Asymptotics for Out of Sample Tests of Causality

Michael W. McCracken*
Louisiana State University

February 25, 1999

Abstract

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 Morgan (1939) and Granger and Newbold (1977), a t-type statistic commonly attributed to either West (1996) or Diebold and Mariano (1995) and an F-type statistic akin to Theil's U. Since the limiting distributions under the null are nonstandard, tables of asymptotically valid critical values are provided. The null 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, testing, hypothesis testing, model comparison.

JEL categories: C12, C32, C52, C53.

* Correspondence: Michael W. McCracken, e-mail: mmccrac@unix1.sncc.lsu.edu, phone: 225.388.3782, address: Louisiana State University, Department of Economics, 2107 CEBA, Baton Rouge, LA, 70803-6306, fax: 225.388.3807.

I'd like to thank Todd Clark, Bruce Hansen, Dek Terrell, Ken West and seminar participants at LSU for helpful comments.

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 just 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 metric or 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 work on analytical results for OOS model evaluation, most notably that of West (1996), by providing asymptotic results for OOS tests comparing the predictive ability of two nested models when parameters are estimated. I provide the limiting distribution theory for three commonly used tests that compare the OOS predictive ability of two nested models: a regression-based test for equal MSE proposed by Morgan (1939) and Granger and Newbold (1977), a similar t-type test commonly attributed to either West (1996) or Diebold and Mariano (1995), and an F-type test similar in spirit to Theil's U (1966) but perhaps closer to in-sample likelihood ratio tests. Since the asymptotics of the former two tests are identical I will frequently reference them simultaneously as "OOS-t" and will reference the latter as the "OOS-F" test.

The limiting 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 derived by Andrews (1993), in the context of testing for changepoints in the regression function, but are distinct. Tables are provided in order to facilitate the use of these distributions. Furthermore, since the limiting distributions do not lie within any well-known class of distributions it is unclear how well the statistics perform in terms of power. I therefore provide a limited collection of results on the power of these tests. This information is comprised of analytical results under a local alternative, as well as tables containing numerical calculations of the power under the local alternative.

There are a number of interesting implications of both the asymptotics under the null and under the local alternative. First and foremost the asymptotics under the null (along with the tables) provide a simple method of constructing asymptotically valid tests of OOS predictive ability between two nested models. One useful result of this theory is that I am able to show that the test statistics are asymptotically pivotal. This implies that for large enough sample sizes the bootstrap provides a refinement to first order asymptotics (Hall, 1992). The null asymptotics also have implications for the Principle of Parsimony and overfitting (Tukey 1961, and Box and Jenkins 1976) when OOS predictive ability is the objective. Assuming for the moment that the predictive model is linear, we know that the more irrelevant regressors used the greater is the in-sample predictive ability. 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 irrelevant 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 (1996) are just 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. When the post-sample size is small relative to the in-sample size and the number of extraneous parameters is small the OOS-F tends to be more powerful. As more of the sample is used for post-sample evaluation and the number of extraneous parameters increase the OOS-t tends to be more powerful. The choice of optimal sample split is less clear and is left to section 4.

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. In section four I provide 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. 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), Mizrach (1991) and Diebold and Mariano (1995).

More recent work on OOS hypothesis testing has also developed. Corradi, Swanson and Olivetti (1998) extend previous work to allow for the comparison of non-nested models in the presence of cointegration. McCracken (1998a) 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 (1997) construct tests of equal predictive ability in the presence of ARCH. Diebold, Gunther and Tay (1997) discuss the evaluation of density forecasts. White (1998) 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 the aforementioned authors is whether two models have the same predictive ability with respect to some loss function L(.). Diebold and Mariano (1995) suggest a test of the form

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

where $\hat{\Omega}$ denotes an 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. $\hat{u}_{i,t+1}$ i = 1,2 is the forecast error observed at time t+1 associated with a forecast from time t. Each forecast is constructed using $\hat{\beta}_{i,t}$ an estimator of the parameters associated with each model. West 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 estimates of the limiting variance, Ω, 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 contribute to the limiting variance[1]. Formulas for the correct limiting variance are provided. 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 Pagan and Hall (1983), Davidson and MacKinnon (1984) and Wooldridge (1990).

Unfortunately it is easy to overlook the most crucial condition for limiting normality. For the OOS t-tests to be limiting standard normal Ω must be positive. If Ω is zero then $P^{-0.5}\sum_{t=R}^{T}[L(\hat{u}_{1,t+1})-L(\hat{u}_{2,t+1})] \to_p 0$. 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 Ω 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 relevant 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 Morgan (1939) and Granger and Newbold (1977) they test the null that advertising causes consumption using the t-statistic (and standard normal tables) associated with α_1 from the artificial OLS regression

(2.2) $\hat{u}_{1,t+1}-\hat{u}_{2,t+1}=\alpha_0+\alpha_1(\hat{u}_{1,t+1}+\hat{u}_{2,t+1})$ + 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

[1] See Randles (1982) for the in-sample analog.

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 (mds) 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. Killian (1997) constructs similar tests but under the null that changes in exchange rates form an mds around a nonzero unconditional mean. It should be mentioned that Mark (1995), Killian (1997) and Berkowitz and Giorgianni (1996) all use the bootstrap when conducting their hypothesis tests in these Long-Horizon regressions and 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 3 of this paper indicate that those distortions may also be due to the fact that the limiting distribution is not standard normal nor is well approximated by a standard normal distribution.

In this paper I focus on constructing asymptotically valid OOS sample tests that compare the predictability of two nested parametric models. As mentioned in the introduction I focus on three different statistics. The first two, those from (2.1) and (2.2), are OOS-t statistics. The third is an OOS-F statistic of the form

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

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 $\hat{u}_{i,t+1}^2 = L(\hat{u}_{i,t+1})$ i = 1,2. As in the descriptions of (2.1) and (2.2) I will denote the restricted model as i = 1 and the unrestricted model as i = 2.

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) have tested 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). See section 3.5 for a discussion of penalty terms in the context of the Principle of Parsimony and its relationship to OOS predictive ability.

I choose to introduce the OOS-F for a number of reasons. It is essentially identical to 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 3 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 to the in sample F-test.

3. Theoretical Results

In this section I derive the limiting distributions of both the OOS-t tests in (2.1) and (2.2) and the OOS-F test in (2.3). I do 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. In section 3.3 I present 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. In section 3.4 I specialize the results to the leading case in which parameters are estimated by NLLS and MSE is used to measure predictive ability. Since these distributions are nonstandard I then provide tables of critical values that can be used to construct an asymptotically valid test for equal predictive ability. In section 3.5 I provide a discussion of the results in sections 3.3 and 3.4 in the context of the Principle of Parsimony. Specifically I show that the traditional in-sample application of the Principle of Parsimony is not necessarily appropriate if OOS predictive ability is the relevant metric for a model to be deemed of value.

3.1 Environment

Throughout it will be assumed that there is an observed random sample $\{X_s\}_{s=1}^{T+1}$ of length $\tilde{T} \equiv T + 1$. Using that sample the researcher wishes to compare the OOS one-step ahead predictive ability of two nested parametric regression models. This structure allows for most relevant applications including those discussed in section 2. It does eliminate applications, like that of Diebold and Nason (1990), that use either kernel-based methods or local-regressions to nonparametrically estimate the regression function and construct forecasts. It also eliminates applications, like Swanson and White (1997b), that use series-based nonparametric methods to estimate the regression function and construct forecasts.

The focus on one-step ahead forecasts rather than τ-step ($\tau > 1$) is both substantive and for purposes of clarity. By limiting the discussion exclusively to one-step ahead forecasts I am able to derive results for a wide range of potential loss functions used to measure predictive ability. Clark and McCracken (1999) are able to derive results for longer horizons but under the restriction that MSE is the relevant measure of predictive ability and the nested models are (vector) autoregressive.

Given the pair of nested parametric regression models, I will allow for three methods of constructing the sequence of P, one-step ahead, forecasts. I will refer to these as the recursive, the rolling and the 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 model and the estimated parameters a sequence of P one-step ahead forecasts is then generated. For more discussion on the use of these three schemes see West (1996), West and McCracken (1998) and McCracken (1998a). 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,…,T the parameter estimate $\hat{\beta}_t$ depends explicitly on all information from s = 1,…,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(.) the loss associated with the first forecast is constructed as $L(y_{R+1} - \hat{y}_{R+1}(\hat{\beta}_{i,R}))$ 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,...,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 $s \in [1, t]$.

Chen and Swanson (1996) use 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. Hence 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,...,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 $s \in [t - R + 1, t]$.

Hoffman and Pagan (1989) introduce 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[2]. 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

---

[2] 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.

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

This test is different than those discussed in both West (1996) and Diebold and Mariano (1995). The alternative is one-sided rather than two sided due to the fact that the two models are nested by construction.

3.2 Assumptions

Before discussing specific assumptions some notation is required. For any function $f_{i,t}(\beta_i)$ that depends upon a parameter β_i let $f_{1,t}$ denote $f_{1,t}(\beta_1^*)$ and f_t denote $f_{2,t}(\beta_2^*)$. For the loss function $L_{i,t}(\beta_i)$ let $h_{i,t}(\beta_i)$ denote $\frac{\partial}{\partial\beta_i}L_{i,t}(\beta_i)$ and $q_{i,t}(\beta_i)$ denote $\frac{\partial^2}{\partial\beta_i\partial\beta_i'}L_{i,t}(\beta_i)$. For any matrix A with elements $a_{i,j}$ let |A| denote $\max_{i,j}|a_{i,j}|$. For any (m×n) matrix A with column vectors a_i let vec(A) denote the (mn×1) vector $[a_1', a_2', ..., a_n']'$.

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 ≡ $(I_{k_1\times k_1}, 0_{k_1\times k_2})$ (k_1×k, k > k_1). Since the two models are nested we then know that $L_{1,t}(\beta_1^*) = L_{2,t}(\beta_2^*) \equiv L_t$, $Jh_t = h_{1,t}$ and $Jq_tJ' = q_{1,t}$ for all t.

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

Assumption 1: The parameter vectors β_1^* and β_2^* are estimated using the aggregate loss functions, $\Lambda_{1,t}(\beta_1)$ and $\Lambda_{2,t}(\beta_2)$. For i = 1,2, and t = R,…,T, $\Lambda_{i,t}(\beta_i) = t^{-1}\sum_{j=1}^{t}L_{i,j}(\beta_i)$ for the recursive scheme, $\Lambda_{i,t}(\beta_i) = R^{-1}\sum_{j=t-R+1}^{t}L_{i,j}(\beta_i)$ for the rolling scheme and $\Lambda_{i,t}(\beta_i) = R^{-1}\sum_{j=1}^{R}L_{i,j}(\beta_i)$ for the fixed scheme.

This first assumption provides us with two pieces of information. Analytically it tells us that the parameter estimates are of the form $\hat{\beta}_{1,t}$ = argmin $\Lambda_{1,t}(\beta_1)$ and $\hat{\beta}_{2,t}$ = argmin $\Lambda_{2,t}(\beta_2)$. This allows for both linear and nonlinear models estimated by OLS, NLLS and maximum likelihood. The substantive part of the first assumption is that I require that the loss function used to estimate the parameters is the same as the loss function used to measure predictive accuracy. An implication of the assumption is that if I want to use MSE as my measure of OOS predictive ability then I must estimate the parameters 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 what loss functions I choose to work with. For example if I estimate the parameters by maximizing a log-likelihood and then I use that log-likelihood as the measure of predictive ability the limiting distribution is essentially the same as if I had estimated the model by OLS and used MSE as my measure of predictive ability[3].

I impose this restriction in order to insure that the test statistics are asymptotically pivotal. Weiss and Andersen (1984) and Weiss (1996) have also noted important links between the way in which the parameters were estimated and the measure of predictive ability. They note that relative to a given loss function, OOS predictive ability is enhanced if the same loss function is used to estimate parameters rather than using some other means of estimating the parameters.

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

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

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

Most of Assumption 2 is imposed in order to insure 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.

<u>Assumption 3:</u> Let $U_t \equiv [h_t^{'}, vec(h_t h_t^{'} - cB^{-1})', vec(q_t - B^{-1})']'$. (a) $EU_t = 0$, (b) U_t is uniformly L^8 bounded, (c) For some 8 > d > 2, U_t is strong mixing with coefficients of size $\frac{-8d}{8-d}$, (d) $\lim_{T\to\infty} T^{-1} E\sum_{j=1}^{T} U_j U_j^{'} = \Omega < \infty$.

There are two ways in which this assumption differs from those presented in previous work on OOS hypothesis testing. The first is that I impose slightly stronger mixing and moment conditions than those in, for example, McCracken (1998a). This is a result of the fact that the models here are nested. If the models are non-nested then the OOS-t statistics in (2.1) and (2.2) will 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 CLT 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 I require conditions sufficient for the joint weak convergence of partial sums and the averages of these partial sums to Brownian

[3] See the discussion in section 3.4.

motion and 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).

The second manner in which this assumption differs from previous work on OOS hypothesis testing is that I impose mixing and moment conditions on the first derivative, h_t, and second derivative, q_t, of the loss function used to measure predictive ability. In West (1996), McCracken (1998a) and Corradi, Swanson and Olivetti (1998) it is sufficient to place conditions on the level, L_t, and the first derivative, h_t, of the loss function used to measure predictive ability. It is instructive to see why this distinction is necessary.

