4.123	3.373	3.089	2.685
	(0.95)	1.472	1.752	2.089	2.042	1.700	1.532	1.100	0.694	0.340	-0.071	-0.471
	(0.90)	1.006	1.074	1.135	0.944	0.617	0.224	-0.174	-0.600	-1.080	-1.529	-1.847
4	(0.99)	2.724	3.649	4.474	4.586	4.432	4.459	4.296	3.621	2.905	2.337	1.699
	(0.95)	1.600	2.078	2.332	1.979	1.536	1.228	0.701	0.116	-0.491	-1.112	-1.487
	(0.90)	1.096	1.284	1.263	0.777	0.356	-0.200	-0.788	-1.341	-1.973	-2.528	-3.182
5	(0.99)	3.008	3.721	4.504	4.710	4.508	4.199	4.042	3.216	2.167	1.370	1.055
	(0.95)	1.768	2.191	2.164	1.783	1.175	0.764	-0.121	-0.542	-1.454	-2.172	-2.765
	(0.90)	1.152	1.315	1.143	0.541	-0.114	-0.774	-1.583	-2.300	-3.102	-3.896	-4.649
6	(0.99)	3.214	4.093	4.532	4.786	4.456	3.899	3.473	2.324	1.500	0.615	0.159
	(0.95)	1.926	2.181	2.117	1.652	1.062	0.318	-0.745	-1.424	-2.302	-3.162	-4.256
	(0.90)	1.224	1.311	0.981	0.365	-0.428	-1.348	-2.392	-3.270	-4.292	-5.180	-6.114
7	(0.99)	3.552	4.340	4.568	4.566	4.181	3.450	3.167	1.696	0.442	-0.300	-1.151
	(0.95)	1.994	2.273	2.321	1.478	0.868	-0.076	-1.162	-2.243	-3.224	-4.261	-5.62
	(0.90)	1.292	1.355	0.939	0.151	-0.783	-1.833	-3.061	-4.153	-5.487	-6.474	-7.583
8	(0.99)	3.480	4.537	4.707	4.681	4.041	3.065	2.488	1.013	-0.144	-1.562	-2.729
	(0.95)	2.080	2.436	2.169	1.213	0.468	-0.626	-1.885	-3.233	-4.430	-5.681	-6.991
	(0.90)	1.386	1.357	0.886	-0.186	-1.278	-2.492	-3.824	-5.225	-6.556	-7.690	-8.939
9	(0.99)	3.552	4.438	4.711	4.443	3.754	2.622	1.723	0.327	-0.877	-2.543	-3.923
	(0.95)	2.164	2.518	1.966	1.075	0.138	-1.113	-2.620	-4.036	-5.258	-6.931	-8.345
	(0.90)	1.378	1.433	0.772	-0.521	-1.835	-3.139	-4.586	-6.133	-7.734	-9.173	-10.558
10	(0.99)	3.728	4.413	4.815	4.589	3.460	2.145	1.121	-0.624	-2.313	-3.795	-5.166
	(0.95)	2.224	2.520	1.893	0.730	-0.235	-1.733	-3.496	-4.940	-6.512	-8.255	-9.863
	(0.90)	1.456	1.411	0.555	-0.701	-2.189	-3.790	-5.566	-7.200	-9.046	-10.574	-12.294

