INTERNATIONAL ECONOMIC REVIEW
Vol. 35, No. 1, February 1994

GENERALIZED PREDICTIVE TESTS AND STRUCTURAL CHANGE ANALYSIS IN ECONOMETRICS*

BY JEAN-MARIE DUFOUR, ERIC GHYSELS, AND ALASTAIR HALL[1]

A generalized predictive testing procedure for structural stability in nonlinear dynamic simultaneous equations models is presented. It has several attractive features: (1) the tests are based on easy-to-compute predicted residuals; (2) the prediction subsample can be arbitrarily small; (3) only consistency is required and allowance is made for data-based model selection; (4) it is possible to analyze the timing and form of structural change equation by equation or globally, allowing an *exploratory analysis* of structural change conveniently summarized in a *predictive analysis table*; and (5) general forms of temporal dependence between model disturbances are allowed.

1. INTRODUCTION

Many models in econometrics can be written in the form

$$f_t[X_t, \beta] = u_t, \; t = 1, \dots, n \tag{1.1}$$

where f_t is an $m \times 1$ (possibly nonlinear) vector function, X_t is a $q \times 1$ vector of variables, β is an $l \times 1$ parameter vector, and u_t is an $m \times 1$ vector of disturbances such that $E(u_t) = 0$. Prominent examples include linear and nonlinear regression models, time series models, nonlinear simultaneous equations models and Euler equations models.

Before a model is used to draw inference about economic phenomena, it is important to assess the adequacy of its specification. While the examples cited above differ in the choice of $f_t[\cdot]$ and the assumptions made about $[X_t, u_t]$, they are all based on the assumption of parameter constancy over the sample. Therefore, in any of these examples, a natural method for assessing model adequacy is to test parameter constancy. Such a test is particularly appropriate for dynamic econometric models. Lucas (1976) has argued that the parameters of traditional econometric models are functions of the economic environment and certain underlying "taste and technology" (or "structural") parameters. This led to a growing

* Manuscript received June 1991.

[1] The authors thank Badi Baltagi, Bryan Campbell, Jeffrey Mills, Adrian Pagan, Pierre Perron, George Tauchen, Quang Vuong, two anonymous referees, as well as seminar participants at the Canadian Econometric Study Group, the Montreal Econometrics Workshop, the Research Triangle Institute, the University of Aarhus, North Carolina State University, Southampton University, the Board of Governors of the Federal Reserve System, the 1990 Conference of the Society of Economic Dynamics and Control, the 1991 European Meeting of the Econometric Society (Cambridge, U.K.) and the Société canadienne de science économique (1989), for several useful comments. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and by the Government of Québec (Fonds FCAR), and by the Canadian Agency for International Development (PARADI Program).

research program whose aim is to estimate these structural parameters from models based on Euler optimality conditions; see, for example, Hansen and Singleton (1982), Pindyck and Rotemberg (1983), Eichenbaum, Hansen, and Singleton (1987) and Gallant and Tauchen (1989). Given the Lucas critique, it appears fundamental to test whether the parameters of models derived from Euler equations exhibit the constancy predicted by economic theory.[2] Of course, it may be useful to test the stability of a model in other contexts as well.

Despite the importance of this problem, work on testing for the presence of structural change has mainly considered linear regression models (see Chow 1960, Dufour 1980, 1982a, 1982b and the references therein) or linear simultaneous equations models (see Erlat 1983, Lo and Newey 1985, Hodoshima 1986 and Honda 1990). Work on nonlinear dynamic simultaneous equations models has been much more limited. Recently, however, Andrews and Fair (1988), Ghysels and Hall (1990a, 1990b) and Andrews (1993) have proposed tests of parameter stability applicable to such models.

Andrews and Fair (1988, henceforth AF) consider the problem of testing parameter stability against the alternative that the sample can be split into two subsamples with known breakpoint and such that coefficients are stable inside the subperiods considered. They propose Wald, likelihood ratio-type (LR) and Lagrange multiplier-type (LM) tests, and they derive the asymptotic distributions of the test statistics as both subsamples become large. Their results apply under weak regularity conditions which allow for general forms of temporal dependence and heteroskedasticity. However, their analysis does require that the ratio of subsample sizes be asymptotically fixed. Further, to implement the Wald and LM tests, the asymptotic distribution of the parameter estimates needs to be known, while LR tests require additional regularity conditions to be valid. Similarly, Ghysels and Hall (1990b) introduce a different LR-type test which is based on procedures suggested by Gallant (1987a) and Eichenbaum, Hansen, and Singleton (1988). All these tests can be interpreted as generalizations of classic analysis-of-covariance tests for the problem of comparing two regressions (Kullback and Rosenblatt 1957, Chow 1960). Ghysels and Hall (1990a, henceforth GH) propose to test structural stability by using a generalized predictive approach. This procedure can be viewed as an extension of the Davis's (1952) approach, originally proposed for single linear regressions, to nonlinear dynamic simultaneous equations models. It is based on evaluating the orthogonality conditions from one subsample at the parameter estimates obtained from the other subsample. If the model is stable, these estimated orthogonality conditions should hold approximately, yielding residuals that are not statistically different from zero.[3] GH derive the asymptotic distribution of their test statistic under the weak assumptions employed by AF. Compared to the AF tests, the predictive test is simple to implement as well as designed against a wider

[2] Furthermore, parameter constancy tests are particularly convenient in Euler equations models because these models are usually estimated under very weak assumptions about $[X_t, u_t]$, so that very few model specification tests are available; see Hansen (1982) and Ghysels and Hall (1990a, 1990b).

[3] The idea of using such residuals to test parameter constancy in nonlinear models was also suggested and applied by Hoffman and Pagan (1989). Other applications appear in Epstein and Zin (1991) and Ghysels and Hall (1990b).

alternative, but shares the disadvantage of requiring each subsample to be sufficiently large to allow the application of conventional asymptotic theory. Further, the asymptotic distribution of the parameter estimates also needs to be known. More recently, Andrews (1993) has extended the AF framework to cover the case where the breakpoint is unknown. His test is based on calculating a given test (Wald, LM or LR) against a two-regime alternative for a subset of possible breakpoints, and then using the maximum of these statistics as the test statistic. Andrews derives and tabulates the limiting distribution of such tests under weak conditions on $[X_t, u_t]$ and $f_t[\cdot]$. While this procedure may represent a significant improvement over previous contributions, because it allows for unknown breakpoint, some undesirable features remain. Firstly, there must be enough observations both before and after every breakpoint to allow the application of conventional asymptotic theory. Secondly, the Wald and LR-type tests require one to reestimate the model over two subperiods for each breakpoint considered, and thus they can be computationally quite expensive. Thirdly, they are clearly designed against a two-regime alternative and do not allow an exploratory analysis of possible structural shifts during a given subperiod. Indeed, it is a common drawback of all the tests mentioned above that they cannot be applied to detect structural instability at either the very beginning or the end of the sample.