Suppose that one is interested in testing for equal predictive ability between two linear parametric regression models, with stationary regressors and disturbances, using MSE as the measure of predictive ability. Let the parameters β_i^* from the two models $y_t = Z_{i,t}^{'}\beta_i^* + u_{i,t}$ be estimated by OLS. For the sake of argument suppose that the parameter estimates are generated recursively and that we use the population level weighting matrices[4] $B_i = (EZ_{i,t}Z_{i,t}^{'})^{-1}$ rather than $B_i(t) = (t^{-1}\sum_{s=1}^{t} Z_{i,s}Z_{i,s}^{'})^{-1}$. The difference between the losses associated with the two forecast errors $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ can be written as the sum of three terms:

$$(3.2.1)\quad \hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2 = [u_{1,t+1}^2 - u_{2,t+1}^2] - 2[u_{1,t+1}Z_{1,t+1}^{'}B_1(t^{-1}\sum_{s=1}^{t} u_{1,s}Z_{1,s}) - u_{2,t+1}Z_{2,t+1}^{'}B_2(t^{-1}\sum_{s=1}^{t} u_{2,s}Z_{2,s})]$$
$$+ [(t^{-1}\sum_{s=1}^{t} u_{1,s}Z_{1,s})^{'}B_1Z_{1,t+1}Z_{1,t+1}^{'}B_1(t^{-1}\sum_{s=1}^{t} u_{1,s}Z_{1,s})$$
$$- (t^{-1}\sum_{s=1}^{t} u_{2,s}Z_{2,s})^{'}B_2Z_{2,t+1}Z_{2,t+1}^{'}B_2(t^{-1}\sum_{s=1}^{t} u_{2,s}Z_{2,s})].$$

If the two models are non-nested then, using Lemma 4.1 in West (1996), we know that the first r.h.s. term in (3.2.1) does, the second term may and the last term does not contribute to the limiting variance of $P^{-0.5}\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]$. It is because of these contributions that West (1996) imposes conditions on, in this example, $[u_{1,t}^2, u_{2,t}^2, u_{1,t}Z_{1,t}^{'}, u_{2,t}Z_{2,t}^{'}]^{'}$ but does not impose direct assumptions on $\mathrm{vec}(Z_{2,t}Z_{2,t}^{'} - B_2^{-1})$.

If the two models are nested these conditions do not suffice. If the two models are nested the limiting variance of $P^{-0.5}\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]$ is zero. None of the r.h.s. terms in (3.2.1) contribute to the limiting variance of $P^{-0.5}\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]$. Moreover the first r.h.s. term is identically zero for all t.

That does not imply that the second and third r.h.s. terms in (3.2.1) do not contribute to the limiting variance of $\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]$. In this case (3.2.1) can be rewritten as

$$(3.2.2)\quad \hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2 = 0 + 2[u_{2,t+1}Z_{2,t+1}^{'}(-J^{'}B_1J + B_2)(t^{-1}\sum_{s=1}^{t} u_{2,s}Z_{2,s})]$$
$$- [-(t^{-1}\sum_{s=1}^{t} u_{2,s}Z_{2,s})^{'}J^{'}B_1JZ_{2,t+1}Z_{2,t+1}^{'}J^{'}B_1J(t^{-1}\sum_{s=1}^{t} u_{2,s}Z_{2,s})$$
$$+ (t^{-1}\sum_{s=1}^{t} u_{2,s}Z_{2,s})^{'}B_2Z_{2,t+1}Z_{2,t+1}^{'}B_2(t^{-1}\sum_{s=1}^{t} u_{2,s}Z_{2,s})].$$

[4] I do this purely for exposition. It is not a requirement for the results of section 3.3.

The second and third r.h.s. terms in (3.2.2) do contribute to the limiting variance of $\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]$. As we will see in section 3.3, $\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]$ rather than $P^{-0.5}\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]$ is the component of interest when constructing tests for OOS predictive ability between two nested models. For this reason, and in this example, I place conditions on $[u_{2,t}Z_{2,t}', vec(u_{2,t}^2 Z_{2,t}Z_{2,t}' - \sigma_u^2 B_2^{-1})', vec(Z_{2,t}Z_{2,t}' - B_2^{-1})']'$.

<u>Assumption 4:</u> Let (a) $Eh_t h_t' = cEq_t \equiv cB^{-1}$ for a positive finite constant c, (b) $E(h_t \mid h_{t-j}, q_{t-j}, j = 1,2,...) = 0$.

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 I must impose some conditions on the loss function L. Here I require 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(.) is a true log-likelihood then that constant is one (Amemiya 1985).

The need for the constant c arises from the fact that MSE is the most common measure of predictability. It is true that 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 σ_u^2 then the OLS estimates of β_2^* are numerically identical to those estimated using the log-likelihood. But that is not the same as saying that h_t and q_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 $h_t^{(OLS)} = -2u_{2,t}Z_{2,t}$, $q_t^{(OLS)} = 2Z_{2,t}Z_{2,t}'$ and hence $Eh_t^{(OLS)}h_t^{(OLS)'} = 4\sigma_u^2 EZ_{2,t}Z_{2,t}' \neq 2EZ_{2,t}Z_{2,t}' = Eq_t^{(OLS)}$. Similarly if we maximize the log-likelihood $h_t^{(MLE)} = -\sigma_u^{-2}u_{2,t}Z_{2,t}$, $q_t^{(MLE)} = \sigma_u^{-2}Z_{2,t}Z_{2,t}'$ and hence $Eh_t^{(MLE)}h_t^{(MLE)'} = \sigma_u^{-2}EZ_{2,t}Z_{2,t}' = Eq_t^{(MLE)}$. It is this difference that generates the need for the constant c. For a more detailed discussion see section 3.4.

<u>Assumption 5:</u> $\lim_{T\to\infty} P/R = \pi$, T = R + P - 1, $0 < \pi < \infty$.

This final assumption introduces the means by which the asymptotics are achieved. As in Hoffman and Pagan (1989), West (1996) and White (1998) the limiting distribution results are derived by imposing a slightly stronger condition than simply that the sample size becomes arbitrarily large. Here I impose the additional condition that both the number of in-sample (R) and OOS (P) observations also become arbitrarily large at the same rate. In this way I insure 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

In this section I provide the null limiting distribution results for the OOS-t statistic in (2.1) and the OOS-F statistic in (2.3). Since the limiting distributions are non-standard I also provide tables of the critical values necessary for the construction of asymptotically valid tests of the null in (3.1.1). All proofs can be found within the Appendix.

There are two main components of the OOS-t and OOS-F statistics. The first, $\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t})-L_{2,t+1}(\hat{\beta}_{2,t})]$, was alluded to in section 3.2 and is a component of all three test statistics. The second, $\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t})-L_{2,t+1}(\hat{\beta}_{2,t})]^2$, 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, from the discussion in section 3.2, we know is equal to zero.

To see how these components affect the OOS-t and OOS-F statistics let's rewrite (2.1). I'll ignore the Morgan/Granger-Newbold statistic from (2.2) for the moment since it is asymptotically equivalent to (2.1) when the loss function is MSE.

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

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

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 estimator of the type suggested by Newey and West (1994)[5]. I emphasize this case because it is most common. Clark (1999) and Harvey, Leybourne and Newbold (1998) consider special cases in which 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 define H(t) as $t^{-1}\sum_{s=1}^{t}h_s$, $R^{-1}\sum_{s=t-R+1}^{t}h_s$ and $R^{-1}\sum_{s=1}^{R}h_s$ for the recursive, rolling and fixed schemes respectively. For matrices C and A defined in Lemma 3.1 let $c^{-0.5}A'CB_2^{0.5}h_t = \tilde{h}_t$ and $c^{-0.5}A'CB_2^{0.5}H(t) = \tilde{H}(t)$.

[5] I also could have estimated Ω using squared deviations from the sample mean. Doing so is asymptotically irrelevant and hence is omitted for notational convenience.

Lemma 3.1: (a) Let $-J'B_1J + B_2 = M$ and $B_2^{-0.5}MB_2^{-0.5} = Q$, 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: $\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]$ =

$$c[\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t))'(T^{-0.5}\tilde{h}_{t+1}) - (0.5)T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t))'(T^{0.5}\tilde{H}(t))] + o_p(1).$$

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

When deriving the limiting distribution of the in-sample F-statistic and other comparable tests comparing nested models, one first shows that the statistic can be written as a weighted quadratic of, say, a $(k\times 1)$ 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 $T^{0.5}\tilde{H}(t) \equiv T^{0.5}t^{-1}\sum_{s=1}^{t}\tilde{h}_s = \frac{T}{t}(T^{-0.5}\sum_{s=1}^{t}\tilde{h}_s)$ for the recursive scheme and let W(s) denote a $(k_2\times 1)$ standard Brownian Motion on $[\lambda,1]$ with W(0) = 0 and $\lambda \equiv (1+\pi)^{-1}$. Since the increments $\tilde{h}_s$ are conditionally homoskedastic vector martingale differences with unit variance and T/t is bounded by Assumption 5, $T^{0.5}\tilde{H}(t)$ is well approximated (in distribution) 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 3.3 shows that the data generating process and loss function are irrelevant to the asymptotics but for the factor c. To eliminate that factor I define the OOS-F as 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 r.h.s. 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 = $\frac{\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]}{\hat{c}} \rightarrow_d F^1$ where F^1 equals

$\int_{\lambda}^{1} s^{-1}W'(s)dW(s) - (0.5)\int_{\lambda}^{1} s^{-2}W'(s)W(s)ds$ for the recursive scheme,

$\lambda^{-1}\{W(1) - W(\lambda)\}'W(\lambda) - (0.5)\pi\lambda^{-1}W'(\lambda)W(\lambda)$ for the fixed scheme,

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 rolling scheme.

Theorem 3.2: OOS-t = $\frac{\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]}{(\sum_{t=R}^{T}[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})]^2)^{0.5}} \rightarrow_d F^2$ where F^2 equals

$$\frac{F^1}{[\int_{\lambda}^{1} s^{-2}W'(s)W(s)ds]^{0.5}}$$ for the recursive scheme,

$$\frac{F^1}{[\pi\lambda^{-1}W'(\lambda)W(\lambda)]^{0.5}}$$ for the fixed scheme.

$$\frac{F^1}{[\lambda^{-2}\int_{\lambda}^{1}\{W(s) - W(s-\lambda)\}'\{W(s) - W(s-\lambda)\}ds]^{0.5}}$$ for the rolling scheme.

There are a number of things to notice about Theorems 3.1 and 3.2. The first is that they are both pivotal. If one is going to use asymptotic critical values from the limiting distributions to construct asymptotically valid tests of the null then that fact is not particularly useful. What does make it useful is the fact that the bootstrap is an increasingly common method of constructing test statistics. Mark (1995) and Killian (1997) bootstrap statistics directly comparable to the OOS-t and OOS-F statistics. Ashley (1998) provides a detailed bootstrap methodology for constructing asymptotically valid tests of the null of equal OOS predictive ability between two nested models. In any of these cases, since the limiting distribution is pivotal, we know from Hall (1992) that the bootstrap provides refinements to first order asymptotics and hence in finite samples may provide more accurate inference[6].

A second fact to note is that the limiting distributions do not depend upon the choice of loss function L(.). 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. McCracken (1998a), in the context of non-nested models, and Corradi, Swanson and Olivetti (1998), in the context of nested models, each show that finite sample performance can be better for some loss functions than others, particularly if the loss functions differ in smoothness.

[6] Killian (1998) offers a passionate argument against the relevance of such refinements.

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 dimensionality of the vector Brownian motion W(s). It is easier to see if we rewrite F^1. Consider the recursive sampling scheme. If we let $W_i(s)$ denote the i^{th} element of W(s) then

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

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

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

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^1 is mean zero for all k_2 while the second term is increasingly negative. See section 3.4 for a discussion of the repercussions of this fact on the empirical application of the Principle of Parsimony. A less important effect due to k_2 is on the variance of F^1. Since F^1 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^1 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 then suggest a transformed version of the F-statistic that they show 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, π, 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 $\lambda = (1+\pi)^{-1}$). It also affects the range of integration on each of the stochastic integrals through λ. Since the parameter π 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 π just as it did with k_2. Once again see section 3.4 for a discussion on repercussions for the Principle of Parsimony. Numerical results indicate that the variance of F^1 is also monotonically increasing in π for a fixed value of k_2. For F^2, numerical results suggest that the mean is decreasing and the variance is increasing in π but to a much 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 I have yet to provide the limiting distribution of the OOS-t statistic from (2.2). The reason I have done this is 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: Morgan/Granger-Newbold statistic: Let $L_{i,t+1}(\beta_i) = u_{i,t+1}^2(\beta_i)$, $a_{0,T} \equiv P^{-1}\sum_{t=R}^{T}[u_{1,t+1}^2(\hat{\beta}_{1,t}) - u_{2,t+1}^2(\hat{\beta}_{2,t})]$, $a_{1,T} \equiv P^{-1}\sum_{t=R}^{T}[u_{1,t+1}(\hat{\beta}_{1,t}) - u_{2,t+1}(\hat{\beta}_{2,t})]^2$ and $a_{2,T} \equiv P^{-1}\sum_{t=R}^{T}[u_{1,t+1}(\hat{\beta}_{1,t}) + u_{2,t+1}(\hat{\beta}_{2,t})]^2$. $\dfrac{P^{0.5}a_{0,T}}{[a_{1,T}a_{2,T} - a_{0,T}^2]^{0.5}} \to_d F^2$

Theorem 3.3, along with Theorem 3.2, states that the two OOS-t statistics are asymptotically equivalent. Hence to construct asymptotically valid tests of equal predictive ability using either of the tests one can use the same critical values.

Tables 1, 2 and 3 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 value of k_2 and π. Generating these draws proceeded as follows. Weights that depend upon π 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+1st observation to the Tth observation. The random number generator was seeded so that all k_2 and π 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 1, 2 and 3. Each of the tables corresponds to either the recursive, rolling or fixed scheme. Within each table there are 300 critical values. Each of these correspond to one permutation of three parameters: k_2 = {1, 2, 3, …,9 , 10}, π = {0.2, 0.4,…, 1.0, 1.2,…, 2.0}[7] and nominal size of the test = {0.01, 0.05, 0.10}. Tables that allow for larger values of both k_2 and π are available from the authors website.