TABLE VI  
 PERCENTILES OF THE (MODIFIED) OOS-F STATISTIC: FIXED SCHEME

$k_2 \setminus \pi$		0.1	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	1.480	1.981	2.681	3.055	3.230	3.377	3.562	3.619	3.816	3.812	3.838
	(0.95)	0.784	1.015	1.345	1.534	1.677	1.667	1.738	1.807	1.812	1.857	1.862
	(0.90)	0.514	0.649	0.835	0.885	0.933	0.964	1.009	0.986	1.050	1.080	1.037
2	(0.99)	1.840	2.554	3.241	3.514	3.944	4.019	4.173	4.364	4.251	4.556	4.414
	(0.95)	1.132	1.421	1.765	1.999	2.077	2.116	2.169	2.232	2.275	2.260	2.195
	(0.90)	0.784	0.914	1.140	1.237	1.299	1.268	1.330	1.250	1.126	1.189	1.151
3	(0.99)	2.324	2.985	3.854	4.103	4.272	4.233	4.549	4.764	4.687	4.915	4.900
	(0.95)	1.408	1.653	2.050	2.322	2.308	2.319	2.325	2.336	2.187	2.283	2.275
	(0.90)	1.000	1.106	1.328	1.367	1.375	1.291	1.289	1.225	1.073	1.043	0.954
4	(0.99)	2.576	3.283	3.999	4.339	4.349	4.629	4.602	5.002	4.793	4.984	5.028
	(0.95)	1.536	1.947	2.374	2.478	2.426	2.238	2.310	2.175	2.063	1.891	1.784
	(0.90)	1.100	1.317	1.472	1.362	1.195	1.109	1.008	0.860	0.667	0.544	0.289
5	(0.99)	2.748	3.437	4.212	4.454	4.330	4.396	4.739	5.044	4.761	4.731	4.560
	(0.95)	1.664	2.018	2.368	2.504	2.337	2.167	2.109	1.862	1.593	1.540	1.249
	(0.90)	1.178	1.387	1.414	1.341	1.038	0.865	0.696	0.445	0.083	-0.075	-0.347
6	(0.99)	3.084	3.755	4.467	4.754	4.559	4.715	4.836	4.515	4.561	4.303	4.365
	(0.95)	1.826	2.164	2.406	2.422	2.267	2.010	1.995	1.654	1.302	1.107	0.744
	(0.90)	1.242	1.417	1.428	1.167	0.962	0.634	0.410	0.014	-0.449	-0.666	-1.113
7	(0.99)	3.294	3.980	4.599	4.683	4.704	5.000	4.828	4.667	4.489	4.367	4.155
	(0.95)	1.962	2.282	2.441	2.410	2.198	1.886	1.535	1.263	0.943	0.356	0.071
	(0.90)	1.342	1.536	1.457	1.057	0.817	0.254	-0.015	-0.668	-1.314	-1.610	-2.23
8	(0.99)	3.364	4.116	4.775	4.724	4.715	4.762	4.480	4.111	4.278	4.482	3.804
	(0.95)	2.078	2.394	2.580	2.244	2.007	1.666	1.266	0.744	0.293	-0.244	-0.658
	(0.90)	1.434	1.501	1.374	0.900	0.622	-0.146	-0.593	-1.157	-1.925	-2.678	-3.109
9	(0.99)	3.430	4.233	4.671	4.640	4.856	4.580	4.112	3.756	3.536	3.648	3.158
	(0.95)	2.090	2.525	2.533	2.064	1.881	1.434	0.892	0.292	-0.344	-0.960	-1.536
	(0.90)	1.474	1.564	1.329	0.727	0.181	-0.578	-1.168	-1.751	-2.653	-3.457	-4.122
10	(0.99)	3.582	4.232	4.750	4.674	4.489	4.251	3.643	3.467	3.373	3.342	3.036
	(0.95)	2.158	2.611	2.481	1.967	1.601	0.936	0.282	-0.360	-1.068	-1.467	-2.404
	(0.90)	1.504	1.583	1.176	0.520	-0.246	-1.125	-1.722	-2.449	-3.606	-4.312	-5.252

TABLE VII  
LOCAL POWER OF THE (MODIFIED) OOS-F AND OOS-t STATISTICS: RECURSIVE SCHEME

A. Nominal Size of 1%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.1292	0.1664	0.2125	0.2446	0.2829
OOS-t	0.2	0.0241	0.0258	0.0405	0.0555	0.0835
OOS-F	1	0.1216	0.1570	0.2207	0.2619	0.3110
OOS-t	1	0.0316	0.0429	0.0847	0.1466	0.2866
OOS-F	2	0.1239	0.1531	0.1929	0.2246	0.2604
OOS-t	2	0.0381	0.0521	0.0990	0.1860	0.3564
OOS-F	50	0.0837	0.0821	0.0590	0.0287	0.0113
OOS-t	50	0.0507	0.0857	0.1578	0.2509	0.4295