The GH predictive test is computationally convenient. However, it is based on estimated orthogonality conditions that are effectively the average of a set of estimated disturbances. When viewed from this perspective, it is clear that one may lose a considerable amount of information by examining only the average rather than the complete vector of the residuals. Clearly, it is possible for certain patterns in the residuals to indicate structural instability, but still average out not to be significantly different from zero.[4]

In this paper, we propose predictive tests for structural stability that utilize all the information contained in the residuals of a nonlinear dynamic simultaneous equations model of the form (1.1). The tests are applicable when the model is structurally stable during a given asymptotically large subperiod (the "estimation subsample") but the form and timing of possible structural changes during the second period (the "prediction subsample") are unknown. Most importantly, our tests do not require the prediction subsample to be asymptotically large. Therefore, unlike the AF, GH and Andrews tests, our procedures can be applied to test structural stability near the end of the sample. Further, the tests are applicable even if the asymptotic distribution of the parameter estimates is unknown: only a consistency assumption is required. Data-based model selection is also allowed. The tests proposed here can be viewed as extensions of Chow's (1960) predictive test, which was originally proposed for linear regression models, to nonlinear dynamic simultaneous equations models. In addition, we present a simple explor-

[4] It is of interest to note here that the predictive test formally derived by GH is not a generalization of the so-called "predictive Chow test" proposed by Chow (1960) for the case where the predicted subsample is undersized in a linear model. Chow's predictive test assesses whether *all* the prediction errors have mean zero, not whether these prediction errors are zero on average. In general, there is no reason to expect that the sum of predicted residuals will be a sufficient statistic to analyze structural change, either in finite samples or asymptotically (except against very specific alternatives).

atory technique for analyzing structural change in the prediction subsample which extends earlier work by Dufour (1980, 1981) for the linear regression model. This exploratory technique can be a useful complement to any global test for the presence of structural change, such as the tests suggested by Andrews and Fair (1988), Ghysels and Hall (1990a, 1990b) and Andrews (1993).

The intuition behind our procedures follows from the observation that if the model is stable over the entire sample, then the laws of motion of u_t within the prediction subsample are the same as those for the estimation subsample. We propose using the estimation subsample to estimate relevant aspects of the laws of motion of u_t in the prediction sample, and then employ these to estimate probabilities of observing values of u_t in the prediction subsample (or bounds on the latter probabilities), when the model is stable. To develop our tests, we must circumvent two problems. The first problem comes from the fact that β is unknown, so that u_t is unobservable. Therefore, we base our tests on the residuals $\hat{u}_t = f_t[X_t, \hat{\beta}]$ where $\hat{\beta}$ is obtained from the estimation subsample. We call such residuals *predicted residuals*. It is important to note that predicted residuals are *not generally identical to prediction errors*, except in special cases (e.g., linear regression models). Indeed, in nonlinear simultaneous equations models, forecasts and forecast errors can be quite difficult to compute (see Bianchi and Calzolari 1980, Brown and Mariano 1984 and Mariano and Brown 1983a, 1983b, 1985), while predicted residuals are usually easy to calculate. We prove two general propositions under which $\hat{u}_t$ and u_t are asymptotically equivalent, and so asymptotically have the same distribution. These propositions allow for data based model selection and may be of separate interest. The second problem pertains to the specification of the laws of motion of u_t. In some circumstances, one may be comfortable with completely specifying the probability law of $\{u_t\}$ up to a finite number of parameters (e.g., with a normality assumption). However, in other cases, such as Euler equations models, economic theory does not provide distributional information and the current practice is to employ robust estimation techniques such as GMM (Hansen 1982). Accordingly, we propose three different methods for estimating the distribution of the predicted residuals, each of which uses different amounts of information about the probability distribution of $\{u_t\}$. First, we consider the case where $\{u_t\}$ has a normal distribution and show that our methods yield convenient normal and χ^2 statistics. Secondly, we consider the case where one has very little information about the distribution of $\{u_t\}$ and propose using Markov inequalities to provide bounds on the desired probabilities. This method only requires an assumption about the constancy of certain conditional or unconditional moments. Finally, we consider the use of semi-nonparametric (SNP) techniques to model the distribution of u_t and propose using this estimated distribution to calculate the desired probabilities. We illustrate our techniques using recent models for the comovements of asset prices and consumption considered by Hansen and Singleton (1982) and Gallant and Tauchen (1989). We find clear evidence of structural instability mostly but not exclusively related to the October 1979 monetary policy shift.

Section 2 describes our general framework, defines the test statistics proposed and shows how they can be used when the distribution of the disturbances is assumed to be known up to a finite number of parameters (e.g., a normal distribution). Section 3 gives the asymptotic equivalence results which ensure the asymptotic validity of the method, under the simple assumption that consistent model selection and parameter estimates are available. Sections 4 and 5 discuss two alternative ways of dealing with nonnormal disturbances: the first one is based on Markov inequalities, while the second one is based on semi-nonparametric estimation. Section 6 presents applications of the techniques suggested to a VAR model and to consumption-based asset pricing models. Finally, Section 7 contains a number of concluding remarks.

2. PREDICTIVE TEST STATISTICS

Consider a general model of the form

$$f_t[k, X_t, \beta(k)] = u_t, \quad t \in T, \tag{2.1}$$

where $f_t[k, X_t, \beta(k)] = (f_{1t}[k, X_t, \beta(k)], \dots, f_{mt}[k, X_t, \beta(k)])'$, f_{jt} is a known real function, X_t is a $q \times 1$ vector of random variables, k is a positive integer parameter, $j = 1, \dots, m$, $\beta(k)$ is an $l(k) \times 1$ real parameter vector whose dimension $l(k)$ can be a function of k, T is a subset of the integers ($T \subseteq \mathbb{I}$) and $u_t = (u_{1t}, \dots, u_{mt})$ is a vector of random disturbances such that $E(u_t) = 0$. In certain cases, we shall also make the stronger assumption $E(u_t|X_{t-1}, X_{t-2}, \dots) = 0$. The integer parameter k can be viewed as an index for the model, to be chosen from the nonempty countable set $M \subseteq \mathbb{N}$, possibly on the basis of the data. Clearly, it is natural to suppose that the dimension $l(k)$ of the parameter vector $\beta(k)$ be a function of k (e.g., the order of an ARMA model). When only one model is considered, equation (2.1) can be written more compactly as

$$f_t[X_t, \beta] = u_t, \quad t \in T \tag{2.2}$$

where β has dimension l.

In general, it will not be assumed that the random disturbances $\{u_t : t \in T\}$ are mutually independent (or uncorrelated). In many cases though, it is appropriate to make such an assumption. Such cases would include simultaneous equations models with uncorrelated errors, as well as Euler equations models where innovations are martingale difference sequences. When the u_t's are autocorrelated, we will assume that we can estimate the autocorrelation structure and orthogonalize the disturbances accordingly (e.g., by using an autoregressive model on the disturbances). Assuming that the disturbances are uncorrelated may therefore not be very restrictive from a theoretical point of view, since it is possible to include the parameters of the autocorrelation structure into the vector $\beta(k)$. On the other hand, it may be simpler in practice not to model the autocorrelation structure. The results below allow both possibilities.