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 I discussed how c was defined. Specifically I showed

[7] These values of π correspond to percentages of in-sample observations λ = {0.833, 0.714, 0.625, 0.555, 0.500, 0.454, 0.417, 0.385, 0.357, 0.333}.

that if OLS is used (or NLLS or any equivalent means) to estimate the parameters and MSE to measure predictive ability then $Eh_t^{(OLS)}h_t^{(OLS)'} = 2\sigma_u^2 Eq_t^{(OLS)}$. This was in contrast to the case where we estimated the parameters by maximizing a log-likelihood and then used the same log-likelihood to measure predictive ability. In this latter case we know that $Eh_t^{(MLE)}h_t^{(MLE)'} = Eq_t^{(MLE)}$.

Introducing 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 $\hat{c} = 2(P^{-1}\sum_{t=R}^{T}\hat{u}_{2,t+1}^2)$[8].

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), Killian (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.

If I modify the definition of the OOS-F in accordance with these applications we have

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

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

Corollary 3.1: Let $L_{i,t+1}(\beta_i) = u_{i,t+1}^2(\beta_i) \equiv (y_{t+1} - g_{i,t+1}(\beta_i))^2$, then $\dfrac{\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]}{P^{-1}\sum_{t=R}^{T}\hat{u}_{2,t+1}^2} \to_d 2F^1$.

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 4, 5 and 6 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 300 values reported on each of the three tables correspond to the same permutations of k_2, π and nominal size of the test that were used in Tables 1, 2 and 3. More detailed tables are available from the authors website.

It should be emphasized that Tables 4, 5 and 6 cannot be directly applied to applications where a 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 4 - 6 can

[8] This estimator is consistent by Theorem 4.1 of West (1996).

be used only after they have been divided by 2. 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 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.

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 1 - 6 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 $2F^1$ 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 both 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 recursive and fixed (modified) OOS-F are -64.018 and -540.728 respectively.

Figure 1 is comprised of four plots. Each show the effect on the density of $2F^1$ when π 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 π 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 $2F^1$ when k_2 increases from 1 to 2 to 5 to 10 to 20 holding π 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. Once 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 π. For the moment focus on the (modified) OOS-F statistic. Algebraically this implies that under the null

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

increases in both k_2 and π.

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 empirical 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. Certainly it is this logic that is behind Akaike's (1969) introduction of penalty terms and information criteria. Rather than rely upon the heuristic Principle of Parsimony, penalty terms provide a statistical mechanism for eliminating the potential for selecting overparameterized parametric regression models.

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 π become arbitrarily large. This implies that the burden of proof, that is on the unrestricted model in-sample, is more so on the restricted model for large enough k_2 and π and when OOS predictive ability is of interest. The MSE from the restricted model must also be "significantly" lower than the MSE from the restricted model[9].

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. For large enough k_2 and π, the critical values from Tables 1 - 6 indicate that it is more likely that the alternative is true than is the null[10].

Another related implication is on the use of OOS information criteria to identify regression models. Swanson and White (1995, 1997a) have applied this methodology. In these, and 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 π 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 π. Then as k_2 and π 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

---

9 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.

10 In other words, I can always choose k_2 and π sufficiently large that zero lies in the rejection region for a given nominal size of the test.

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 π. 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 3.5.1. In Theorem 3.5.1 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.

4.0 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 1 - 6 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, π, or the sampling scheme in order to maximize power of the test. In this section I provide a limited set of evidence concerning the 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 π. 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 π.

Certainly given two tests of equal size we prefer the test with greater power. There are at least two common methods of providing evidence concerning the power of a given test. One is to conduct a series of Monte Carlo experiments. Clark (1998), Diebold and Mariano (1995) and Corradi, Swanson and Olivetti (1998) provide evidence concerning the finite sample size and power properties of OOS tests of equal predictive ability between two non-nested models. Sullivan and White (1998) do the same for White's (1998) bootstrap adjustment for data-snooping. McCracken (1998b) shows analytically and through Monte Carlo evidence how data-mining can affect model-based in-sample and OOS hypothesis testing. Swanson, Ozyildirium and Pisu (1996) conduct an extensive series of Monte Carlo experiments concerning both in-sample and OOS tests for equal predictive ability between two nested models.

In this paper I will focus on using a second method. Here I will derive the limiting distribution of both the OOS-t and (modified) OOS-F under a sequence of local alternatives akin to those discussed by Kendall and Stuart (1979) and Davidson and MacKinnon (1987). Clark and McCracken (1999) provide Monte Carlo evidence on the finite sample size and power performance of these statistics 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 will focus on the most relevant application. I will 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 alternative models can then be specified as

(4.1) H_0: $y_t = Z_{1,t}'\beta_1^* + u_{1,t}$ vs.

$$H_A:\ y_t = Z_{1,t}'\beta_1^* + \frac{\sigma_u Z_{22,t}'\beta_{2,2}^*}{T^{0.5}} + u_{2,t} = Z_{2,t}'\beta_2^* + (\frac{\sigma_u}{T^{0.5}} - 1)Z_{22,t}'\beta_{22}^* + u_{2,t}$$

where $Z_{2,t} = (Z_{1,t}', Z_{22,t}')'$, $\beta_2^* = (\beta_1^{*'}, \beta_{22}^{*'})'$, $\beta_{22}^* \neq 0$ and $u_{2,t}$ is i.i.d. $(0, \sigma_u^2)$.

I choose this alternative specification for two main reasons. The first is that linear models are the parametric regression model of choice in most applications, including 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.

In the following Assumptions 1 - 5 are maintained.

Theorem 4.1: Let $L_{i,t+1}(\beta_i) = u_{i,t+1}^2(\beta_i)$ 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 $\to_d$ F^3 and OOS-t $\to_d$ F^4 where

$$F^3 = 2F^1 + \pi\lambda\beta_2^{*'}J_2'B_2^{-0.5}QB_2^{-0.5}J_2\beta_2^* + 2\beta_2^{*'}J_2'B_2^{-0.5}CA[W(1) - W(\lambda)]$$

and

$$F^4 = \frac{F^3}{2[\chi_2 + \pi\lambda\beta_2^{*'}J_2'B_2^{-0.5}QB_2^{-0.5}J_2\beta_2^* + 2\beta_2^{*'}J_2'B_2^{-0.5}CA\chi_3]^{0.5}}$$

for χ_2 defined as the square of the denominator term in Theorem 3.2 and for χ_3 defined 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.

The first thing that should be noted about Theorem 4.1 is that the limiting distributions are not pivotal. The local power of the tests depends upon the data generating process through B_2 and β_2^*. Under the null $\beta_2^* = 0$ and B_2 was irrelevant. Under the local alternative both affect the limiting distribution. 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 7, 8 and 9 provide a brief list of local power characteristics. The calculations apply the same methods used to construct the critical values in Tables 1 - 6. 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 I needed to choose a particular data generating process. In order to simplify presentation I chose one for which the regressors are i.i.d. orthonormal and are of equal relevance to the conditional mean function so that $\beta_2^* = (1,1,\ldots,1)'$. 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

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

where W_i and $\chi_{i,3}$ represent the i^{th} components of W and χ_3 respectively.

Tables 7, 8 and 9 report the percentage of 5000 draws of F^3 and F^4 that were greater than the relevant critical values reported in Tables 1 - 6. In each table the local power of the OOS-t and OOS-F is reported for a range of π = (0.2, 1.0, 2.0, 50.0), k_2 = (1, 2, 5, 10, 20) and nominal size of the test = (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 7 relates to the recursive scheme, 8 to the rolling and 9 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 π are smaller the OOS-F is more powerful than the OOS-t. As either k_2 or π become sufficiently large the OOS-t becomes more powerful. Hence given a particular k_2 or π pair, Tables 7 – 9 provide some guideline on the choice between the OOS-F and OOS-t statistics.

Comparing local power across Tables 7 – 9 we can also draw some conclusions on the choice of sampling scheme. Of the 60 possible (k_2, π, nominal size) comparisons among the three sampling schemes the recursive scheme is most powerful 57 and the rolling 3 times when the OOS-t is used. When the OOS-F is used the recursive is most powerful 39 and the fixed 21 times. In this latter case, the fixed scheme is most powerful only when both k_2 and π are smaller. It therefore seems that when choosing a sampling scheme the recursive scheme should be the first choice unless both k_2 or π are small in which case perhaps the fixed should be considered.

Deciding upon the optimal sample split parameter π 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 π (π = 50.0) are best when the OOS-t is used. For the OOS-F the optimal split is usually small (π = 0.2) when k_2 is small and moderate (π = 1.0) when k_2 is larger. For both the fixed and rolling schemes smaller values of π (π = 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 (π = 1.0) but when using the rolling scheme power is highest at slightly higher levels (π = 2.0).

5 Conclusion

In this paper I derive the limiting distributions of three statistics commonly used for testing the null of equal predictive ability between two nested models. These statistics include a regression-based statistic proposed by Morgan (1939) and Granger and Newbold (1977), a t-type statistic often attributed to either West (1996) or Diebold and Mariano (1995) and a likelihood ratio-type statistic akin to Theil's U and the in-sample F-statistic (or likelihood ratio test). The limiting distributions of these three statistics are non-standard. I therefore provide numerically calculated asymptotic critical values so that asymptotically valid tests of equal predictive ability can be constructed.

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

Perhaps the most interesting results concern implications for the Principle of Parsimony and overfitting. In section 3.5 I show that the probability of the MSE of an overparameterized model being larger than the MSE of a less parameterized model is increasing in the number of excess parameters in the overparameterized model. This simple result implies that the common in-sample application of the Principle of Parsimony is inappropriate when applied OOS. Moreover it provides the first analytical evidence for why overfit models might tend to show poor OOS predictive ability (Diebold p.47, 1998).

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) or kernel-based (Chung and Zhou, 1996) 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 to 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 3 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.

Amemiya, Takeshi (1985): Advanced Econometrics, (Cambridge, Harvard University Press).

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.

Ashley, Richard (1998): “A New Technique for Postsample Model Selection and Validation”, *Journal of Economic Dynamics and Control*, 22, 647-665.

Ashley, R., 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 (1996): “Long-Horizon Exchange Rate Predictability”, manuscript, Federal Reserve Board of Governors and International Monetary Fund.

Box, G.E.P. and Jenkins (1976): Time Series Analysis, Forecasting and Control, (San Francisco, Holden Day).

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.

Chen, Xiaohong and Norman R. Swanson (1996): “Semiparametric ARX Neural Network Models with an Application to Forecasting Inflation”, Working Paper, University of Chicago and Pennsylvania State University.

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.

Clark, Todd E. and Michael W. McCracken (1999): "Granger Causality Tests of Equal Forecast Accuracy and Encompassing", manuscript, Federal Reserve Bank of Kansas City.

Corradi, V., N.R. Swanson and C. Olivetti (1998): “Predictive Ability with Cointegrated Variables”, manuscript, University of Pennsylvania and Pennsylvania State University.

Davidson, Russell and James G. MacKinnon (1984): "Model Specification Tests Based on Artificial Linear Regressions", *International Economic Review*, 25, 485-502.

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).

Diebold, Francis X., Todd A. Gunther and Anthony S. Tay (1997): "Evaluating Density Forecasts", NBER Technical Working Paper #215.

Diebold, Francis X. and Lutz Kilian (1997): "Measuring Predictability: Theory and Macroeconomic Applications", NBER Technical Working Paper #213.

Diebold, Francis X. and Robert S. Mariano (1995): "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, Francis X. 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.

Fair, Ray C. and Robert Shiller (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).

Granger, Clive W.J. 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.

Kendall, M. and A. Stuart (1979): The Advanced Theory of Statistics, Vol. II, (London, Charles Griffin).

Kilian, Lutz (1997): "Exchange Rates and Monetary Fundamentals: What Do We Learn From Long-Horizon Regressions?", manuscript, University of Michigan.

Kilian, Lutz (1998): "Pitfalls in Constructing Bootstrap Confidence Intervals for Asymptotically Pivotal Statistics", manuscript, University of Michigan.

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.

Mark, Nelson C. (1995): “Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability”, *The American Economic Review*, 85, 201-218.

Mark, Nelson C. and Donggyu Sul (1998): “Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Seventeen Country Panel”, manuscript, Ohio State University.

McCracken, Michael W. (1998): "Robust Out of Sample Inference", submitted to Journal of Econometrics, Louisiana State University.

McCracken, Michael W. (1998): “Data Mining and Out-of-Sample Inference”, 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.

Meese, Richard A. and Kenneth Rogoff (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.

Newey, 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).

Newey, Whitney K. and Kenneth D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation", *Review of Economic Studies*, 61, 631-654.

Pagan, Adrian R. and Anthony D. Hall (1983): "Diagnostic Tests as Residual Analysis", *Econometric Reviews*, 2, 159-218.

Pagan, Adrian R. 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.

Randles, 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., 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.

Swanson, Norman R. 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.

Swanson, Norman R. 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.

Swanson, Norman R. 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): <u>Applied Economic Forecasting</u>, (Amsterdam, North-Holland).

Tukey, John W. (1961): "Discussion, Emphasizing the Connection Between Analysis of Variance and Spectrum Analysis", *Technometrics*, 3, 191-219.

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.

Weiss, Andrew A. 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.

West, Kenneth D. and Dongchul Cho (1995): "The Predictive Ability of Several Models of Exchange Rate Volatility", *Journal of Econometrics*, 69, 367-391.

West, Kenneth D. and Michael W. McCracken (1998): "Regression-Based Tests of Predictive Ability", *International Economic Review*, 39, 817-840.