B. Nominal Size of 5%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.2266	0.2680	0.3072	0.3321	0.3788
OOS-t	0.2	0.0920	0.1018	0.1275	0.1587	0.2375
OOS-F	1	0.2182	0.2660	0.3201	0.3505	0.3989
OOS-t	1	0.1157	0.1484	0.2379	0.3329	0.5049
OOS-F	2	0.2168	0.2489	0.2870	0.3095	0.3491
OOS-t	2	0.1300	0.1658	0.2634	0.3848	0.5852
OOS-F	50	0.1386	0.1277	0.0929	0.0505	0.0218
OOS-t	50	0.1538	0.2096	0.3249	0.4649	0.6587

C. Nominal Size of 10%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.2850	0.3255	0.3625	0.3886	0.4306
OOS-t	0.2	0.1622	0.1760	0.2167	0.2665	0.3649
OOS-F	1	0.2772	0.3217	0.3716	0.4001	0.4457
OOS-t	1	0.1973	0.2478	0.3531	0.4581	0.6255
OOS-F	2	0.2682	0.3063	0.3390	0.3600	0.3982
OOS-t	2	0.2167	0.2674	0.3878	0.5230	0.6991
OOS-F	50	0.1739	0.1591	0.1118	0.0639	0.0293
OOS-t	50	0.2429	0.3064	0.4399	0.5883	0.7648

Notes: Each element of Panels A, B and C is the local power of either the OOS-t or (modified) OOS-F test for a given permutation of the choice of sample split  $\pi$ , number of excess parameters in the unrestricted model  $k_2$ , and the nominal size of the test. For a description of how the results were generated see Section four of the text.

TABLE VIII  
LOCAL POWER OF THE (MODIFIED) OOS-F AND OOS-t STATISTICS: ROLLING SCHEME

A. Nominal Size of 1%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.1249	0.1566	0.2009	0.2345	0.2626
OOS-t	0.2	0.0222	0.0277	0.0414	0.0584	0.0776
OOS-F	1	0.1012	0.1294	0.1534	0.1740	0.1795
OOS-t	1	0.0288	0.0356	0.0702	0.1277	0.2365
OOS-F	2	0.0985	0.1059	0.1048	0.0950	0.0697
OOS-t	2	0.0325	0.0495	0.0884	0.1578	0.2690
OOS-F	50	0.0001	0.0000	0.0000	0.0000	0.0000
OOS-t	50	0.0162	0.0171	0.0216	0.0280	0.0459

B. Nominal Size of 5%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.2165	0.2565	0.2941	0.3182	0.3573
OOS-t	0.2	0.0886	0.1010	0.1226	0.1549	0.2298
OOS-F	1	0.1862	0.2183	0.2434	0.2504	0.2575
OOS-t	1	0.1011	0.1407	0.2095	0.2992	0.4622
OOS-F	2	0.1706	0.1808	0.1627	0.1464	0.1072
OOS-t	2	0.1191	0.1574	0.2252	0.3301	0.4843
OOS-F	50	0.0004	0.0000	0.0000	0.0000	0.0000
OOS-t	50	0.0664	0.0765	0.0925	0.1140	0.1501

C. Nominal Size of 10%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.2795	0.3160	0.3509	0.3720	0.4115
OOS-t	0.2	0.1557	0.1763	0.2077	0.2580	0.3549
OOS-F	1	0.2428	0.2722	0.2941	0.2982	0.3035
OOS-t	1	0.1806	0.2351	0.3250	0.4258	0.5941
OOS-F	2	0.2134	0.2209	0.2002	0.1784	0.1350
OOS-t	2	0.1959	0.2438	0.3358	0.4445	0.6154
OOS-F	50	0.0009	0.0000	0.0000	0.0000	0.0000
OOS-t	50	0.1277	0.1442	0.1630	0.1936	0.2534

Notes: Each element of Panels A, B and C is the local power of either the OOS-t or (modified) OOS-F test for a given permutation of the choice of sample split  $\pi$ , number of excess parameters in the unrestricted model  $k_2$ , and the nominal size of the test. For a description of how the results were generated see Section four of the text.