Suppose now that $T = \{-n_1 + 1, \dots, 0, 1, \dots, n_2\}$ and the sample is split into two parts: the estimation subsample $T_1 = \{-n_1 + 1, \dots, 0\}$ and the prediction subsample $T_2 = \{1, \dots, n_2\}$. The first sample is assumed to be large enough to allow the estimation of the model. In contrast, the prediction subsample may not be large (possibly containing only one observation). Further, we suppose that the model is stable over the first period T_1, while it is not necessarily stable over the second period T_2. We are interested in detecting the presence of structural change during this second period. We would also like to analyze the timing and form of possible shifts over the latter period. Note we are not assuming that there is a structural break at the point where the sample is split: we simply assume that some form of structural change may have taken place during the second period. More explicitly, the null hypothesis will be defined as

$$H_0 : E(f_t[k_0, X_t, \beta_0]) = 0, \quad \forall\, t \in T, \tag{2.3}$$

where $k_0 \in M$ and $\beta_0 \in \mathbb{R}^{l(k_0)}$, while the alternatives considered are subsets of the general alternative

$$H_1 : E(f_t[k_0, X_t, \beta_0]) = 0, \ \forall\, t \in T_1, \quad \text{and } E(f_t[k_0, X_t, \beta_0]) \neq 0, \text{ for some } t \in T_2. \tag{2.4}$$

As pointed out by Ghysels and Hall (1990a), a natural way of testing structural stability consists in estimating the model from the first sample and then checking whether the estimated disturbances from the second sample are "large." We shall call the latter *predicted residuals*. More precisely, if $\hat{k}_{n_1}$ and $\tilde{\beta}(\hat{k}_{n_1})$ are estimators of k and $\beta(k)$ obtained from sample T_1, we check whether the predicted residuals

$$\tilde{u}_t(n_1) \equiv f_t[\hat{k}_{n_1}, X_t, \tilde{\beta}(\hat{k}_{n_1})], \quad t \in T_2, \tag{2.5}$$

are statistically "large." We call such tests *generalized predictive tests* for structural change. Moreover, by looking at individual equations and individual values of $t \in T_2$, we can assess which equations and which observations in the second sample exhibit discrepancies with the rest of the sample. Note that predicted residuals are not generally identical to prediction errors, except in special cases (e.g., linear regression models). Indeed, in nonlinear models such as (2.1), forecasts can be difficult to compute; see Bianchi and Calzolari (1980), Brown and Mariano (1984) and Mariano and Brown (1983a, 1983b, 1985). By contrast, it is usually straightforward to compute predicted residuals.

We propose two types of predictive stability tests. The first one examines individual values of $\tilde{u}_t(n_1) \equiv f_t[\hat{k}_{n_1}, X_t, \tilde{\beta}(\hat{k}_{n_1})]$ and their components, where $t \in T_2$. Such tests are called *individual* or *sequential predictive tests*. The second one considers several or all the values of $\tilde{u}_t(n_1)$, $t \in T_2$, for evidence of instability: these tests are referred to as *joint predictive tests*. Both tests can be applied under fairly general circumstances, provided we can rely on: (1) an asymptotic equivalence result, and (2) a distributional stability assumption under the null hypothesis appearing in (2.3).

To obtain the asymptotic equivalence result needed, it will be shown in Section 3 that, under very general conditions, $\tilde{u}_t(n_1)$ and $u_t = f_t(k_0, X_t, \beta_0)$ have the same asymptotic distribution as $n_1 \to \infty$, provided $\text{plim}_{n_1\to\infty} \hat{k}_{n_1} = k_0$ and $\text{plim}_{n_1\to\infty} \tilde{\beta}(k_0) = \beta_0$. In addition, we need to use a distributional stability argument in order to decide whether $\tilde{u}_t(n_1)$ is "large." This requires being able to determine the distribution of u_t (under H_0), where $t \in T_2$, on the basis of the first sample T_1 (e.g., by appropriate "stationarity" assumptions). The distribution considered may be the unconditional distribution of u_t or the conditional distribution given an appropriate information set Ψ. Natural choices for the set Ψ are the history of the process up to the end of the first sample (Ψ_0) or up to time $t - 1$ (Ψ_{t-1}), where we denote $\Psi_\tau = \{X_s : s \leq \tau\}$. Depending on the conditioning information set, we distinguish: (1) predictive tests based on the unconditional distribution, i.e., Ψ is the empty set; (2) predictive tests conditioned on data from the period T_1, i.e. $\Psi = \Psi_0$; and (3) tests for $\tilde{u}_t(n_1)$, $t \in T_2$ using information up to $t - 1$, hence, possibly including observations from the period T_2. In the second case, we assume also that $E(u_t|\Psi_0) = 0$ for $t \in T_2$, while in the third case, we assume $E(u_t|\Psi_{t-1}) = 0$ for $t \in T_2$. Tests based on the unconditional distribution are often the easiest to implement, in comparison with conditional tests. The latter, however, are most likely to be more powerful than unconditional tests because they use more information. This, in fact, emphasizes an important point already noted in Section 1. Namely, since we look at individual predicted residuals $\tilde{u}_t(n_1)$, instead of averages, distributional assumptions about u_t will play a crucial role, even if n_1 is large.

In order to analyze the information contained in the predicted residual vectors $\tilde{u}_t(n_1)$, it will usually be convenient to scale these by appropriate factors or matrices, such as standard errors or covariance matrix estimates. Let $\tilde{\sigma}_{ijst}(n_1|\Psi)$, where $i, j = 1, \ldots, m$ and $s, t = 1, \ldots, n_2$, be a collection of such factors. The latter will typically be functions of the first sample (of size n_1). Ψ represents a (possibly empty) set of conditioning variables. Let also

$$\tilde{\Sigma}_{st}(n_1|\Psi) = [\tilde{\sigma}_{ijst}(n_1|\Psi)]_{i,j=1,\ldots,m}, \quad \tilde{\Delta}_{ij}(n_1|\Psi) = [\tilde{\sigma}_{ijst}(n_1|\Psi)]_{s,t=1,\ldots,n_2}, \tag{2.6}$$

$$\tilde{\Delta}(n_1|\Psi) = [\tilde{\Sigma}_{st}(n_1|\Psi)]_{s,t=1,\ldots,n_2}. \tag{2.7}$$

For example, if it is assumed that $E(u_{jt}^2) < \infty$, for all j and t, a natural choice for $\tilde{\sigma}_{ijst}(n_1|\Psi)$ is to take a weakly consistent estimator (as $n_1 \to \infty$) of the covariance $\sigma_{ijst}(\Psi)$ between u_{is} and u_{jt} (conditional on Ψ), i.e., $\text{plim}_{n_1 \to \infty} [\tilde{\sigma}_{ijst}(\Psi) - \sigma_{ijst}(\Psi)] = 0$. Note that $\Psi = \varnothing$ (empty conditioning set) yields unconditional covariances. Assumptions under which such estimators can be obtained will be discussed in Section 3. In this case, $\tilde{\Sigma}_{st}(n_1|\Psi)$ is a consistent estimator of the covariance matrix $\Sigma_{st}(\Psi)$ between u_s and u_t, $\tilde{\Delta}_{ij}(n_1|\Psi)$ is a consistent estimator of the covariance matrix $\Delta_{ij}(\Psi)$ between $u^i = (u_{i1}, \ldots, u_{in_2})'$ and $u^j = (u_{j1}, \ldots, u_{jn_2})'$; and $\tilde{\Delta}(n_1|\Psi)$ is a consistent estimator of the covariance matrix $\Delta(\Psi)$ of $u = (u_1', u_2', \ldots, u_{n_2}')'$. In most cases studied below, we shall adopt this covariance interpretation. However, other interpretations may be required (e.g., when second moments do not exist). We can now define the basic test statistics that will be considered.