White, Halbert (1998): "A Reality Check for Data Snooping", manuscript, University of California San Diego and QuantMetrics R&D Associates, LLC.

Wolff, Christian C.P. (1987): “Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models”, *Journal of Business and Statistics*, 5, 87-97.

Wooldridge, Jeffrey M. (1990): "A Unified Approach to Robust, Regression-Based Specification Tests", *Econometric Theory*, 6, 17-43.

Wooldridge, Jeffrey M. and Halbert White (1989): "Central Limit Theorems for Dependent, Heterogeneous Processes with Trending Moments", manuscript, Michigan State University.

Appendix

Lemma A1: For all $a \in [0,0.5)$ and each i = 1,2, (a) $\sup_t |\hat{\beta}_{i,t} - \beta_i^*|$ = $o_p(1)$, (b) $\sup_t |B_i(t) - B_i|$ = $o_p(1)$, (c) $\sup_t t^a |U(t)| = o_p(1)$, (d) $\sup_t t^a |\hat{\beta}_{i,t} - \beta_i^*|$ = $o_p(1)$, (e) $\sup_t |T^{0.5}(\text{vec}[B_i(t)] - \text{vec}[B_i])|$ = $O_p(1)$.

Proof of Lemma A1: (a) Given Assumptions 2 and 3, the result follows from Theorem 2.1 of Newey and McFadden (p.2121, 1994) adjusted for a.s. convergence.

(b) I will show this for the recursive scheme. The fixed scheme follows immediately from the recursive and the rolling follows from a proof similar to that for the recursive.

First notice that $g(\text{vec}[B_i^{-1}(t)]) - g(\text{vec}[B_i^{-1}])$ = $\text{vec}[B_i(t)] - \text{vec}[B_i]$ for a continuously differentiable function g(.) with $\frac{\partial g(\nu)}{\partial \nu} \equiv g_\nu(\nu)$. Second notice that there exists a neighborhood $N(\text{vec}[B_i^{-1}])$ of $\text{vec}[B_i^{-1}]$ and a finite positive constant D such that $\sup_{\nu \in N(\text{vec}[B_i^{-1}])} |g_\nu(\nu)| < D$. Hence taking a first order Taylor expansion of g(.) around $\text{vec}[B_i^{-1}]$ we have

$$\sup_t |\text{vec}[B_i(t)] - \text{vec}[B_i]| = \sup_t |g_\nu(\tilde{\nu}_t)'(\text{vec}[B_i^{-1}(t)] - \text{vec}[B_i^{-1}])| \tag{1}$$

$$\leq k \sup_t |g_\nu(\tilde{\nu}_t)| \sup_t |(\text{vec}[B_i^{-1}(t)] - \text{vec}[B_i^{-1}])|$$

for some $\tilde{\nu}_t$ on the line between $\text{vec}[B_i^{-1}(t)]$ and $\text{vec}[B_i^{-1}]$.

Now consider $\sup_t |(\text{vec}[B_i^{-1}(t)] - \text{vec}[B_i^{-1}])|$ and recall that for the recursive scheme $\text{vec}[B_i^{-1}(t)] = \text{vec}[t^{-1}\sum_{j=1}^t q_{i,j}(\tilde{\beta}_{i,t})]$. Adding and subtracting $\text{vec}[t^{-1}\sum_{j=1}^t q_{i,j}]$ we have

$$\sup_t |\text{vec}[B_i^{-1}(t)] - \text{vec}[B_i^{-1}]| \leq \tag{2}$$

$$\frac{T^{0.5}}{R} \sup_t |T^{-0.5}\sum_{j=1}^t \text{vec}[q_{i,j} - Eq_{i,j}]| + \sup_t |t^{-1}\sum_{j=1}^t \text{vec}[q_{i,j}(\tilde{\beta}_{i,t}) - q_{i,j}]|.$$

That the first r.h.s. term in (2) is $o_p(1)$ follows from the fact that $\frac{T^{0.5}}{R}$ is o(1) by Assumption 5 and by Theorem 3.1 of Hansen (1992), $\sup_t |T^{-0.5}\sum_{j=1}^t \text{vec}[q_{i,j} - Eq_{i,j}]|$ = $O_p(1)$. Consider the second r.h.s. term in (2). By Assumption 3 and the triangle inequality we know that

$$\sup_t |t^{-1}\sum_{j=1}^t \text{vec}[q_{i,j}(\tilde{\beta}_{i,t}) - q_{i,j}]| \leq \sup_t t^{-1}\sum_{j=1}^t |\text{vec}[q_{i,j}(\tilde{\beta}_{i,t}) - q_{i,j}]|$$

$$\leq \sup_t t^{-1}\sum_{j=1}^t m_j |\tilde{\beta}_{i,t} - \beta_i^*|^\varphi \leq (\sup_t |\tilde{\beta}_{i,t} - \beta_i^*|)^\varphi (\sup_t t^{-1}\sum_{j=1}^t m_j)$$

$$\leq (\sup_t |\hat{\beta}_{i,t} - \beta_i^*|)^\varphi (\frac{T}{R})(T^{-1}\sum_{j=1}^T m_j). \tag{3}$$

The last inequality uses the fact that for all t, $m_t > 0$ and t > R. Then since Lemma A1 (a) implies that

$\sup_t |\hat{\beta}_{i,t} - \beta_i^*| = o_p(1)$, $\frac{T}{R}$ is bounded, and Markov's inequality implies that $T^{-1}\sum_{j=1}^{T} m_j = O_p(1)$, we know

that $\sup_t |(\text{vec}[B_i^{-1}(t)] - \text{vec}[B_i^{-1}])| = o_p(1)$.

The result will then follow if $\sup_t |g_v(\tilde{v}_t)| = O_p(1)$. By the previous paragraph we know that for all ε > 0 there exists R sufficiently large that $\Pr ob(\tilde{v}_t \in N(\text{vec}[B_i^{-1}])) > 1 - \varepsilon$. Then since $g_v(v)$ is continuous and $D < \infty$ we know that $\sup_t |g_v(\tilde{v}_t)| = O_p(1)$.

(c) Follows from Lemma A3 of West (1996) and Lemma A1 of West and McCracken (1998).

(d) By definition, $\sup_t t^a |\hat{\beta}_{i,t} - \beta_i^*| = \sup_t t^a |B_i(t)H_i(t)|$. Hence by the triangle inequality we have

$$\sup_t t^a |B_i(t)H_i(t)| \leq k(\sup_t |B_i(t) - B_i|)(\sup_t t^a |H_i(t)|) + k|B_i|(\sup_t t^a |H_i(t)|).$$

The result follows from Lemma A1 (b) and (c).

(e) I will show this for the recursive scheme. The fixed scheme follows immediately and the rolling follows from a proof similar to that for the recursive. Using the expansion in (1) we have

$$\sup_t |T^{0.5}(\text{vec}[B_i(t)] - \text{vec}[B_i])| = \sup_t |T^{0.5} g_v(\tilde{v}_t)'(\text{vec}[B_i^{-1}(t)] - \text{vec}[B_i^{-1}])| \quad (4)$$

$$= \sup_t |g_v(\tilde{v}_t)'(\frac{T}{t} T^{-0.5}\sum_{j=1}^{t} \text{vec}[q_{i,j}(\tilde{\beta}_{i,t}) - Eq_{i,j}])|$$

$$\leq k^2 \sup_t |g_v(\tilde{v}_t)| \{\frac{T}{R}\sup_t |\{T^{-0.5}\sum_{j=1}^{t} \text{vec}[q_{i,j} - Eq_{i,j}]| + \frac{T}{R}\sup_t |T^{-0.5}\sum_{j=1}^{t} \text{vec}[q_{i,j}(\tilde{\beta}_{i,t}) - q_{i,j}]|\}.$$

That $\sup_t |g_v(\tilde{v}_t)| = O_p(1)$ follows from the proof of (b) above. That $\sup_t |T^{-0.5}\sum_{j=1}^{t} \text{vec}[q_{i,j} - Eq_{i,j}]| = O_p(1)$ follows from Assumption 5 and Theorem 3.1 of Hansen (1992). By assumption $\frac{T}{R}$ is bounded. It then suffices to show $\sup_t |T^{-0.5}\sum_{j=1}^{t} \text{vec}[q_{i,j}(\tilde{\beta}_{i,t}) - q_{i,j}]| = O_p(1)$. By Assumption 3 and (3) we know that

$$\sup_t |T^{-0.5}\sum_{j=1}^{t} \text{vec}[q_{i,j}(\tilde{\beta}_{i,t}) - q_{i,j}]| \leq (\sup_t T^{0.5} |\hat{\beta}_{i,t} - \beta_i^*|)^{\varphi}(\frac{T}{R})(T^{-1}\sum_{j=1}^{T} m_j).$$

That $T^{-1}\sum_{j=1}^{T} m_j = O_p(1)$ follows by Markov's inequality. If we then add and subtract B_i we have

$\sup_t T^{0.5} |\hat{\beta}_{i,t} - \beta_i^*| \leq k\frac{T}{R}\sup_t |B_i(t) - B_i| \sup_t |T^{-0.5}\sum_{j=1}^{t} h_{i,j}| + k\frac{T}{R}|B_i| \sup_t |T^{-0.5}\sum_{j=1}^{t} h_{i,j}|$. Given (b) it suffices to show $\sup_t |T^{-0.5}\sum_{j=1}^{t} h_{i,j}| = O_p(1)$. The result then follows from Assumption 5 and Theorem 3.1 of Hansen (1992).

Lemma A2: $\sum_t H_i(t)'B_i(t)q_{i,t+1}(\dot{\beta}_{i,t})B_i(t)H_i(t) = \sum_t H_i(t)'B_i q_{i,t+1} B_i H_i(t) + o_p(1)$.

Proof of Lemma A2: Add and subtract B_i as well as $q_{i,t+1}$ to get

$$\sum_t H_i(t)' B_i(t) q_{i,t+1}(\dot{\beta}_{i,t}) B_i(t) H_i(t) =$$

$$\sum_t H_i(t)' B_i q_{i,t+1} B_i H_i(t) + \sum_t H_i(t)' B_i q_{i,t+1}[B_i(t) - B_i] H_i(t) +$$

$$\sum_t H_i(t)' B_i [q_{i,t+1}(\dot{\beta}_{i,t}) - q_{i,t+1}] B_i H_i(t) + \sum_t H_i(t)' B_i [q_{i,t+1}(\dot{\beta}_{i,t}) - q_{i,t+1}][B_i(t) - B_i] H_i(t) +$$

$$\sum_t H_i(t)' [B_i(t) - B_i] q_{i,t+1} B_i H_i(t) + \sum_t H_i(t)' [B_i(t) - B_i] q_{i,t+1} [B_i(t) - B_i] H_i(t) +$$

$$\sum_t H_i(t)' [B_i(t) - B_i][q_{i,t+1}(\dot{\beta}_{i,t}) - q_{i,t+1}] B_i H_i(t) +$$

$$\sum_t H_i(t)' [B_i(t) - B_i][q_{i,t+1}(\dot{\beta}_{i,t}) - q_{i,t+1}][B_i(t) - B_i] H_i(t) .$$

I must then show that the latter seven terms are each $o_p(1)$. I will show this for the final term, the others follow from similar arguments. The absolute value of the final term is less than or equal to

$$k^4 (\sup_t t^{1/4} \mid H(t) \mid)^2 (\sup_t t^{1/4} \mid B_i(t) - B_i \mid)^2 (P/R)(P^{-1} \sum_t \mid q_{i,t+1}(\dot{\beta}_{i,t}) - q_{i,t+1} \mid) .$$

Since by Assumption 6, $\pi < \infty$, and Lemma A1 implies both $\sup_t t^{1/4} |B_i(t) - B_i| = o_p(1)$ and $\sup_t t^{1/4} |H(t)| = o_p(1)$, the result will follow if $P^{-1} \sum_t | q_{i,t+1}(\dot{\beta}_{i,t}) - q_{i,t+1} | = o_p(1)$. But by Assumption 3

$$P^{-1} \sum_t | q_{i,t+1}(\dot{\beta}_{i,t}) - q_{i,t+1} | \le P^{-1} \sum_t m_t | \dot{\beta}_{i,t} - \beta_i^* |^{\varphi}$$

which in turn is less than or equal to $(\sup_t | \dot{\beta}_{i,t} - \beta_i^* |)^{\varphi} (P^{-1} \sum_t m_t) \le (\sup_t | \hat{\beta}_{i,t} - \beta_i^* |)^{\varphi} (P^{-1} \sum_t m_t)$. The result then follows from Lemma A1 and Markov's inequality .

Lemma A3: $\sum_t h_{i,t+1}' B_i(t) H_i(t) = \sum_t h_{i,t+1}' B_i H_i(t) + o_p(1)$.

Proof of Lemma A3: I will show this for the recursive scheme. The fixed and rolling schemes follow proofs similar to that for the recursive. Add and subtract B_i to get

$$\sum_t h_{i,t+1}' B_i(t) H_i(t) = \sum_t h_{i,t+1}' B_i H_i(t) + \sum_t h_{i,t+1}' (B_i(t) - B_i) H_i(t) =$$

$$\sum_t h_{i,t+1}' B_i H_i(t) + T^{-0.5} \sum_t \frac{T}{t} \mathrm{vec}[T^{0.5}(B_i(t) - B_i)]' [(T^{-0.5} h_{i,t+1}) \otimes (T^{-0.5} \sum_{j=1}^{t} h_{i,j})] .$$

Given Assumption 5 and Lemma A1 (f), Theorem 3.1 of Hansen (1992) implies

$\sum_t \frac{T}{t} \mathrm{vec}[T^{0.5}(B_i(t) - B_i)]' [(T^{-0.5} h_{i,t+1}) \otimes (T^{-0.5} \sum_{j=1}^{t} h_{i,j})] = O_p(1)$. Since $T^{-0.5} = o(1)$ the result follows.