TABLE IX  
LOCAL POWER OF THE (MODIFIED) OOS-F AND OOS-t STATISTICS: FIXED SCHEME

A. Nominal Size of 1%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.1428	0.1846	0.2313	0.2563	0.2710
OOS-t	0.2	0.0237	0.0226	0.0328	0.0505	0.0716
OOS-F	1	0.1392	0.1822	0.2101	0.2045	0.1918
OOS-t	1	0.0291	0.0364	0.0632	0.0979	0.1627
OOS-F	2	0.1418	0.1625	0.1492	0.1231	0.0952
OOS-t	2	0.0287	0.0437	0.0569	0.0855	0.1471
OOS-F	50	0.0808	0.0461	0.0093	0.0031	0.0000
OOS-t	50	0.0299	0.0250	0.0208	0.0197	0.0205

B. Nominal Size of 5%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.2492	0.2935	0.3209	0.3343	0.3624
OOS-t	0.2	0.0843	0.0926	0.1171	0.1491	0.2115
OOS-F	1	0.2475	0.2800	0.2816	0.2717	0.2597
OOS-t	1	0.1067	0.1253	0.1767	0.2396	0.3468
OOS-F	2	0.2399	0.2431	0.2111	0.1784	0.1473
OOS-t	2	0.1045	0.1223	0.1701	0.2274	0.3347
OOS-F	50	0.1067	0.0681	0.0244	0.0099	0.0013
OOS-t	50	0.0866	0.0756	0.0704	0.0742	0.0790

C. Nominal Size of 10%		$k_2$				
Test	$\pi$	1	2	5	10	20
OOS-F	0.2	0.3204	0.3507	0.3704	0.3846	0.4160
OOS-t	0.2	0.1489	0.1608	0.2015	0.2469	0.3367
OOS-F	1	0.3149	0.3336	0.3233	0.3137	0.3055
OOS-t	1	0.1784	0.2110	0.2728	0.3514	0.4718
OOS-F	2	0.2939	0.2881	0.2504	0.2163	0.1833
OOS-t	2	0.1711	0.2073	0.2639	0.3390	0.4651
OOS-F	50	0.1344	0.0926	0.0409	0.0184	0.0035
OOS-t	50	0.1379	0.1305	0.1292	0.1352	0.1450

Notes: Each element of Panels A, B and C is the local power of either the OOS-t or (modified) OOS-F test for a given permutation of the choice of sample split  $\pi$ , number of excess parameters in the unrestricted model  $k_2$ , and the nominal size of the test. For a description of how the results were generated see Section four of the text.