Consider first the individual predictive test statistics. These are defined as follows:

$$\tilde{v}_{jt}(n_1|\Psi) = \tilde{u}_{jt}(n_1|\Psi)/\tilde{\sigma}_{jt}(n_1|\Psi), \quad j = 1, \ldots, m, \ t \in T_2, \tag{2.8}$$

(2.9) $$\tilde{w}_t(n_1|\Psi) = \tilde{u}_t(n_1)'[\tilde{\Sigma}_t(n_1|\Psi)]^{-1}\tilde{u}_t(n_1), \quad t \in T_2,$$

where $\tilde{\sigma}_{jt}(n_1|\Psi)^2 = \tilde{\sigma}_{jjtt}(n_1|\Psi)$ and $\tilde{\Sigma}_t(n_1|\Psi) = \tilde{\Sigma}_{tt}(n_1|\Psi)$. If $\text{plim}_{n_1\to\infty} [\tilde{\sigma}_{ijst}(\Psi) - \sigma_{ijst}(\Psi)] = 0$ and appropriate regularity conditions hold (see Section 3), the asymptotic distributions (as $n_1 \to \infty$) of $\tilde{v}_{jt}(n_1|\Psi)$ and $\tilde{w}_t(n_1|\Psi)$ are identical to the distributions of $v_{jt}(\Psi) = u_{jt}/\sigma_{jt}(\Psi)$ and $w_t(\Psi) = u_t'[\Sigma_t(\Psi)]^{-1}u_t$ respectively, where $\sigma_{jt}^2(\Psi) = \sigma_{jjtt}(\Psi)$ and $\Sigma_t(\Psi) = \Sigma_{tt}(\Psi)$. The statistics $\tilde{v}_{jt}$, $j = 1, \ldots, m$, allow one to assess discrepancies at observation t for each equation, while $\tilde{w}_t$ provides an overall indication based on the m equations taken jointly. Note also that p-values based on $\tilde{v}_{jt}$ may be combined by using Bonferroni's inequality.

The joint predictive test statistics for individual equations are

(2.10) $$\tilde{U}_j(n_1|\Psi) = \tilde{u}^j(n_1)'[\tilde{\Delta}_j(n_1|\Psi)]^{-1}\tilde{u}^j(n_1), \quad j = 1, \ldots m,$$

where $\tilde{u}^j(n_1) = [\tilde{u}_{j1}(n_1), \tilde{u}_{j2}(n_1), \ldots, \tilde{u}_{jn_2}(n_1)]'$ is the vector of predicted residuals for the jth equation, and $\tilde{\Delta}_j(n_1|\Psi) = \tilde{\Delta}_{jj}(n_1|\Psi)$ is a consistent estimator (as $n_1 \to \infty$) of the covariance matrix of u^j. $\tilde{U}_j(n_1|\Psi)$ is a portmanteau statistic that allows one to assess whether the n_2 predicted residuals for the jth equation show evidence of structural change. A portmanteau statistic that pools the m equations considered takes the form

(2.11) $$\tilde{W}(n_1|\Psi) = \tilde{u}(n_1)'[\tilde{\Delta}(n_1|\Psi)]^{-1}\tilde{u}(n_1),$$

where $\tilde{u}(n_1) = [\tilde{u}_1(n_1)', \tilde{u}_2(n_1)', \ldots, \tilde{u}_{n_2}(n_1)']'$ is the $(mn_2) \times 1$ vector stacking all the predicted residuals and $\tilde{\Delta}(n_1|\Psi)$ is given by (2.7). $\tilde{U}_j(n_1|\Psi)$ is a natural generalization of Chow's (1960) "predictive test" (for a single linear regression) to the case of a nonlinear dynamic equation, while $\tilde{W}(n_1|\Psi)$ provides a similar extension to systems of nonlinear simultaneous dynamic equations.

In the special (but important) case where $u_1, \ldots, u_{n_2}$ are uncorrelated, the statistics $\tilde{U}_j(n_1|\Psi)$ and $\tilde{W}(n_1|\Psi)$ simplify considerably:

(2.12) $$\tilde{U}_j(n_1|\Psi) = \sum_{t=1}^{n_2} \tilde{v}_{jt}(n_1|\Psi)^2 = \sum_{t=1}^{n_2} [\tilde{u}_{jt}(n_1)/\tilde{\sigma}_{jt}(n_1|\Psi)]^2, \quad j = 1, \ldots, m,$$

(2.13) $$\tilde{W}(n_1|\Psi) = \sum_{t=1}^{n_2} \tilde{w}_t(n_1|\Psi) = \sum_{t=1}^{n_2} \tilde{u}_t(n_1)'[\tilde{\Sigma}_t(n_1|\Psi)]^{-1}\tilde{u}_t(n_1).$$

$\tilde{U}_j(n_1|\Psi)$ is the sum of the squared $\tilde{v}_{jt}(n_1|\Psi)$ statistics while $\tilde{W}(n_1|\Psi)$ is the sum of the $\tilde{w}_t(n_1|\Psi)$ statistics.

Since there are m equations and n_2 predicted residuals for each equation, a convenient way of presenting the statistics defined in (2.8) through (2.11) consists in building a table where the lines (or columns) correspond to equations ($j = 1, \ldots, m$) and the columns (or lines) to observations ($t = 1, \ldots, n_2$), and where $\tilde{v}_{jt}$ is presented at the corresponding position. Furthermore, $\tilde{w}_t$ is given at the end of tth column (line), $\tilde{U}_j$ at the end of jth line (column), and $\tilde{W}$ in the corner of the last line and column. Because it is akin to an analysis-of-variance table, such a table will be called a *predictive analysis table* (PAT). The general form of such a table is

TABLE 2.1
PREDICTIVE ANALYSIS TABLE

	$t = 1$	2	$\cdots$	n_2	
$j = 1$	$\bar{v}_{11}$	$\bar{v}_{12}$	$\cdots$	$\bar{v}_{1n_2}$	$\bar{U}_1$
2	$\bar{v}_{21}$	$\bar{v}_{22}$	$\cdots$	$\bar{v}_{2n_2}$	$\bar{U}_2$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
m	$\bar{v}_{m1}$	$\bar{v}_{m2}$	$\cdots$	$\bar{v}_{mn_2}$	$\bar{U}_m$
	$\bar{w}_1$	$\bar{w}_2$	$\cdots$	$\bar{w}_{n_2}$	$\bar{W}$

described by Table 2.1. Depending on the way the various statistics are conditioned, several variants of the predictive analysis table are possible. The most natural choices consist in taking either $\Psi = \varnothing$ or $\Psi = \Psi_0$ for all four statistics appearing in the table. In the first case, we are led to consider unconditional tests while, in the second one, the tests will be conditional on the history of the process up to the end of the first sample. Another possibility consists in taking $\Psi = \Psi_{t-1}$ for $\bar{v}_{it}$ and $\bar{w}_t$, and $\Psi = \Psi_0$ for $\bar{U}_j$ and $\bar{W}$, i.e., $\bar{v}_{jt} = \bar{v}_j(n_1|\Psi_{t-1})$, $\bar{w}_t = \bar{w}_t(n_1|\Psi_{t-1})$, $\bar{U}_j = \bar{U}_j(n_1|\Psi_0)$ and $\bar{W} = \bar{W}(n_1|\Psi_0)$. In some situations, this choice is considerably more convenient for obtaining p-values than the two previous possibilities (see Sections 5 and 6).