Lemma A4: $\sum_t H_i(t)' B_i q_{i,t+1} B_i H_i(t) = \sum_t H_i(t)' B_i H_i(t) + o_p(1)$.

Proof of Lemma A4: I will show this for the recursive scheme. The fixed and rolling schemes follow proofs similar to that for the recursive. In order to simplify notation let $B_i = I$ and $\bar{q}_{i,t+1} = q_{i,t+1} - I$. Add and subtract B_i^{-1} to get

$$\sum_t H_i(t)' q_{i,t+1} H_i(t) = \sum_t H_i(t)' H_i(t) + \sum_t H_i(t)' \bar{q}_{i,t+1} H_i(t)$$

$$= \sum_t H_i(t)' H_i(t) + \sum_t (H_i(t)' \otimes H_i(t)') \mathrm{vec}[\bar{q}_{i,t+1}]$$

$$= \sum_t H_i(t)' H_i(t) \; + \; T^{-0.5} \sum_t (\frac{T}{t})^2 ((T^{-0.5} \sum_{j=1}^{t} h_{i,j})' \otimes (T^{-0.5} \sum_{j=1}^{t} h_{i,j})')(T^{-0.5} vec[\bar{q}_{i,t+1}]) \; .$$

That $T^{-0.5} \sum_t (\frac{T}{t})^2 ((T^{-0.5} \sum_{j=1}^{t} h_{i,j})' \otimes (T^{-0.5} \sum_{j=1}^{t} h_{i,j})')(T^{-0.5} vec[\bar{q}_{i,t+1}]) = O_p(1)$ follows from Assumption 5 and Theorem 3.1 of Hansen (1992). Since $T^{-0.5} = o(1)$ the result follows.

Lemma A5: For $s \in [\lambda, 1]$, (a) $T^{-1/2} \sum_{j=1}^{t} \tilde{h}_j \Rightarrow W(s)$, (b) $(\frac{T}{t}) T^{-1/2} \sum_{j=1}^{t} \tilde{h}_j \Rightarrow s^{-1} W(s)$, (c) $(\frac{T}{R}) T^{-1/2} \sum_{j=t-R+1}^{t} \tilde{h}_j \Rightarrow \lambda^{-1} \{W(s) - W(s-\lambda)\}$.

Proof of Lemma A5: (a) Given Assumptions 4 and 5 and the fact that $T^{-1} E \sum_{j=1}^{T} \tilde{h}_j \tilde{h}_j' \to I_{k_2 \times k_2} < \infty$ the result follows from Theorem 2.1 of Hansen (1992). (b) Given (a) the result follows from the Continuous Mapping Theorem. (c) That $\frac{T}{R} \to \lambda^{-1}$ is immediate. Write $T^{-1/2} \sum_{j=t-R+1}^{t} \tilde{h}_j$ as $T^{-1/2} \sum_{j=1}^{t} \tilde{h}_j - T^{-1/2} \sum_{j=1}^{t-R} \tilde{h}_j$. That $T^{-1/2} \sum_{j=1}^{t} \tilde{h}_j \Rightarrow W(s)$ follows from (a). For the second piece, if I simply define $s' = s - \lambda$ then Theorem 2.1 of Hansen (1992) implies $T^{-1/2} \sum_{j=1}^{t-R} \tilde{h}_j \Rightarrow W(s')$ (even though we're only interested in $s' \in [0, \pi\lambda]$, or $s \in [\lambda, 1]$).

Lemma A6: $\sum_t \tilde{H}'(t) \tilde{h}_{t+1} \to_d \chi_1$ where χ_1 equals

$\int_{\lambda}^{1} s^{-1} W'(s) dW(s)$	for the recursive scheme,
$\lambda^{-1} \{W(1) - W(\lambda)\}' W(\lambda)$	for the fixed scheme,
$\lambda^{-1} \int_{\lambda}^{1} \{W(s) - W(s-\lambda)\}' dW(s)$	for the rolling scheme.

Proof of Lemma A6: The result is simple for the fixed. For the recursive and rolling write $\sum_t \tilde{H}'(t) \tilde{h}_{t+1}$ as $\sum_t \frac{T}{t} (T^{-0.5} \sum_{j=1}^{t} \tilde{h}_j)' (T^{-0.5} \tilde{h}_{t+1})$ and $\frac{T}{R} \sum_t (T^{-0.5} \sum_{j=1}^{t} \tilde{h}_j - T^{-0.5} \sum_{j=1}^{t-R} \tilde{h}_j)' (T^{-0.5} \tilde{h}_{t+1})$ respectively. Given Assumptions 4, 5 and Lemma A5 the results follow from Theorem 3.1 of Hansen (1992).

Lemma A7: $\sum_t \tilde{H}'(t) \tilde{H}(t) \to_d \chi_2$ where χ_2 equals

$\int_{\lambda}^{1} s^{-2} W'(s) W(s) ds$	for the recursive scheme,
$\pi \lambda^{-1} W'(\lambda) W(\lambda)$	for the fixed scheme,
$\lambda^{-2} \int_{\lambda}^{1} \{W(s) - W(s-ë)\}' \{W(s) - W(s-ë)\} ds$	for the rolling scheme.

Proof of Lemma A7: The result is immediate for the fixed. Given Lemma A5 the results for the recursive and rolling follow from the Continuous Mapping Theorem.

Lemma A8: Both $\sum_t\{-H(t)'J'B_1(t)q_{1,t+1}(\dot{\beta}_{1,t})B_1(t)JH(t)+H(t)'B(t)q_{t+1}(\dot{\beta}_{2,t})B(t)H(t)\}^2$ and $\sum_t\{-H(t)'J'B_1(t)q_{1,t+1}(\dot{\beta}_{1,t})B_1(t)JH(t)+H(t)'B(t)q_{t+1}(\dot{\beta}_{2,t})B(t)H(t)\}\{-h_{t+1}'J'B_1(t)JH(t)+h_{t+1}'B(t)H(t)\}$ are $o_p(1)$.

Proof of Lemma A8: The proof for each is largely the same. I will show the result for the first term. Immediately we know the first term is less than or equal to

$$\sup_t\{-P^{0.25}H(t)'J'B_1(t)q_{1,t+1}(\dot{\beta}_{1,t})B_1(t)JP^{0.25}H(t)+P^{0.25}H(t)'B(t)q_{t+1}(\dot{\beta}_{2,t})B(t)P^{0.25}H(t)\}^2.$$

Using the triangle inequality the above expression is less than or equal to

$$\begin{aligned}&\sup_t|P^{0.25}H(t)'J'B_1(t)q_{1,t+1}(\dot{\beta}_{1,t})B_1(t)JP^{0.25}H(t)|^2 + \\ &\quad\sup_t|P^{0.25}H(t)'B(t)q_{t+1}(\dot{\beta}_{2,t})B(t)P^{0.25}H(t)|^2 + \\ &\quad 2\{\sup_t|P^{0.25}H(t)'J'B_1(t)q_{1,t+1}(\dot{\beta}_{1,t})B_1(t)JP^{0.25}H(t)|\times \\ &\qquad\sup_t|P^{0.25}H(t)'B(t)q_{t+1}(\dot{\beta}_{2,t})B(t)P^{0.25}H(t)|\}\end{aligned} \tag{5}$$

It then suffices to show both $\sup_t|P^{0.25}H(t)'J'B_1(t)q_{1,t+1}(\dot{\beta}_{1,t})B_1(t)JP^{0.25}H(t)|$ and $\sup_t|P^{0.25}H(t)'B(t)q_{t+1}(\dot{\beta}_{2,t})B(t)P^{0.25}H(t)|$ are $o_p(1)$. I will do so for the latter. Adding and subtracting both B and q_{t+1} we immediately know that

$$\begin{aligned}&\sup_t|P^{0.25}H(t)'B(t)q_{t+1}(\dot{\beta}_{2,t})B(t)P^{0.25}H(t)|\leq \\ &\quad k^4(\sup_t|P^{0.25}H(t)|)^2(\sup_t|B(t)-B|)^2\sup_t|q_{t+1}(\dot{\beta}_{2,t})-q_{t+1}| + \\ &\quad 2k^4(\sup_t|P^{0.25}H(t)|)^2\sup_t|B(t)-B||B|\sup_t|q_{t+1}(\dot{\beta}_{2,t})-q_{t+1}| + \\ &\quad k^4(\sup_t|P^{0.25}H(t)|)^2|B|^2\sup_t|q_{t+1}(\dot{\beta}_{2,t})-q_{t+1}| + \\ &\quad k^4(\sup_t|P^{0.25}H(t)|)^2(\sup_t|B(t)-B|)^2\sup_t|q_{t+1}| + \\ &\quad 2k^4(\sup_t|P^{0.25}H(t)|)^2\sup_t|B(t)-B||B|\sup_t|q_{t+1}| + \\ &\quad k^4(\sup_t|P^{0.25}H(t)|)^2|B|^2\sup_t|q_{t+1}|.\end{aligned}$$

That $\sup_t|P^{0.25}H(t)|$ and $\sup_t|B(t)-B|$ are $o_p(1)$ follows from Lemma A1. That $\sup_t|q_{t+1}| = O_p(1)$ follows from Assumption 4. To complete the proof notice that by Assumption 3 $\sup_t|q_{t+1}(\dot{\beta}_{2,t})-q_{t+1}|\leq$ $(\sup_t|m_t|)(\sup_t|\dot{\beta}_{2,t}-\beta_2^*|)^{\varphi}\leq(\sup_t|m_t|)(\sup_t|\hat{\beta}_{2,t}-\beta_2^*|)^{\varphi}$. The desired result follows since $\sup_t|m_t|$ $= O_p(1)$ by Assumption 3 and $\sup_t|\hat{\beta}_{2,t}-\beta_2^*| = o_p(1)$ by Lemma A1.

Lemma A9: $\sum_t\{-h_{t+1}'J'B_1(t)JH(t)+h_{t+1}'B(t)H(t)\}^2 = \sum_t\{-h_{t+1}'J'B_1JH(t)+h_{t+1}'BH(t)\}^2 + o_p(1)$.

Proof of Lemma A9: Add and subtract both B_1 and B to obtain

$$\sum_t\{-h_{t+1}'J'B_1(t)JH(t)+h_{t+1}'B(t)H(t)\}^2 = \sum_t\{-h_{t+1}'J'B_1JH(t)+h_{t+1}'BH(t)\}^2 + \tag{6}$$

$$\sum_t \{-h'_{t+1}J'(B_1(t)-B_1)JH(t)+h'_{t+1}(B(t)-B)H(t)\}^2 \ +$$

$$2\sum_t \{-h'_{t+1}J'B_1JH(t)+h'_{t+1}BH(t)\}\{-h'_{t+1}J'(B_1(t)-B_1)JH(t)+h'_{t+1}(B(t)-B)H(t)\}.$$

It then must be shown that the second and third terms on the r.h.s. of (6) are $o_p(1)$. I will do so for the second term. The third follows from a similar argument. The second term is less than or equal to

$$P\sup_t \{-h'_{t+1}J'(B_1(t)-B_1)JH(t)+h'_{t+1}(B(t)-B)H(t)\}^2$$

$$\leq \{\sup_t P^{0.5}\,|-h'_{t+1}J'(B_1(t)-B_1)JH(t)+h'_{t+1}(B(t)-B)H(t)|\}^2$$

$$\leq k^4 \sup_t |h_{t+1}||J|^2 (\sup_t P^{0.25}|B_1(t)-B_1|)(\sup_t P^{0.25}|H(t)|) \ +$$

$$k^3 \sup_t |h_{t+1}|(\sup_t P^{0.25}|B(t)-B|)(\sup_t P^{0.25}|H(t)|).$$

By Assumption 5 $\sup_t|h_{t+1}| = O_p(1)$. The result then follows from Lemma A1 since both $\sup_t P^{0.25}|B(t)-B|$ and $\sup_t P^{0.25}|H(t)|$ are $o_p(1)$.

Lemma A10: $\sum_t\{-h'_{t+1}J'B_1JH(t)+h'_{t+1}BH(t)\}^2 \ = \ c^2\sum_t\{\tilde{H}(t)\tilde{h}_{t+1}\}^2$.

Proof of Lemma A10: The result is immediate since $c^{-0.5}A'CB^{1/2}h_t = \tilde{h}_t$ and $c^{-0.5}A'CB^{1/2}H(t) = \tilde{H}(t)$.

Lemma A11: $\sum_t\{\tilde{H}(t)\tilde{h}_{t+1}\}^2 \ \to_d \chi_2$ for χ_2 defined in Lemma A7.

Proof of Lemma A11: Define $\gamma_{t+1} \equiv \tilde{h}_{t+1}\tilde{h}'_{t+1} - I$[11]. Then $\sum_t\{\tilde{H}(t)\tilde{h}_{t+1}\}^2 \ = \ \sum_t \tilde{H}'(t)\tilde{H}(t) \ +$ $\sum_t(\tilde{H}'(t)\otimes\tilde{H}'(t))\mathrm{vec}(\gamma_{t+1})$. Given Lemma A7 it suffices to show that the latter r.h.s. term is $o_p(1)$. I will do so for the recursive scheme, the others follow similar arguments. By definition the latter term is equal to

$$T^{-0.5}\sum_t (\frac{T}{t})^2\{(T^{-0.5}\sum_{j=1}^t \tilde{h}_{t+1})'\otimes(T^{-0.5}\sum_{j=1}^t \tilde{h}_{t+1})'\}(T^{-0.5}\mathrm{vec}(\gamma_{t+1})). \qquad (7)$$

That (7) is $O_p(1)$ follows from Assumption 5 and Theorem 3.1 of Hansen (1992). The result then follows from the fact that $T^{-0.5}$ is $o(1)$.