Figure 1  
Density Plots for OOS-F: Recursive

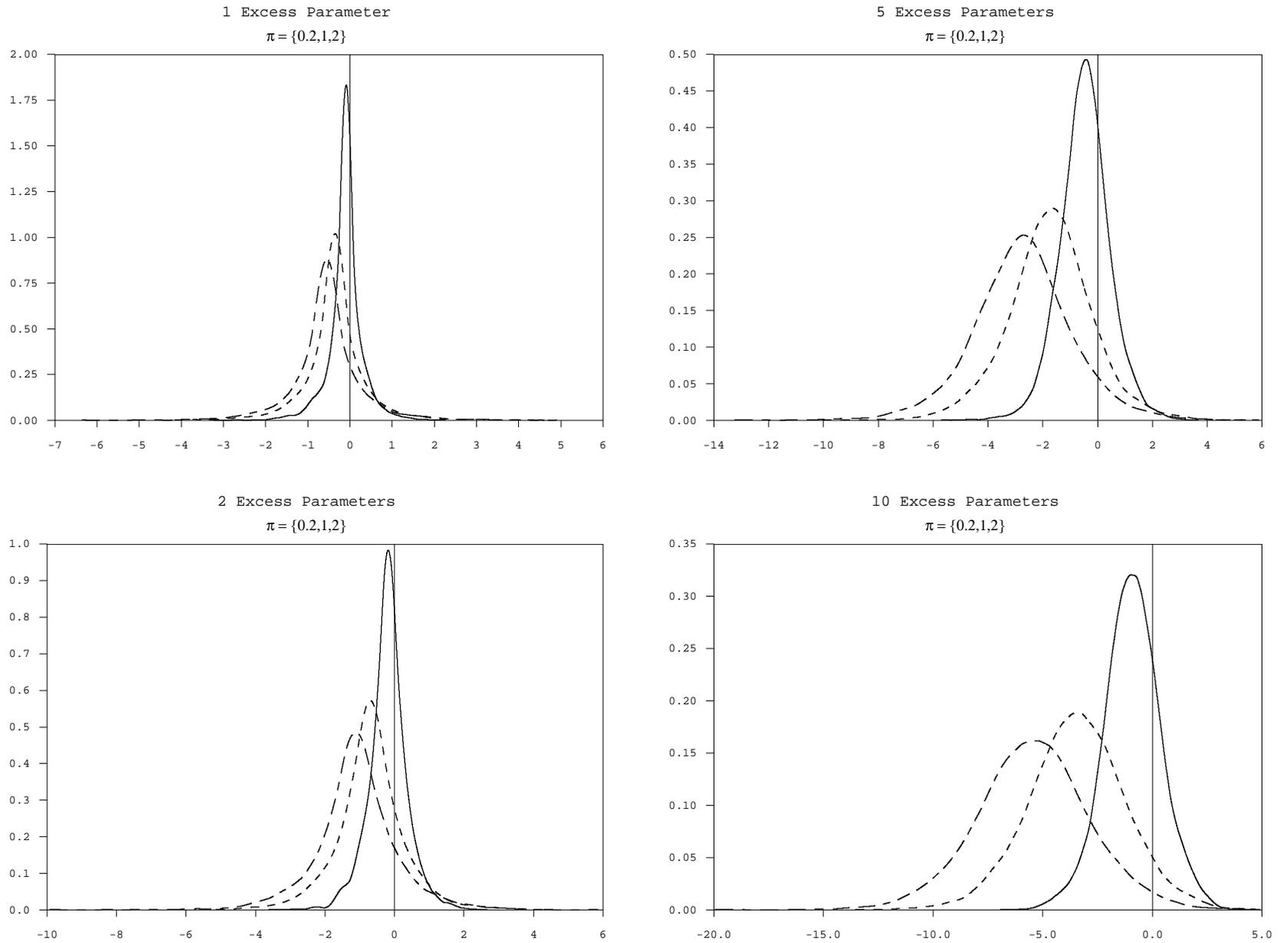


Figure 2  
Density Plots for OOS-F: Recursive