Because $\bar{v}_{jt}$ is asymptotically equivalent to $v_{jt} = u_{jt}/\sigma_{jt}$, it is clear that distributional assumptions about u_{jt} can play an important role when deciding whether $\bar{v}_{jt}$ is too large; and similarly for the statistics $\bar{w}_t$, $\bar{U}_j$ and $\bar{W}$. If it is assumed that $u = (u_1', u_2', \ldots, u_{n_2}')'$ follows a multinormal distribution (either unconditionally, or conditional on Ψ, depending on the type of test considered), it is natural to define $\sigma_{ijst} = E(u_{is}u_{jt})$ or $\sigma_{ijst} = E(u_{is}u_{jt}|\Psi)$. In this case, the results of the next section (see Proposition 3.1) demonstrate that the asymptotic distributions as $n_1 \to \infty$ (under H_0) of $\bar{v}_{jt}(n_1)$, $\bar{w}_t(n_1)$, $\bar{U}_j(n_1)$ and $\bar{W}(n_1)$ are respectively $N(0, 1)$, $\chi^2(m)$, $\chi^2(n_2)$ and $\chi^2(mn_2)$, provided $\operatorname{plim}_{n_1\to\infty}(\bar{\sigma}_{ijst} - \sigma_{ijst}) = 0$. When the process u_t is stationary and ergodic, a natural estimator of the unconditional covariance σ_{ijst} is

$$\bar{\sigma}_{ijst} = \frac{1}{n_1} \sum_{\tau = -n_1+1}^{-|t-s|} \bar{u}_{i\tau}(n_1)\bar{u}_{j,\tau+|t-s|}(n_1), \; s \le t. \tag{2.14}$$

The consistency of $\bar{\sigma}_{ijst}$ for σ_{ijst} as $n_1 \to \infty$ follows from Proposition 3.2. It is important to note, however, that the chi-squared approximations for $\bar{U}_j(n_1)$ and $\bar{W}(n_1)$ may not be accurate even asymptotically if n_2 is not small with respect to n_1 (e.g., if we let the number of predicted residuals grow as $n_1 \to \infty$).

More generally, if it is assumed that the distributions of v_{jt}, w_t, etc., are not normal but known up to a finite-dimensional parameter vector θ (moments, for example), then asymptotically valid p-values can be computed, provided a consistent estimator $\bar{\theta}(n_1)$ of θ is available and the distribution function is continuous in its arguments. For example, if the (possibly conditional) distribution function of $|v_{jt}|$ is $G_{jt}(x;\, \theta_0)$, $x \in \mathbb{R}$, where $G_{jt}(x;\, \theta)$ is continuous in (x, θ), then

$$\operatorname*{plim}_{n_1 \to \infty} \{G_{jt}[|\bar{v}_{jt}(n_1)|,\ \bar{\theta}(n_1)] - G_{jt}[|v_{jt}|,\ \theta_0]\} = 0$$

provided $\text{plim}_{n_1 \to \infty}\ \bar{\theta}(n_1) = \theta_0$; the validity of this argument is ensured by Proposition 3.1 below.[5]

Note finally that the assumed distribution function may be validated by the data, using for example the conditional moment tests of Newey (1985) and Tauchen (1985). If these tests reject or no information is available about the functional form of $G_{jt}(\cdot)$, then it is more appropriate to use the procedures outlined in Sections 4 and 5.

3. ASYMPTOTIC EQUIVALENCE

In this section, we give two propositions on asymptotic equivalence. The first proposition will allow us to conclude that $f_t[\hat{k}_n, X_t, \bar{\beta}_n(\hat{k}_n)]$ and $f_t[k_0, X_t, \beta_0]$ have the same asymptotic distribution, provided $\text{plim}_{n\to\infty}\ \hat{k}_n = k_0$ and $\text{plim}_{n\to\infty}\ \bar{\beta}_n(k_0) = \beta_0$ (where $n = n_1$). The second proposition will do the same for statistics of the form $(1/n)\sum_{t=1}^{n} g(f_t[\hat{k}_n, X_t, \bar{\beta}_n(\hat{k}_n)])$ and $(1/n)\sum_{t=1}^{n} g(f_t[k_0, X_t, \beta_0])$, where $g(\cdot)$ is a function of f_t. Important examples of such functions $g(\cdot)$ include moment estimators such as $g(\cdot) = |\cdot|^r$, where $r > 0$. The regularity conditions used are very weak, only involving local continuity or Lipschitz conditions. In addition, no information will be needed on the asymptotic distribution of the consistent estimators employed. The first proposition is based on the following assumptions, where $\|\cdot\|_k$ refers to the usual Euclidean norm in $\mathbb{R}^k$ and $\mathbb{N}$ is the set of positive integers.

ASSUMPTION B1 (CONSISTENT ESTIMATORS). *Let M be a nonempty subset of the positive integers ($\varnothing \neq M \subseteq \mathbb{N}$), $k \in M$, $l(k) \in \mathbb{N}$, $q \in \mathbb{N}$ and $n \in \mathbb{N}$. $\{X \in \mathbb{R}^q$, $X_t \in \mathbb{R}^q$, $\bar{\beta}_n(k) \in \mathbb{R}^{l(k)} : t \in \mathbb{I}$, $n \in \mathbb{N}$, $k \in M\}$ is a collection of random vectors, and $\hat{k}_n$ is a positive integer-valued random variable for each $n \in \mathbb{N}$, all defined on a common probability space $(\Omega, \mathcal{A}, P)$ and such that $\lim_{n\to\infty} P[\hat{k}_n = k_0] = 1$ and $\text{plim}_{n\to\infty}\ \bar{\beta}_n(k_0) = \beta_0$, where $k_0 \in M$ and $\beta_0 \in \mathbb{R}^{l(k_0)}$.*

ASSUMPTION B2 (MEASURABILITY). *Let $x \in \mathbb{R}^q$ and $\beta(k) \in \mathbb{R}^{l(k)}$. $\{g_t[k, x, \beta(k)] : t \in \mathbb{I}\}$ is a collection of functions from $\mathbb{N} \times \mathbb{R}^{q+l(k)}$ to $\mathbb{R}^m$ such that $g_t[k, X, \beta(k)]$, $g_t[\hat{k}_n, X, \bar{\beta}_n(\hat{k}_n)]$, $g_t[k, X_t, \beta(k)]$ and $g_t[\hat{k}_n, X_t, \bar{\beta}_n(\hat{k}_n)]$ are random vectors for all $k \in M$, $t \in \mathbb{I}$ and $n \in \mathbb{N}$.*

ASSUMPTION B3 (JOINT CONTINUITY). *For all $t \in \mathbb{I}$, $g_t[k_0, x, \beta]$ is a continuous function of (x, β) for $\|\beta - \beta_0\|_{l(k_0)} \leq \tau(\beta_0)$ and some $\tau(\beta_0) > 0$.*

ASSUMPTION B4 (LIPSCHITZ CONDITION). *There is a constant $\tau(\beta_0) > 0$ such that $\beta \in \mathbb{R}^{l(k_0)}$ and $\|\beta - \beta_0\|_{l(k_0)} \leq \tau(\beta_0)$ imply $\|g_t[k_0, X, \beta] - g_t[k_0, X, \beta_0]\|_m \leq$*

[5] Note that θ is usually distinct from the vector of structural parameters β. Namely, β contains the parameters that index the moment condition, while θ is a vector of nuisance parameters that determine the distribution G_{jt} required to perform the tests.

$B_t(k_0, X, \beta_0)h[d(\beta, \beta_0), \beta_0]$, $\forall t$, a.s., where B_t, h and d are nonnegative real-valued functions such that $B_t(k_0, X, \beta_0)$ is a random variable, $d(\beta, \beta_0) \to 0$ as $\|\beta - \beta_0\|_{l(k_0)} \to 0$, and $h(y, \beta_0) \to 0$ as $y \downarrow 0$.

Assumption B1 states that the model (or discrete parameter) estimator $\hat{k}_n$ and the parameter estimator $\tilde{\beta}_n(k_0)$ obtained under the assumption that $k = k_0$ are consistent as $n \to \infty$. B2 simply assumes the measurability of the relevant functions of the data. B3 assumes that g_t is continuous with respect to x and β in a neighborhood of β_0. B4 assumes that g_t is Lipschitz continuous with respect to β in a neighborhood of β_0, and thus relaxes the joint continuity on x and β. B3 and B4 will be taken as alternative assumptions. In applications, we will take g_t either as identical to f_t or equal to appropriate transformations of it, e.g., $g_t = f_{jt}^2$, $g_t = f_{jt}^4$, etc. Using the Assumptions B1 to B4, we can show the following proposition. The proofs of the propositions are given in the Appendix.