Lemma A12: $\sum_t u_{t+1}g'_{i,\beta,t+1}(\dot{\beta}_{i,t})B_i(t)H_i(t) \ = \ \sum_t h'_{i,t+1}B_iH_i(t) \ + o_p(1)$.

Proof of Lemma A12: If we add and subtract B_i, and $g_{i,\beta,t+1} \equiv g_{i,\beta,t+1}(\beta_i^*)$ we have

$$\sum_t u_{t+1}g'_{i,\beta,t+1}(\dot{\beta}_{i,t})B_i(t)H_i(t) \ = \ \sum_t h'_{i,t+1}B_iH_i(t) \ + \ \sum_t h'_{i,t+1}(B_i(t)-B_i)H_i(t) \ +$$

$$\sum_t u_{t+1}(g'_{i,\beta,t+1}(\dot{\beta}_{i,t})-g'_{i,\beta,t+1})(B_i(t)-B_i)H_i(t) \ + \ \sum_t u_{t+1}(g'_{i,\beta,t+1}(\dot{\beta}_{i,t})-g'_{i,\beta,t+1})B_iH_i(t).$$

That the second r.h.s. term is $o_p(1)$ follows from the proof of Lemma A3. We then need only show that the latter two terms are $o_p(1)$. I will do so for the final term. The proof of the other is similar. Let

[11] I'd like to thank Bruce Hansen for suggesting this approach.

$g_{i,\beta\beta,t+1}(\beta_i) \equiv \frac{\partial}{\partial \beta_i} g_{i,\beta,t+1}(\beta_i)$. If we take a first order Taylor expansion of $g_{i,\beta,t+1}(\beta_i)$ around β_i^* and add and subtract $g_{i,\beta\beta,t+1} = g_{i,\beta\beta,t+1}(\beta_i^*)$ we then have

$$\Sigma_t u_{t+1}(g^{'}_{i,\beta,t+1}(\dot{\beta}_{i,t}) - g^{'}_{i,\beta,t+1})B_i H_i(t) = \Sigma_t (\dot{\beta}_{i,t} - \beta_i^*)' g_{i,\beta\beta,t+1}(\ddot{\beta}_{i,t})B_i H_i(t) u_{t+1} = \quad (8)$$

$$P^{-1}\Sigma_t T^{0.5}(\dot{\beta}_{i,t} - \beta_i^*)'(g_{i,\beta\beta,t+1}(\ddot{\beta}_{i,t}) - g_{i,\beta\beta,t+1})B_i \frac{P}{t}(T^{-0.5}\Sigma_{s=1}^t h_{i,s})u_{t+1} +$$

$$T^{-0.5}\Sigma_t T^{0.5}(\dot{\beta}_{i,t} - \beta_i^*)'(T^{-0.5}u_{t+1}g_{i,\beta\beta,t+1})B_i \frac{T}{t}(T^{-0.5}\Sigma_{s=1}^t h_{i,s}).$$

To show that the first r.h.s. term in (8) is $o_p(1)$ take its absolute value. It is then less than or equal to

$$\frac{P}{R}|B_i|(\sup_t T^{0.5}|\dot{\beta}_{i,t} - \beta_i^*|)(\sup_t |g_{i,\beta\beta,t+1}(\ddot{\beta}_{i,t}) - g_{i,\beta\beta,t+1}|)(\sup_t |T^{-0.5}\Sigma_{s=1}^t h_{i,s}|)(|\sup_t |u_{t+1}|)$$

That $\sup_t |u_{t+1}|$ and $\sup_t |T^{-0.5}\Sigma_{s=1}^t h_{i,s}|$ are $O_p(1)$ follows from Assumption 4 and Theorem 2.1 of Hansen (1992) respectively. Let's now show that $\sup_t T^{0.5}|\dot{\beta}_{i,t} - \beta_i^*|$ is $O_p(1)$. Notice that $\sup_t T^{0.5}|\dot{\beta}_{i,t} - \beta_i^*| \leq \sup_t T^{0.5}|\hat{\beta}_{i,t} - \beta_i^*| \leq \frac{T}{R}k(\sup_t |B_i(t) - B_i|)(\sup_t |T^{-0.5}\Sigma_{s=1}^t h_{i,s}|) + \frac{T}{R}k|B_i|(\sup_t |T^{-0.5}\Sigma_{s=1}^t h_{i,s}|)$. The desired result follows since $\sup_t |T^{-0.5}\Sigma_{s=1}^t h_{i,s}|$ is $O_p(1)$ by Theorem 2.1 of Hansen (1992), $\sup_t |B_i(t) - B_i|$ is $o_p(1)$ by Lemma A1 and T/R is bounded. That the first r.h.s. term in (8) is $o_p(1)$ then follows since by Assumption 3 and Lemma A1 $\sup_t |g_{i,\beta\beta,t+1}(\ddot{\beta}_{i,t}) - g_{i,\beta\beta,t+1}| \leq (\sup_t |m_t|)(\sup_t |\ddot{\beta}_{i,t} - \beta_i^*|) \leq (\sup_t |m_t|)(\sup_t |\hat{\beta}_{i,t} - \beta_i^*|) = o_p(1)$.

To show that the second term on the r.h.s. of (8) is $o_p(1)$ first note that $vec[u_{t+1}g_{i,\beta\beta,t+1}]$ forms a martingale difference sequence and by the previous paragraph $\sup_t T^{0.5}|\dot{\beta}_{i,t} - \beta_i^*|$ is $O_p(1)$. Hence by Theorem 3.1 of Hansen (1992) $\Sigma_t \frac{T}{t}\{(T^{-0.5}\Sigma_{s=1}^t h^{'}_{i,s}B_i) \otimes T^{0.5}(\dot{\beta}_{i,t} - \beta_i^*)'\}vec[T^{-0.5}u_{t+1}g_{i,\beta\beta,t+1}]$ is $O_p(1)$. The result then follows since $T^{-0.5}$ is $o(1)$.

Lemma A13: Let $g_{i,\beta,t+1} \equiv g_{i,\beta,t+1}(\beta_i^*)$. For i,j = 1,2, $\Sigma_t H^{'}_i(t)B_i(t)g_{i,\beta,t+1}(\dot{\beta}_{i,t})g^{'}_{j,\beta,t+1}(\dot{\beta}_{j,t})B_j(t)H_j(t) = \Sigma_t H^{'}_i(t)B_i E(g_{i,\beta,t+1}g^{'}_{j,\beta,t+1})B_j H_j(t) + o_p(1)$.

Proof of Lemma A13: Follows from arguments almost identical to those in Lemmas A2 and A4.

Lemma A14: Let $g_{i,\beta,t+1} \equiv g_{i,\beta,t+1}(\beta_i^*)$. $\Sigma_t(u_{1,t+1}(\hat{\beta}_{1,t}) - u_{2,t+1}(\hat{\beta}_{2,t}))^2 \to_d c\chi_2$.

Proof of Lemma A14: If we take first order Taylor expansions of both $u_{1,t+1}(\hat{\beta}_{1,t})$ and $u_{2,t+1}(\hat{\beta}_{2,t})$ around β_1^* and β_2^* respectively, we have $\Sigma_t(u_{1,t+1}(\hat{\beta}_{1,t}) - u_{2,t+1}(\hat{\beta}_{2,t}))^2 =$

$$\Sigma_t H^{'}_1(t)B_1(t)g_{1,\beta,t+1}(\dot{\beta}_{1,t})g^{'}_{1,\beta,t+1}(\dot{\beta}_{1,t})B_1(t)H_1(t) +$$

$$\sum_t H_2'(t)B_2(t)g_{2,\beta,t+1}(\dot{\beta}_{2,t})g_{2,\beta,t+1}'(\dot{\beta}_{2,t})B_2(t)H_2(t) \ -$$

$$2\sum_t H_1'(t)B_1(t)g_{1,\beta,t+1}(\dot{\beta}_{1,t})g_{2,\beta,t+1}'(\dot{\beta}_{2,t})B_2(t)H_2(t)\,.$$

But by Lemma A13 this equals

$$\sum_t H'(t)J'B_1JH(t) \ + \ \sum_t H'(t)BH(t) \ - 2\sum_t H'(t)J'B_1Eg_{1,\beta,t+1}g_{2,\beta,t+1}'BH(t) \ + o_p(1).$$

The result then follows from Lemma 3.1, Lemma A7 and the fact that $Jg_{2,\beta,t+1} = g_{1,\beta,t+1}$.

Proof of Lemma 3.1: (a) For i,j = 1,2, let $q_{t,i,j}$ represent the ij-block of the matrix q_t. Since

$$B_2 = \begin{pmatrix} B_1(I+[Eq_{t,1,2}]N_2[Eq_{t,2,1}]B_1^{-1}) & -B_1^{-1}[Eq_{t,1,2}]N_2 \\ -N_2[Eq_{t,2,1}]B_1^{-1} & N_2 \end{pmatrix}$$

$$N_2 = [Eq_{t,2,2} - Eq_{t,2,1}B_1Eq_{t,1,2}]^{-1}$$

and

$$M = \begin{pmatrix} B_1[Eq_{t,1,2}]N_2[Eq_{t,2,1}]B_1 & -B_1[Eq_{t,1,2}]N_2 \\ -N_2[Eq_{t,2,1}]B_1 & N_2 \end{pmatrix}$$

the result follows from straightforward algebra. (b) Since Q is idempotent the result follows from Schur's Decomposition Theorem (p.16, Magnus and Neudecker, 1988).

Proof of Lemma 3.2: If we take second order Taylor expansions of both $L_{1,t+1}(\hat{\beta}_{1,t})$ and $L_{2,t+1}(\hat{\beta}_{2,t})$ around β_1^* and β_2^* respectively, we have

$$\sum_t[L_{1,t+1}(\hat{\beta}_{1,t}) - L_{2,t+1}(\hat{\beta}_{2,t})] = \sum_t\{-h_{t+1}'J'B_1(t)JH(t) + h_{t+1}'B(t)H(t)\} \ -$$

$$(0.5)\sum_t\{-H(t)'J'B_1(t)q_{1,t+1}(\dot{\beta}_{1,t})B_1(t)JH(t) + H(t)'B(t)q_{t+1}(\dot{\beta}_{2,t})B(t)H(t)\}$$

for $\dot{\beta}_{i,t}$ on the line between $\hat{\beta}_{i,t}$ and β_i^* respectively. The result then follows from from Lemmas A1-A4 and Lemma 3.1.

Proof of Lemma 3.3: If we square the argument of the summation in the proof of Lemma 3.2, this follows from Lemmas A8-A11.

Proof of Theorem 3.1: Given Lemma 3.2, the result follows from Lemmas A5-A7.

Proof of Theorem 3.2: Given Lemma 3.3, Lemma A11 and Theorem 3.1, the result follows from Theorem 2.1 of Hansen (1992) and the Continuous Mapping Theorem.

Proof of Theorem 3.3: Given Theorem 1 and the Continuous Mapping Theorem it suffices to show $a_{1,T}a_{2,T} - a_{0,T}^2 \to_d 4c^2\chi_2$ for χ_2 defined in Lemma A7. That $a_{2,T} \to_p 4c$ follows from Theorem 4.1 of

West (1996). To show that $Pa_{0,T}^2 = o_p(1)$ note that from Lemmas A10-A13, $Pa_{0,T} = O_p(1)$. The result follows since by Lemma A14, $Pa_{1,T} \rightarrow_d c\chi_2$.

Proof of Corollary 3.1: Follows immediately from Theorem 3.1.

Proof of Theorem 4.1: Define $u_{T,t} \equiv \frac{\sigma_u Z_{22,t}^{'}\beta_{22}^{*}}{T^{0.5}} + u_t$, $h_{T,t} \equiv u_{T,t}Z_t$, $\tilde{Z}_t \equiv \sigma_u^{-1}A^{'}CB_2^{0.5}Z_t$ and $\tilde{h}_{T,t} \equiv u_{T,t}\tilde{Z}_t$. Define $H_T(t)$ and $\tilde{H}_T(t)$ as in section 3.3 but relative to $h_{T,t}$ and $\tilde{h}_{T,t}$ respectively.