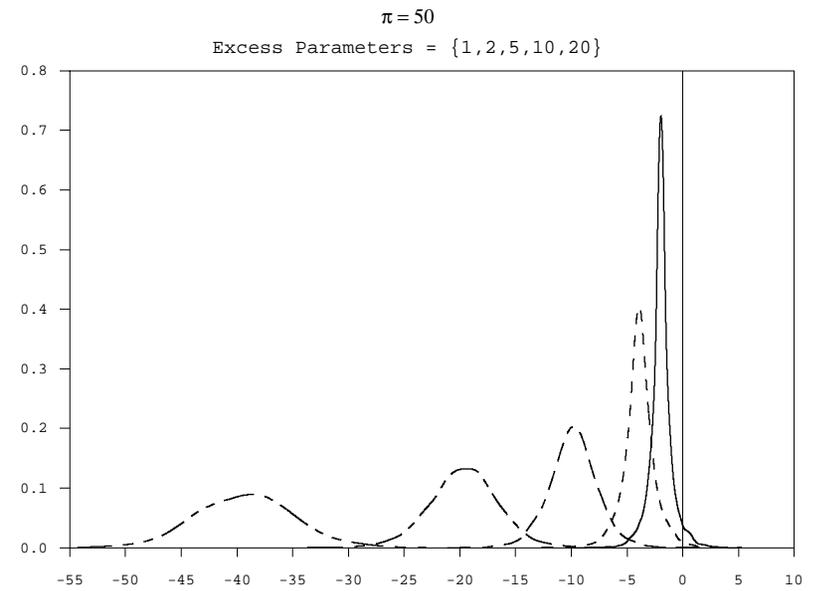
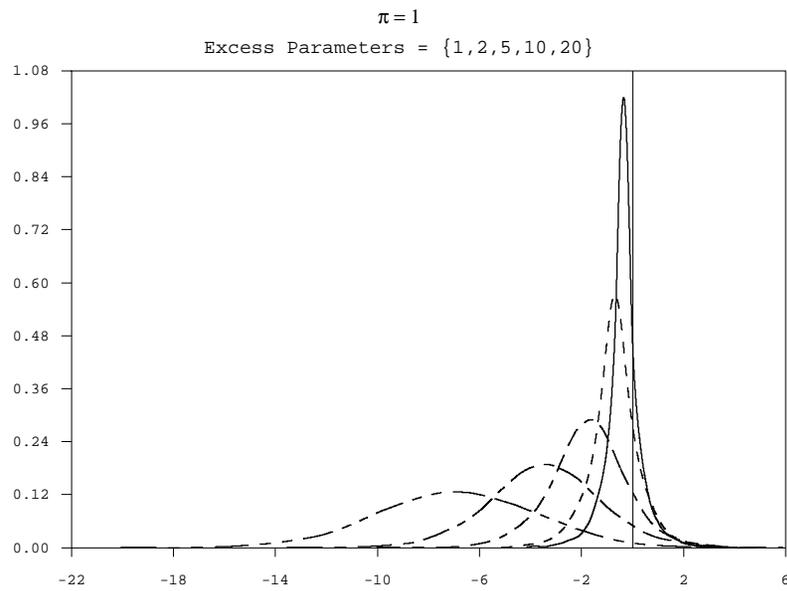
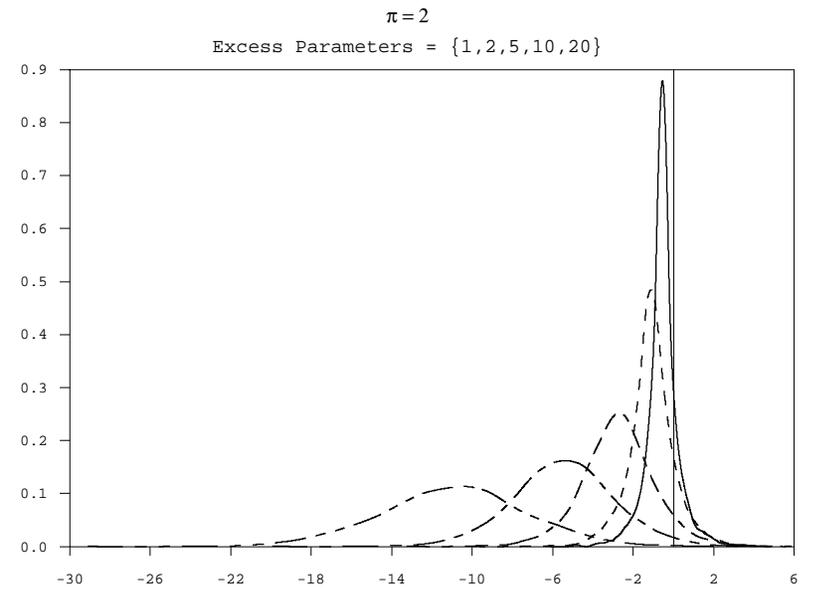
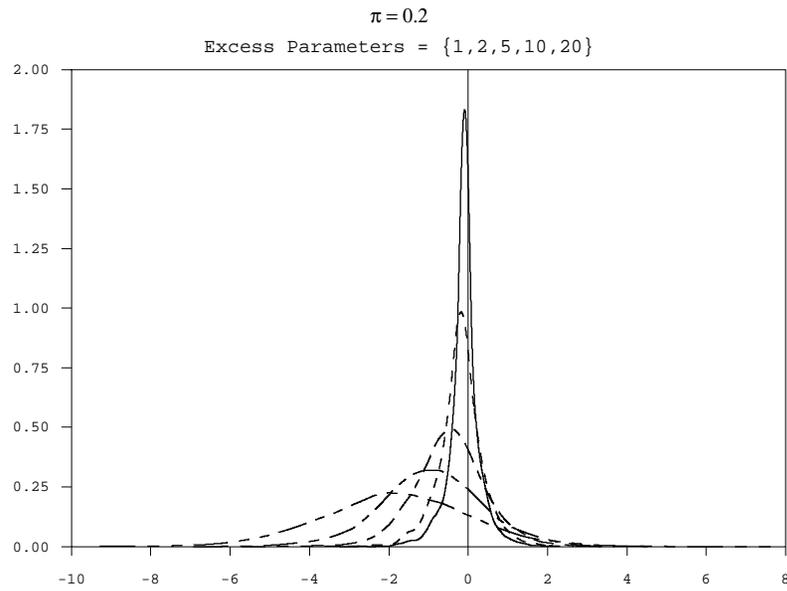