PROPOSITION 3.1 (POINTWISE EQUIVALENCE). *Let Assumptions B1 and B2 hold. If B3 or B4 holds, then*

$$\operatorname*{plim}_{n\to\infty} \{g_t[\hat{k}_n, X, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X, \beta_0]\} = 0, \ \forall t. \tag{3.1}$$

By taking $g_t = f_{jt}$, where f_t is given in Section 2, we see that $\tilde{u}_t \equiv f_t[\hat{k}_n, X_t, \tilde{\beta}_n(\hat{k}_n)]$ and $u_t \equiv f_t[k_0, X_t, \beta_0]$ must have the same asymptotic distribution as $n \to \infty$, provided $\operatorname{plim}_{n\to\infty} \hat{k}_n = k_0$ and $\operatorname{plim}_{n\to\infty} \tilde{\beta}_n(k_0) = \beta_0$. Further, $\operatorname{plim}_{n\to\infty} [S(\tilde{u}_1, \ldots, \tilde{u}_{n_2}) - S(u_1, \ldots, u_{n_2})] = 0$ for any continuous function $S: \mathbb{R}^{mn_2} \to \mathbb{R}^s$, so that $S(\tilde{u}_1, \ldots, \tilde{u}_{n_2})$ and $S(u_1, \ldots, u_{n_2})$ have the same asymptotic distribution as $n \to \infty$ (for n_2 and s fixed).

An interesting special case of Proposition 3.1 is the one where only one model is considered and is identical to the "true" model ($\hat{k}_n = k_0$). Since the result takes a simple form in this case and because of its importance, we describe it extensively (without explicit reference to Assumptions B1 to B4) in the following corollary.

COROLLARY 3.1. *Let $X \in \mathbb{R}^q$ and $\tilde{\beta}_n \in \mathbb{R}^l$, $n = 1, 2, \ldots$, be random vectors defined on a common probability space $(\Omega, \mathcal{A}, P)$ such that* $\operatorname{plim}_{n\to\infty} \tilde{\beta}_n = \beta_0 \in \mathbb{R}^l$. *Let $g(x, \beta)$ be a function from $\mathbb{R}^{q+l}$ to $\mathbb{R}^m$ ($x \in \mathbb{R}^q$, $\beta \in \mathbb{R}^l$), such that $g(X, \tilde{\beta}_n)$ is a random vector and such that the following condition holds: there is a constant $\tau(\beta_0) > 0$ such that $\|\beta - \beta_0\|_l < \tau(\beta_0)$ implies*

(i) *$g(x, \beta)$ is a continuous function of (x, β), or*

(ii) *$\|g(X, \beta) - g(X, \beta_0)\|_m \le B(X, \beta_0)h[d(\beta, \beta_0), \beta_0]$ a.s.,*

where B, h and d are nonnegative real-valued functions such that $B(X, \beta_0)$ is a random variable, $d(\beta, \beta_0) \to 0$ as $\|\beta - \beta_0\|_l \to 0$, $h(y, \beta_0) \to 0$ as $y \downarrow 0$. Then

$$\operatorname*{plim}_{n\to\infty} [g(X, \tilde{\beta}_n) - g(X, \beta_0)] = 0. \tag{3.2}$$

In order to apply the tests described in Section 2, we shall also need consistent estimators of the variances and covariances of u_t and, eventually, of higher-order moments. Stronger assumptions are needed in order to do this. For this reason, we introduce two additional assumptions.

ASSUMPTION B5 (UNIFORM CONVERGENCE). *There is a constant* $\tau(\beta_0) > 0$ *and a function* $c_n(\beta_0)$ *of* n *and* β_0 *such that* $\beta \in \mathbb{R}^{l_0}$ *and* $\|\beta - \beta_0\|_{l_0} \leq \tau(\beta_0)$ *imply*

(a) $\text{plim}_{n\to\infty} \{\sup_{\beta\in B_0} \|c_n(\beta_0) \sum_{t=1}^n (g_t[k_0, X_t, \beta] - Eg_t[k_0, X_t, \beta])\|_m\} = 0$, *and*

(b) $c_n(\beta_0) \sum_{t=1}^n Eg_t[k_0, X_t, \beta]$ *is continuous on* $\beta \in B_0$, *uniformly in* n, *where* $l_0 = l(k_0)$ *and* $B_0 = \{\beta \in \mathbb{R}^{l_0}: \|\beta - \beta_0\|_{l_0} \leq \tau(\beta_0)\}$.

ASSUMPTION B6 (LIPSCHITZ BOUNDEDNESS). *There is a function* $c_n(\beta_0)$ *of* n *and* β_0 *and a function* $G(x, \beta_0)$ *such that*

$$\overline{\lim_{n\to\infty}}\ P\Big[|c_n(\beta_0)| \sum_{t=1}^n B_t(k_0, X_t, \beta_0) \geq x\Big] \leq G(x, \beta_0),$$

$$\forall x \geq 0, \text{ and } \lim_{x\to\infty} G(x, \beta_0) = 0.$$

Assumption B5 is a uniform convergence assumption. For the case $c_n(\beta_0) = n^{-1}$, fairly general conditions under which B5 holds are given by Andrews (1987), Gallant (1987a, Chap. 7), Gallant and White (1988, Chap. 2) and Pötscher and Prucha (1986, 1989). Assumption B6 sets an additional restriction on the Lipschitz condition B4 by assuming that $|c_n(\beta_0)| \sum_{t=1}^n B_t(k_0, X_t, \beta_0)$ is bounded in probability. Note that B6 is weaker than the more familiar condition $\overline{\lim}_{n\to\infty} c_n(\beta_0) \sum_{t=1}^n EB_t(k_0, X_t, \beta_0) < \infty$; see Andrews (1987). Using these conditions, we have the following asymptotic equivalence result.

PROPOSITION 3.2 (LAW OF LARGE NUMBERS EQUIVALENCE). *Let Assumptions B1 and B2 hold. If B5 holds or B4 and B6 hold, then*

$$\text{(3.3)}\qquad \text{plim}_{n\to\infty}\ c_n(\beta_0) \sum_{t=1}^n \{g_t[\hat{k}_n, X_t, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X_t, \beta_0]\} = 0.$$

Taking $c_n(\beta_0) = n^{-1}$ and $g_t = |f_{jt}|^r$, where $r > 0$, it is easy to see that

$$\text{(3.4)}\qquad \text{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \{|\tilde{u}_{jt}|^r - |u_{jt}|^r\} = 0, \quad j = 1, \dots, m,$$

where $\tilde{u}_{jt} = f_t[\hat{k}_n, X_t, \tilde{\beta}_n(\hat{k}_n)]$ and $u_t = f_t[k_0, X_t, \beta_0]$. Provided the law of large numbers holds for $|u_{jt}|^r$, $(1/n) \sum_{t=1}^n |\tilde{u}_{jt}|^r$ and $(1/n) \sum_{t=1}^n |u_{jt}|^r$ will converge to the same limit. If the process u_t is strictly stationary and ergodic with $\gamma_{jr} = E|u_{jt}|^r < \infty$, we thus have

$$\text{(3.5)} \qquad \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} |\bar{u}_{jt}|^r = \gamma_j .$$

It is straightforward to see that a similar condition holds for the estimation of the covariances or any other (cross-) moment of the u_t process. For the special case where only one model is considered ($\hat{k}_n = k_0$), we get the following useful corollary.