With these definitions in place, and keeping in mind that the nested models are linear and estimated by OLS, it is straightforward to show that Lemmas A1-A4, A8-A10, A12, A13 and Lemmas 3.1-3.3 continue to hold but with u_t, h_t, $\tilde{h}_t$, H(t) and $\tilde{H}(t)$ replaced by $u_{T,t}$, $h_{T,t}$, $\tilde{h}_{T,t}$, $H_T(t)$ and $\tilde{H}_T(t)$ respectively. Hence we know that

$$\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2] =$$

$$\sigma_u^2[2\sum_{t=R}^{T}(T^{0.5}\tilde{H}_T(t))^{'}(T^{-0.5}\tilde{h}_{T,t+1}) - T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}_T(t))^{'}(T^{0.5}\tilde{H}_T(t))] + o_p(1), \quad (9)$$

$$\sum_{t=R}^{T}[\hat{u}_{1,t+1}^2 - \hat{u}_{2,t+1}^2]^2 =$$

$$4\sigma_u^4 T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}_T(t))^{'}(T^{0.5}\tilde{H}_T(t)) + o_p(1), \quad (10)$$

and

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

$$4\sigma_u^4 T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}_T(t))^{'}(T^{0.5}\tilde{H}_T(t)) + o_p(1). \quad (11)$$

Note that the rhs of both (10) and (11) are identical. Hence all we need to derive is the limiting behavior of the rhs's of both (9) and (10) in order to reach the desired result. Applying the definition of $u_{T,t}$ we then know that

$$\sum_{t=R}^{T}(T^{0.5}\tilde{H}_T(t))^{'}(T^{-0.5}\tilde{h}_{T,t+1}) =$$

$$\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t))^{'}(T^{-0.5}\tilde{h}_{t+1}) + \sum_{t=R}^{T}(t^{-1}\sum_{s=1}^{t}\sigma_u\beta_{22}^{*'}Z_{22,s}\tilde{Z}_s^{'})(T^{-0.5}\tilde{h}_{t+1}) +$$

$$T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t))^{'}(\tilde{Z}_{t+1}Z_{22,t+1}^{'}\beta_{22}^{*}\sigma_u) + T^{-1}\sum_{t=R}^{T}(t^{-1}\sum_{s=1}^{t}\sigma_u\beta_{22}^{*'}Z_{22,s}\tilde{Z}_s^{'})(\tilde{Z}_{t+1}Z_{22,t+1}^{'}\beta_{22}^{*}\sigma_u)$$

and

$$T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}_T(t))^{'}(T^{0.5}\tilde{H}_T(t))] =$$

$$T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t))^{'}(T^{0.5}\tilde{H}(t)) + 2T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t))^{'}(t^{-1}\sum_{s=1}^{t}\tilde{Z}_s Z_{22,s}^{'}\beta_{22}^{*}\sigma_u) +$$

$$T^{-1}\sum_{t=R}^{T}(t^{-1}\sum_{s=1}^{t}\sigma_u\beta_{22}^{*'}Z_{22,s}\tilde{Z}_s^{'})(t^{-1}\sum_{s=1}^{t}\tilde{Z}_s Z_{22,s}^{'}\beta_{22}^{*}\sigma_u).$$

Using arguments akin to those in Lemmas A1-A4 it can then be shown that

$$\sum_{t=R}^{T}(T^{0.5}\tilde{H}_T(t))^{'}(T^{-0.5}\tilde{h}_{T,t+1}) =$$

$$\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t))'(T^{-0.5}\tilde{h}_{t+1}) + \beta_2^{*'}J_2B_2^{-0.5}CA(T^{-0.5}\sum_{t=R}^{T}\tilde{h}_{t+1}) +$$

$$\beta_2^{*'}J_2B_2^{-0.5}CA(T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t)) + \pi\lambda\beta_2^{*'}J_2B_2^{-0.5}QB_2^{-0.5}J_2'\beta_2^{*'} + o_p(1) \tag{12}$$

and

$$T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}_T(t))'(T^{0.5}\tilde{H}_T(t))] = T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t))'(T^{0.5}\tilde{H}(t)) +$$

$$2\beta_2^{*'}J_2B_2^{-0.5}CA(T^{-1}\sum_{t=R}^{T}(T^{0.5}\tilde{H}(t)) + \pi\lambda\beta_2^{*'}J_2B_2^{-0.5}QB_2^{-0.5}J_2'\beta_2^{*'} + o_p(1). \tag{13}$$

If we substitute (12) and (13) back into (9) and (10) the results follows from Theorem 2.1 of Hansen (1992) and the Continuous Mapping Theorem.

Table 1

Percentiles of the OOS-t statistic: Recursive scheme

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

Table 2

Percentiles of the OOS-t statistic: Rolling scheme

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

Table 3

Percentiles of the OOS-t statistic: Fixed scheme

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

Table 4
Percentiles of the (modified) OOS-F statistic: Recursive scheme

$k_2\backslash\pi$		0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	2.129	2.768	3.179	3.459	3.584	3.771	3.589	3.838	3.882	3.951
	(0.95)	1.038	1.298	1.554	1.567	1.548	1.583	1.623	1.599	1.553	1.518
	(0.90)	0.659	0.814	0.796	0.798	0.751	0.759	0.698	0.685	0.687	0.616
2	(0.99)	2.691	3.426	3.907	4.129	4.200	4.362	4.304	4.309	4.278	4.250
	(0.95)	1.453	1.733	1.891	1.820	1.802	1.819	1.752	1.734	1.692	1.706
	(0.90)	0.912	1.029	1.077	1.008	0.880	0.785	0.697	0.666	0.587	0.506
3	(0.99)	3.092	4.080	4.136	4.322	4.341	4.337	4.192	4.089	4.365	4.184
	(0.95)	1.710	2.062	2.073	1.978	1.909	1.930	1.795	1.715	1.710	1.612
	(0.90)	1.064	1.117	1.121	0.960	0.857	0.691	0.599	0.386	0.276	0.127
4	(0.99)	3.440	4.541	4.609	4.378	4.202	4.586	4.477	4.337	4.247	4.096
	(0.95)	1.964	2.246	2.194	1.900	1.809	1.578	1.376	1.256	1.122	1.029
	(0.90)	1.225	1.313	1.184	0.829	0.545	0.354	0.197	-0.058	-0.234	-0.456
5	(0.99)	3.673	4.466	4.434	4.249	4.351	4.349	4.187	3.945	3.783	3.783
	(0.95)	2.082	2.235	2.242	1.773	1.449	1.316	1.045	0.718	0.502	0.459
	(0.90)	1.277	1.228	0.958	0.614	0.241	-0.099	-0.361	-0.656	-0.820	-1.072
6	(0.99)	3.846	4.545	4.676	4.637	4.703	4.286	4.144	3.981	3.525	3.321
	(0.95)	2.124	2.217	2.121	1.660	1.360	1.181	0.761	0.413	0.299	-0.109
	(0.90)	1.313	1.164	0.890	0.419	-0.044	-0.405	-0.776	-1.072	-1.395	-1.664
7	(0.99)	4.098	4.508	4.419	4.271	4.312	4.150	3.677	3.155	3.090	2.880
	(0.95)	2.239	2.424	2.057	1.604	1.282	0.928	0.378	-0.008	-0.199	-0.591
	(0.90)	1.333	1.118	0.799	0.242	-0.363	-0.728	-1.194	-1.657	-2.033	-2.507
8	(0.99)	4.130	4.645	4.625	4.202	4.147	3.912	3.185	2.933	2.952	2.484
	(0.95)	2.312	2.373	1.895	1.390	0.943	0.587	0.131	-0.372	-0.680	-1.140
	(0.90)	1.369	1.058	0.552	0.014	-0.632	-1.076	-1.633	-2.174	-2.731	-3.160
9	(0.99)	4.388	4.703	4.873	4.122	4.066	3.753	3.027	2.925	2.802	2.186
	(0.95)	2.440	2.219	1.714	1.286	0.631	0.198	-0.356	-0.851	-1.241	-1.696
	(0.90)	1.432	0.920	0.393	-0.327	-1.007	-1.595	-2.229	-2.666	-3.250	-3.794
10	(0.99)	4.433	4.813	4.718	3.944	3.645	3.194	2.578	2.282	2.152	1.436
	(0.95)	2.489	2.157	1.536	1.055	0.205	-0.431	-1.071	-1.459	-1.988	-2.378
	(0.90)	1.401	0.884	0.155	-0.600	-1.341	-2.008	-2.782	-3.348	-3.839	-4.437

Table 5
Percentiles of the (modified) OOS-F statistic: Rolling scheme

$k_2 \backslash \pi$		0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	2.230	2.812	3.300	3.634	3.811	3.688	3.721	3.924	3.612	3.765
	(0.95)	1.112	1.394	1.644	1.627	1.583	1.574	1.469	1.488	1.378	1.215
	(0.90)	0.667	0.838	0.865	0.773	0.693	0.602	0.482	0.390	0.355	0.276
2	(0.99)	2.819	3.544	3.988	4.066	4.398	4.403	4.109	4.293	4.046	3.566
	(0.95)	1.481	1.802	1.889	1.841	1.695	1.495	1.264	1.015	0.783	0.504
	(0.90)	0.920	1.028	1.004	0.806	0.468	0.399	0.095	-0.198	-0.394	-0.623
3	(0.99)	3.128	4.135	4.120	4.264	4.519	4.386	4.123	3.373	3.089	2.685
	(0.95)	1.752	2.089	2.042	1.700	1.532	1.100	0.694	0.340	-0.071	-0.471
	(0.90)	1.074	1.135	0.944	0.617	0.224	-0.174	-0.600	-1.080	-1.529	-1.847
4	(0.99)	3.649	4.474	4.586	4.432	4.459	4.296	3.621	2.905	2.337	1.699
	(0.95)	2.078	2.332	1.979	1.536	1.228	0.701	0.116	-0.491	-1.112	-1.487
	(0.90)	1.284	1.263	0.777	0.356	-0.200	-0.788	-1.341	-1.973	-2.528	-3.182
5	(0.99)	3.721	4.504	4.710	4.508	4.199	4.042	3.216	2.167	1.370	1.055
	(0.95)	2.191	2.164	1.783	1.175	0.764	-0.121	-0.542	-1.454	-2.172	-2.765
	(0.90)	1.315	1.143	0.541	-0.114	-0.774	-1.583	-2.300	-3.102	-3.896	-4.649
6	(0.99)	4.093	4.532	4.786	4.456	3.899	3.473	2.324	1.500	0.615	0.159
	(0.95)	2.181	2.117	1.652	1.062	0.318	-0.745	-1.424	-2.302	-3.162	-4.256
	(0.90)	1.311	0.981	0.365	-0.428	-1.348	-2.392	-3.270	-4.292	-5.180	-6.114
7	(0.99)	4.340	4.568	4.566	4.181	3.450	3.167	1.696	0.442	-0.300	-1.151
	(0.95)	2.273	2.321	1.478	0.868	-0.076	-1.162	-2.243	-3.224	-4.261	-5.620
	(0.90)	1.355	0.939	0.151	-0.783	-1.833	-3.061	-4.153	-5.487	-6.474	-7.583
8	(0.99)	4.537	4.707	4.681	4.041	3.065	2.488	1.013	-0.144	-1.562	-2.729
	(0.95)	2.436	2.169	1.213	0.468	-0.626	-1.885	-3.233	-4.430	-5.681	-6.991
	(0.90)	1.357	0.886	-0.186	-1.278	-2.492	-3.824	-5.225	-6.556	-7.690	-8.939
9	(0.99)	4.438	4.711	4.443	3.754	2.622	1.723	0.327	-0.877	-2.543	-3.923
	(0.95)	2.518	1.966	1.075	0.138	-1.113	-2.620	-4.036	-5.258	-6.931	-8.345
	(0.90)	1.433	0.772	-0.521	-1.835	-3.139	-4.586	-6.133	-7.734	-9.173	-10.558
10	(0.99)	4.413	4.815	4.589	3.460	2.145	1.121	-0.624	-2.313	-3.795	-5.166
	(0.95)	2.520	1.893	0.730	-0.235	-1.733	-3.496	-4.940	-6.512	-8.255	-9.863
	(0.90)	1.411	0.555	-0.701	-2.189	-3.790	-5.566	-7.200	-9.046	-10.574	-12.294

Table 6
Percentiles of the (modified) OOS-F statistic: Fixed scheme

$k_2\backslash\pi$		0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
1	(0.99)	1.981	2.681	3.055	3.230	3.377	3.562	3.619	3.816	3.812	3.838
	(0.95)	1.015	1.345	1.534	1.677	1.667	1.738	1.807	1.812	1.857	1.862
	(0.90)	0.649	0.835	0.885	0.933	0.964	1.009	0.986	1.050	1.080	1.037
2	(0.99)	2.554	3.241	3.514	3.944	4.019	4.173	4.364	4.251	4.556	4.414
	(0.95)	1.421	1.765	1.999	2.077	2.116	2.169	2.232	2.275	2.260	2.195
	(0.90)	0.914	1.140	1.237	1.299	1.268	1.330	1.250	1.126	1.189	1.151
3	(0.99)	2.985	3.854	4.103	4.272	4.233	4.549	4.764	4.687	4.915	4.900
	(0.95)	1.653	2.050	2.322	2.308	2.319	2.325	2.336	2.187	2.283	2.275
	(0.90)	1.106	1.328	1.367	1.375	1.291	1.289	1.225	1.073	1.043	0.954
4	(0.99)	3.283	3.999	4.339	4.349	4.629	4.602	5.002	4.793	4.984	5.028
	(0.95)	1.947	2.374	2.478	2.426	2.238	2.310	2.175	2.063	1.891	1.784
	(0.90)	1.317	1.472	1.362	1.195	1.109	1.008	0.860	0.667	0.544	0.289
5	(0.99)	3.437	4.212	4.454	4.330	4.396	4.739	5.044	4.761	4.731	4.560
	(0.95)	2.018	2.368	2.504	2.337	2.167	2.109	1.862	1.593	1.540	1.249
	(0.90)	1.387	1.414	1.341	1.038	0.865	0.696	0.445	0.083	-0.075	-0.347
6	(0.99)	3.755	4.467	4.754	4.559	4.715	4.836	4.515	4.561	4.303	4.365
	(0.95)	2.164	2.406	2.422	2.267	2.010	1.995	1.654	1.302	1.107	0.744
	(0.90)	1.417	1.428	1.167	0.962	0.634	0.410	0.014	-0.449	-0.666	-1.113
7	(0.99)	3.980	4.599	4.683	4.704	5.000	4.828	4.667	4.489	4.367	4.155
	(0.95)	2.282	2.441	2.410	2.198	1.886	1.535	1.263	0.943	0.356	0.071
	(0.90)	1.536	1.457	1.057	0.817	0.254	-0.015	-0.668	-1.314	-1.610	-2.230
8	(0.99)	4.116	4.775	4.724	4.715	4.762	4.480	4.111	4.278	4.482	3.804
	(0.95)	2.394	2.580	2.244	2.007	1.666	1.266	0.744	0.293	-0.244	-0.658
	(0.90)	1.501	1.374	0.900	0.622	-0.146	-0.593	-1.157	-1.925	-2.678	-3.109
9	(0.99)	4.233	4.671	4.640	4.856	4.580	4.112	3.756	3.536	3.648	3.158
	(0.95)	2.525	2.533	2.064	1.881	1.434	0.892	0.292	-0.344	-0.960	-1.536
	(0.90)	1.564	1.329	0.727	0.181	-0.578	-1.168	-1.751	-2.653	-3.457	-4.122
10	(0.99)	4.232	4.750	4.674	4.489	4.251	3.643	3.467	3.373	3.342	3.036
	(0.95)	2.611	2.481	1.967	1.601	0.936	0.282	-0.360	-1.068	-1.467	-2.404
	(0.90)	1.583	1.176	0.520	-0.246	-1.125	-1.722	-2.449	-3.606	-4.312	-5.252

Table 7
Local Power of the (modified) OOS-F and OOS-t statistics: Recursive Scheme

A. Nominal Size of 1%

		k_2				
Test	π	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	π	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	π	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 π, 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 4 of the text. Recall that under the local alternative, the limiting distributions are not pivotal. Hence the local power results do not necessarily reflect power of the test under an environment that is different from that in (4.1).