Figure 3  
Density Plots for OOS-t: Recursive

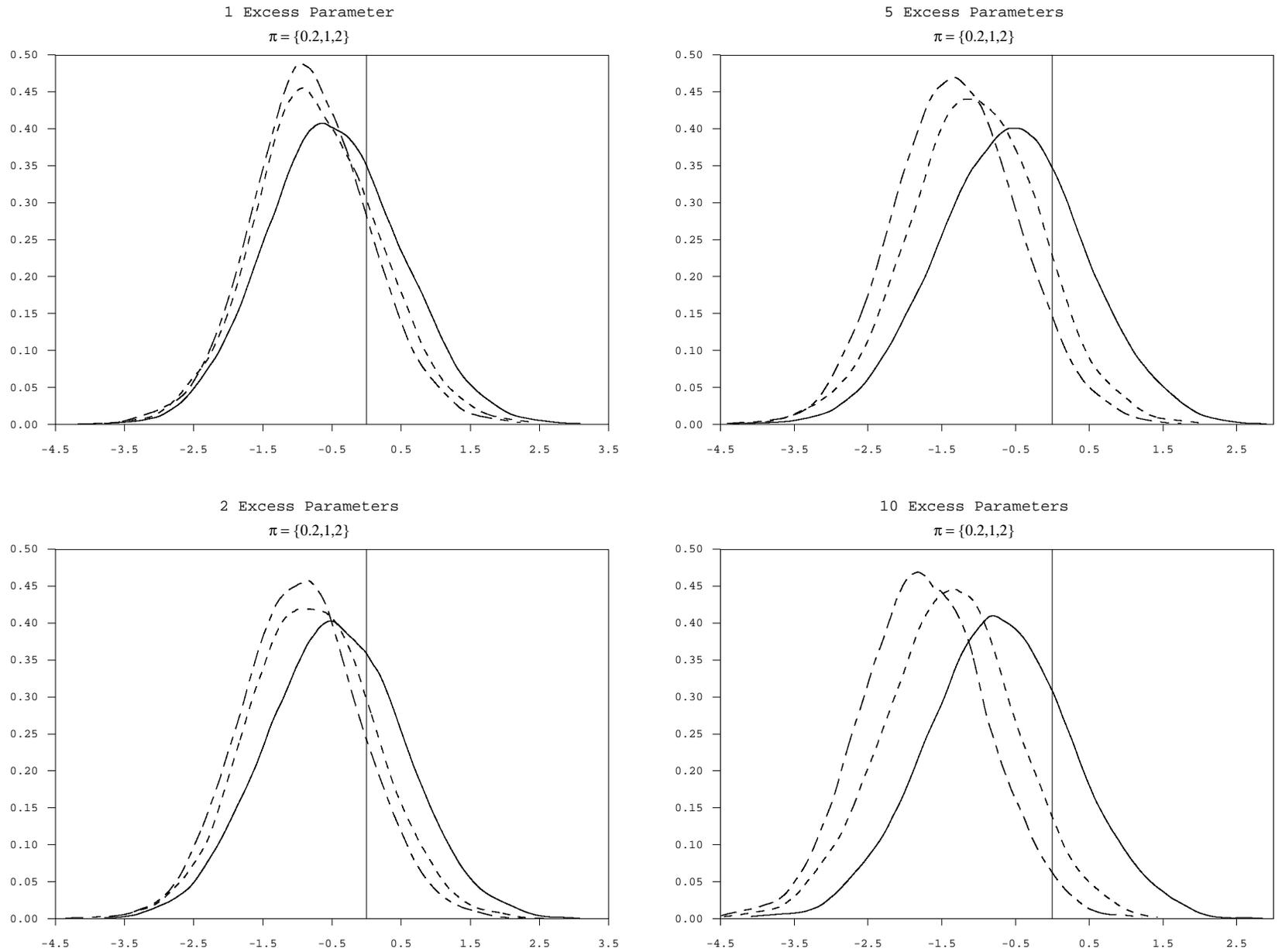


Figure 4  
Density Plots for OOS-t: Recursive