COROLLARY 3.2. *Let $\{X_t : t \in \mathbb{I}\}$ and $\{\tilde{\beta}_n : n \in \mathbb{N}\}$ be collections of random vectors defined on a common probability space $(\Omega, \mathcal{A}, P)$ and such that* $\operatorname{plim}_{n\to\infty} \tilde{\beta}_n = \beta_0 \in \mathbb{R}^l$. *Let $\{g_t(x, \beta) : t \in \mathbb{I}\}$ be a collection of functions from $\mathbb{R}^{q+l}$ to $\mathbb{R}^m$ ($x \in \mathbb{R}^q$, $\beta \in \mathbb{R}^l$), such that $g_t(X_t, \beta)$ and $g_t(X_t, \tilde{\beta}_n)$ are random vectors for all $\beta \in \mathbb{R}^l$. Further, suppose there is a constant $\tau(\beta_0) > 0$ such that $\beta \in B_0 = \{\beta \in \mathbb{R}^l : \|\beta - \beta_0\|_l \leq \tau(\beta_0)\}$ implies one of the following two conditions:*

(i) $\operatorname{plim}_{n\to\infty} \{\sup_{\beta \in B_0} \|(1/n) \sum_{t=1}^{n} [g_t(X_t, \beta) - Eg_t(X_t, \beta)]\|_m\} = 0$ *and $(1/n) \sum_{t=1}^{n} Eg_t(X_t, \beta)$ is continuous on $\beta \in B_0$ uniformly in n; or*

(ii) $\|g_t(X_t, \beta) - g_t(X_t, \beta_0)\|_m \leq B_t(X_t, \beta_0) h[d(\beta, \beta_0), \beta_0]$, $\forall t$, *a.s., and* $\overline{\lim}_{n\to\infty} P[(1/n) \sum_{t=1}^{n} B_t(X_t, \beta_0) \geq x] \leq G(x, \beta_0)$, $\forall x \geq 0$,

where B_t, h, d and G are nonnegative real-valued functions such that $B_t(X_t, \beta_0)$ is a random variable ($\forall t$), $d(\beta, \beta_0) \to 0$ as $\|\beta - \beta_0\|_l \to 0$, $\lim_{y \downarrow 0} h(y, \beta_0) = 0$ and $\lim_{x\to\infty} G(x, \beta_0) = 0$. Then

$$\text{(3.6)} \qquad \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} [g_t(X_t, \tilde{\beta}_n) - g_t(X_t, \beta_0)] = 0.$$

4. MARKOV PREDICTIVE TESTS

The predictive test statistics introduced in Section 2 have well-known asymptotic distributions when observations are generated by a Gaussian process. Assuming normality is not always appropriate, however. To deal with the possibility of non-Gaussian disturbances, we study two approaches, namely: (1) methods based on Markov inequalities, and (2) methods based on a semi-nonparametric estimation of the (conditional) distribution of the process $\{u_t\}$. The first approach is studied in this section, while the second one will be considered in the next section.

Let X be a real random variable and $g : \mathbb{R} \to \mathbb{R}$ a nonnegative Borel function such that $g(x) = g(-x)$ and $g(x)$ is nondecreasing for $x \geq 0$. Then we have the inequality

$$\text{(4.1)} \qquad P[|X| \geq \lambda] \leq E[g(X)]/g(\lambda), \quad \forall \lambda \geq 0,$$

where we set $0/0 = 1$; see Loève (1977, p. 159). Natural choices for $g(x)$ include $g_r(x) = |x|^r$ and

(4.2) $$g_r^c(x) = |x|^r, \quad \text{if } |x| < c,$$
$$= c^r, \quad \text{if } |x| \geq c,$$

where $0 < r < \infty$ and $0 < c \leq \infty$ are fixed constants. Clearly $g_r^\infty(x) = g_r(x)$. The first choice leads to Chebyshev-type inequalities and requires the evaluation of an absolute moment. The second choice leads one to consider truncated moments. An important advantage of the latter is that truncated moments always exist, provided $0 < c < \infty$. The probability and the expected value in (4.1) may also be taken conditionally on any appropriate set of variables (e.g., past observations).

In the present context, the random variable X may be any component of u_t for $t \in T_2$, or any of the predictive test statistics introduced in Section 2. For simplicity, let us discuss first the case where $X = u_{jt}$. Then

(4.3) $$p_{jt}(\lambda, g) \equiv E[g(u_{jt})]/g(\lambda) = \nu_{jt}(g)/g(\lambda),$$

where $u_{jt} = f_{jt}[k_0, X_t, \beta_0]$, is an upper bound on $P[|u_{jt}| > \lambda]$. If the realization $u_{jt} = \bar{u}_{jt}$ is observed, $p_{jt}(|\bar{u}_{jt}|, g)$ yields an upper bound on the p-value for a test that rejects the null hypothesis of stability when $|\bar{u}_{jt}|$ is large. But, of course, neither u_{jt} nor the moment $\nu_{jt}(g)$ can typically be observed. If $\bar{u}_{jt}$ (the realized value of u_{jt}) and $\nu_{jt}(g)$ are replaced by consistent estimates $\tilde{u}_{jt}$ and $\tilde{\nu}_{jt}(g)$ (as $n_1 \to \infty$), it is clear that

(4.4) $$\operatorname*{plim}_{n_1 \to \infty} \{[\tilde{\nu}_{jt}(g)/g(\tilde{u}_{jt})] - p_{jt}(|\bar{u}_{jt}|, g)\} = 0$$

for all values $\bar{u}_{jt}$ in a set with probability 1. Similarly, if Ψ is some appropriate information set (which may depend on t), the quantity

(4.5) $$p_{jt}(\lambda, g|\Psi) \equiv E[g(u_{jt})|\Psi]/g(\lambda) \equiv \nu_{jt}(g|\Psi)/g(\lambda)$$

is an upper bound on $P[|u_{jt}| > \lambda|\Psi]$. Again, provided $\tilde{\nu}_{jt}(g|\Psi)$ is a consistent estimator of $\nu_{jt}(g|\Psi)$ as $n_1 \to \infty$, we have

(4.6) $$\operatorname*{plim}_{n_1 \to \infty} \{[\tilde{\nu}_{jt}(g|\Psi)/g(\tilde{u}_{jt})] - p_{jt}(|\bar{u}_{jt}|, g|\Psi)\} = 0.$$

For example, consider unconditional tests and assume that the u_t process is strictly stationary and ergodic. Then, using the asymptotic equivalence results of Proposition 3.2, one sees easily that

(4.7) $$\tilde{\nu}_{jt}(n_1, g_r^c) = \frac{1}{n_1} \sum_{s=-n_1+1}^{0} g_r^c[\tilde{u}_{js}(n_1)]$$

is a consistent estimator of $\nu_{jt}(g_r^c) = E[g_r^c(u_{jt})]$ as $n_1 \to \infty$, where g_r^c is given by (4.2). Here, under the assumption of stationarity, the first sample T_1 can be used to estimate the moments of u_{jt} for $t \in T_2$. Thus the critical region $\tilde{p}_{jt}(\tilde{u}_{jt}(n_1), g_r^c) \leq \alpha$, where

$$\bar{p}_{jt}(\bar{u}_{jt}(n_1),\, g_r^c) = \bar{\nu}_{jt}(n_1,\, g_r^c)/g_r^c[\bar{u}_{jt}(n_1)], \tag{4.8}$$

is conservative at level α.