Table 8
Local Power of the (modified) OOS-F and OOS-t statistics: Rolling Scheme

A. Nominal Size of 1%

Test	π	k_2 = 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%

Test	π	k_2 = 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%

Test	π	k_2 = 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 π, 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 4 of the text. Recall that under the local alternative, the limiting distributions are not pivotal. Hence the local power results do not necessarily reflect power of the test under an environment that is different from that in (4.1).

Table 9
Local Power of the (modified) OOS-F and OOS-t statistics: Fixed Scheme

A. Nominal Size of 1%

Test	π	k_2 = 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%

Test	π	k_2 = 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%

Test	π	k_2 = 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 π, 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 4 of the text. Recall that under the local alternative, the limiting distributions are not pivotal. Hence the local power results do not necessarily reflect power of the test under an environment that is different from that in (4.1).

Figure 1

Density Plots for OOS-F: Recursive

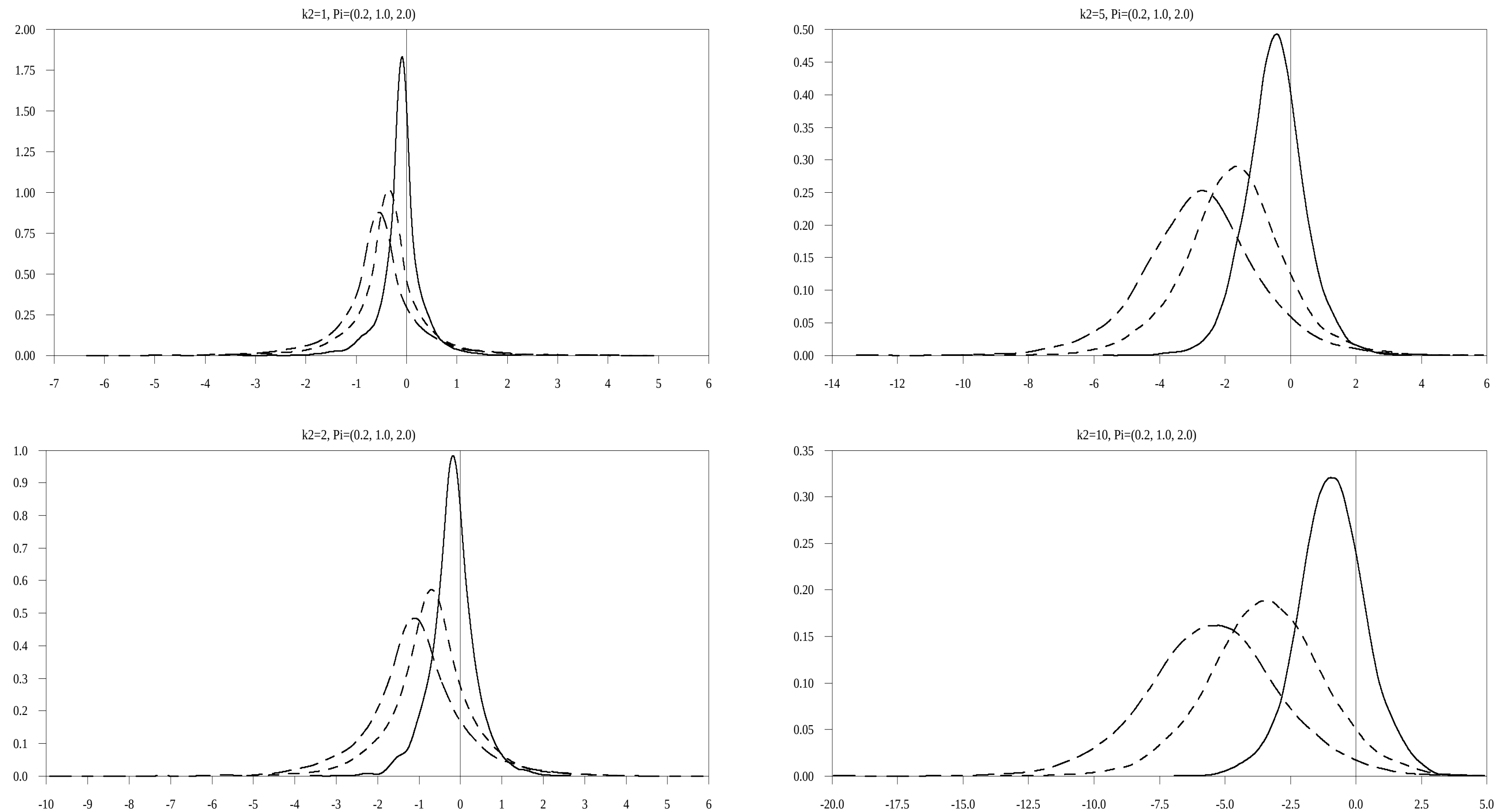

Figure 2

Density Plots for OOS-F: Recursive

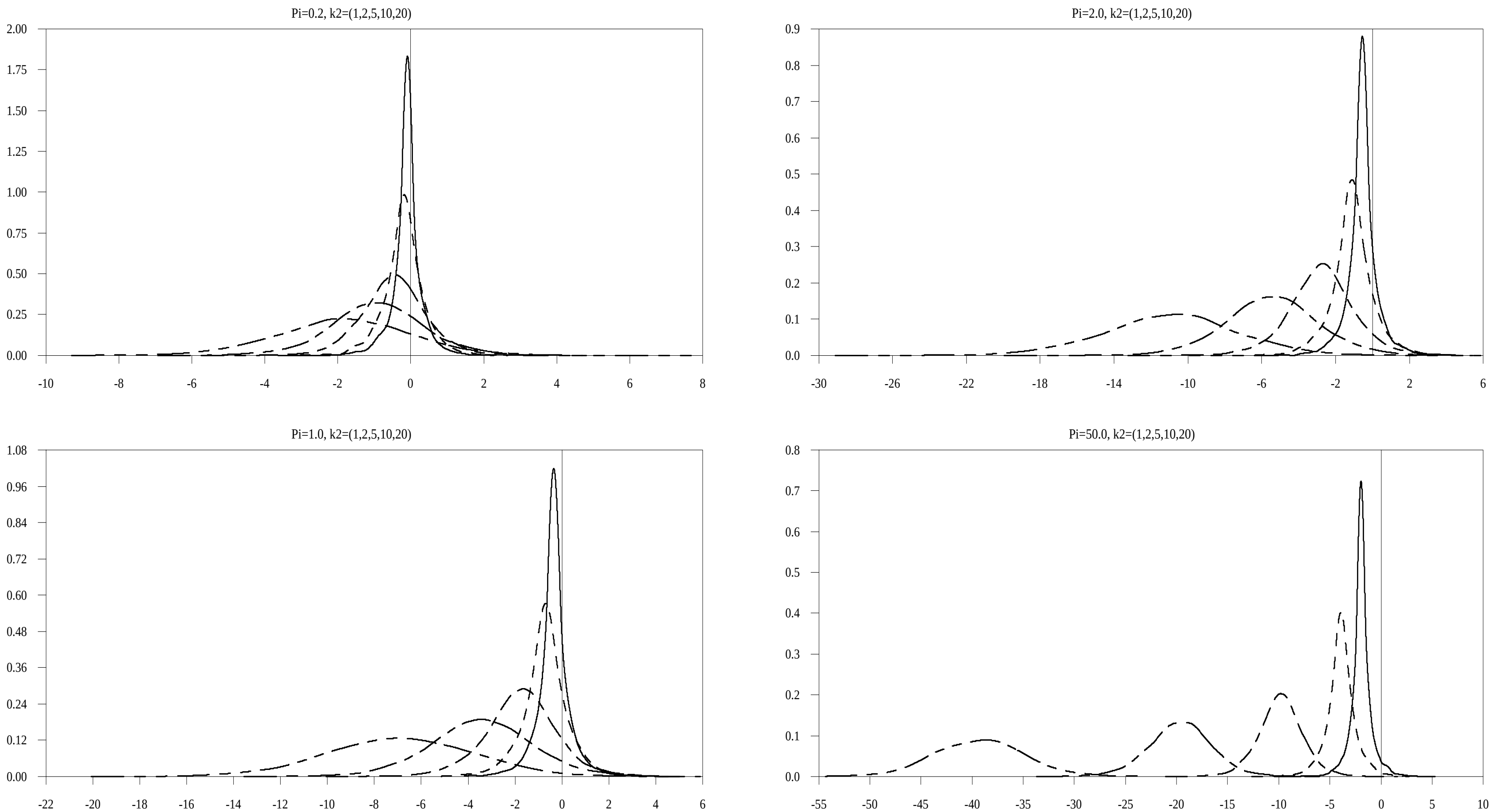

Figure 3

Density Plots for OOS-t: Recursive

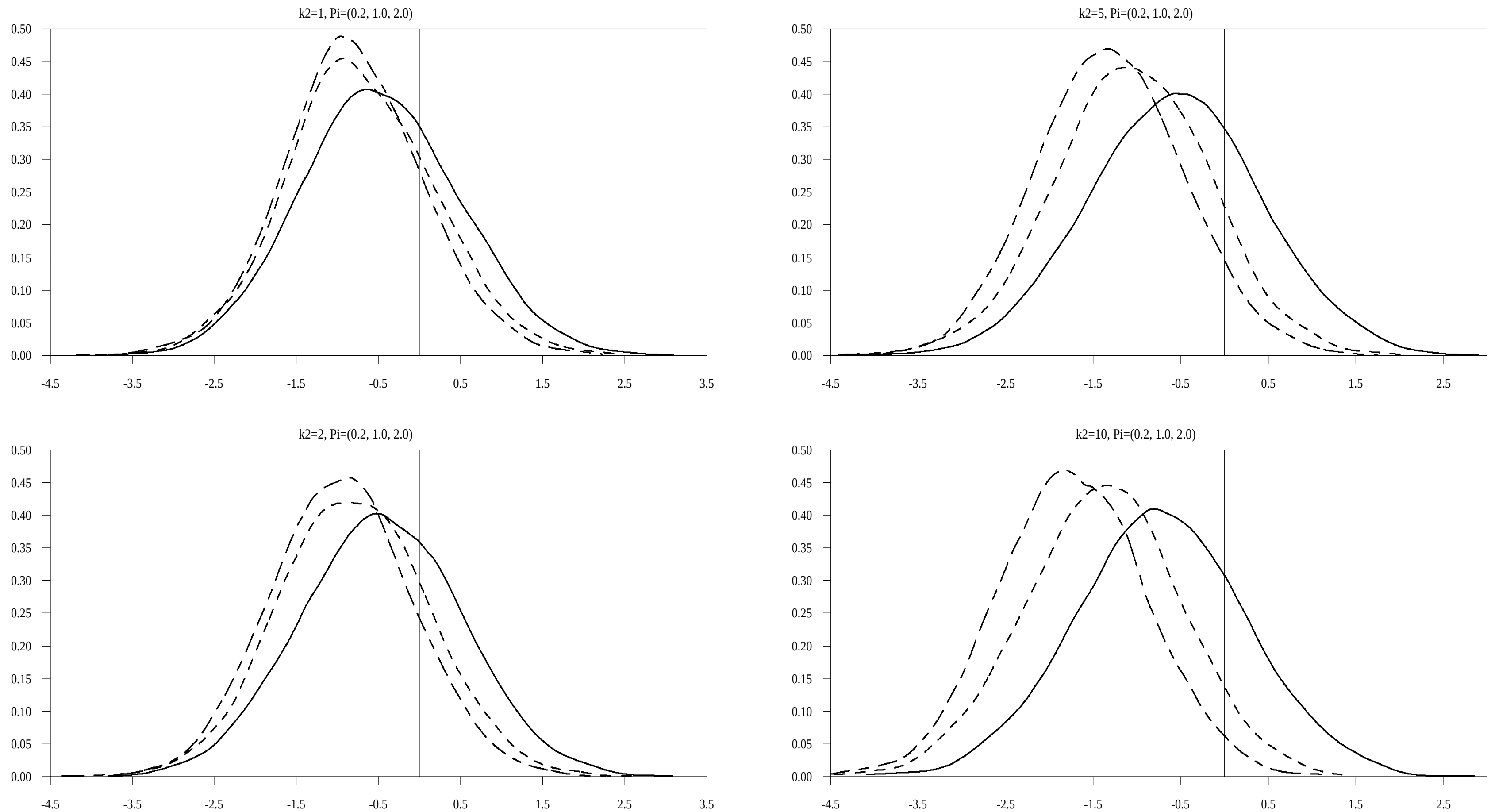

Figure 4

Density Plots for OOS-t: Recursive