Naturally, the choice of r and c is arbitrary. Suppose, however, that the pairs (r, c) are selected from a finite set $S = \{(r_j, c_j): j = 1, \dots, N\}$. Since (4.1) holds for all $g = g_r^c$ such that $0 < r < \infty$ and $0 < c \le \infty$, it follows that

$$P[|X| \ge \lambda] \le \min\{E[g_r^c(X)]/g_r^c(\lambda): (r, c) \in S\}. \tag{4.9}$$

Consequently, if we compute several bounds based on different functions g_r^c, we can obtain an improved bound by taking the lowest of these bounds. Consequently, $\min\{\bar{\nu}_{jt}(g_r^c)/g_r^c(\bar{u}_{jt}): (r, c) \in S\}$ is an asymptotically valid bound on the (unconditional) p-value for $|\bar{u}_{jt}|$, and similarly for $\min\{\bar{\nu}_{jt}(g_r^c|\Psi)/g_r^c(\bar{u}_{jt}): (r, c) \in S\}$ for the conditional p-value (given Ψ).[6]

The above method for constructing upper bounds on p-values for realizations of $\{u_t: t \in T_2\}$ can be extended to $\bar{v}_{jt}$, $\bar{w}_t$, $\bar{U}_j$ and $\bar{W}$ as defined in Section 2. For $\bar{v}_{jt}$ and $\bar{w}_t$, the approach is exactly the same except that $\bar{u}_{jt}$ is replaced by $\bar{v}_{jt}$ and $\bar{w}_t$. For $\bar{U}_j(n_1)$ and $\bar{W}(n_1)$, the construction of *upper bounds on p-values* is slightly more complicated because each of these statistics uses n_2 predicted residuals. To apply an approach similar to the one just described for $\bar{u}_{jt}$, $\bar{v}_{jt}$ and $\bar{w}_t$, we need to estimate the expected values of $g(U_j)$ and $g(W)$ in a consistent way from the sample T_1, where U_j and W are defined as $\bar{U}_j(n_1)$ and $\bar{W}(n_1)$ with $\bar{u}_{jt}(n_1)$, $\bar{\Delta}_j(n_1)$ and $\bar{\Delta}(n_1)$ replaced by their probability limits (as $n_1 \to \infty$). This can be done, in particular, if n_1 is sufficiently large with respect to n_2 ($n_2 \ll n_1$). From T_1, one can form $n_1 - n_2 + 1$ subsamples of size n_2. Then an asymptotically valid upper bound on the p-value for $\bar{U}_j(n_1)$ is

$$\bar{p}_j(\bar{U}_j(n_1),\, g_r^c) = \left[\frac{1}{n_1 - n_2 + 1}\sum_{\tau=-n_1+n_2}^{0} g_r^c[\bar{U}_j(n_1, \tau)]\right]\Big/ g_r^c[\bar{U}_j(n_1, \tau)] \tag{4.10}$$

where $\bar{U}_j(n_1, \tau) = \bar{u}_j(n_1, \tau)'\bar{\Delta}_j(n_1)^{-1}\bar{u}_j(n_1, \tau)$ and $\bar{u}_j(n_1, \tau) = [\bar{u}_{j,\tau-n_2+1}(n_1), \dots, \bar{u}_{j\tau}(n_1)]'$. Note that $\bar{u}_j(n_1, n_2) = \bar{u}_j(n_1)$. An upper bound on the p-value for the joint test $\bar{W}$ is constructed along the same principles:

$$\bar{p}_W(\bar{W},\, g_r^c) = \left(\frac{1}{n_1 - n_2 + 1}\sum_{\tau=-n_1+n_2}^{0} g_r^c[\bar{u}(n_1, \tau)'\bar{\Delta}(n_1)^{-1}\bar{u}(n_1, \tau)]\right)\Big/ g_r^c[\bar{W}(n_1)] \tag{4.11}$$

[6] When using higher order moments with moderately sized samples, some caution may need to be exercised because higher order moments converge slowly to their limits. There has recently been some discussion of this issue in the context of the information matrix test, and simulation evidence indicates this slow convergence can have a nonnegligible impact on the finite sample properties of the test; see Orme (1990), Chesher and Spady (1991) and Kennan and Neumann (1988). However, higher order moments are used here to *bound p*-values: since the resulting tests are generally conservative, the fact that estimated higher order moments have a large variance does not imply that the resulting p-values will loose their conservative character. Using truncated moments as suggested above can also have an additional stabilizing effect. Further research on these issues would certainly be of interest, but is beyond the scope of the present paper.

where $\tilde{u}(n_1, \tau) = [\tilde{u}_{\tau-n_2+1}(n_1)', \dots, \tilde{u}_\tau(n_1)']'$ and $\tilde{u}(n_1, n_2) \equiv \tilde{u}(n_1)$.

5. SEMI-NONPARAMETRIC PREDICTIVE TESTS

The Markov inequalities used in the previous section provide conservative probability statements. To avoid this problem (but at the expense of further regularity conditions and computation costs), we propose now to calculate the p-values of the predictive tests via a semi-nonparametric (SNP) estimation of the *conditional* distribution of the test statistics.

In Section 2, we observed that if we knew the distributions (conditional or not) of the test statistics up to a finite-dimensional parameter vector θ, we would be able to exploit the asymptotic equivalence results of Section 3 to compute the p-values of our tests. The situation we consider here is different, in the sense that we assume a flexible functional form for the density function proposed by Phillips (1983), Gallant and Nychka (1987) and Gallant and Tauchen (1989). Let us denote by $G_X(x; K, \theta(K)|\Psi)$ the conditional probability density function (p.d.f.) of X given Ψ, where X is a random variable that may stand for any of the test statistics considered. The parameter vector $\theta(K)$ describes the polynomials of the SNP expansion.[7] Specific examples would be: $G_{v_{jt}}(x; K, \theta(K)|\Psi_{t-1})$ and $G_{w_t}(x; K, \theta(K)|\Psi_{t-1})$ which are conditioned on Ψ_{t-1}, and $G_{U_j}(x; K, \theta(K)|\Psi_0)$ and $G_W(x; K, \theta(K)|\Psi_0)$ conditioned on Ψ_0. The index parameter K appears as one of the arguments because the SNP approach uses a Hermite polynomial as a general approximation to the conditional density. The polynomial expansions are truncated, with the truncation rule depending on the sample size. Following Gallant and Tauchen (1989), we consider the following generic representation of the conditional density:

$$G(x; K, \theta(K)|\Psi) = \frac{\left[\sum_{|\alpha|=0}^{K_1(n_1)} a_\alpha(\Psi)x^\alpha\right]^2 \phi(x|\Psi)}{\int_{-\infty}^{\infty}\left[\sum_{|\alpha|=0}^{K_1(n_1)} a_\alpha(\Psi)x^\alpha\right]^2 \phi(x|\Psi)\,dx} \tag{5.1}$$

where (i) $a_\alpha(\Psi)$ is a linear polynomial of degree $K_2(n_1)$ defined on a finite subset of Ψ; (ii) the index K represents the pair $(K_1(n_1), K_2(n_1))$ of truncation parameters depending on n_1; (iii) $\theta(K)$ is the unknown parameter vector describing the parameters of the $(K_1(n_1), K_2(n_1))$ Hermite polynomial expansion; (iv) $\phi(x|\Psi)$ is the "lead term" of the expansion which is itself a probability density function.

Several aspects of this approximation are worth noting. Firstly, Gallant and Nychka (1987) verify a set of conditions under which $G_X(\cdot|\Psi)$ perfectly approximates the true underlying conditional p.d.f. as $n_1 \to \infty$ for a wide class of distributions.[8] Gallant and Tauchen (1989) provide a data-based strategy for

[7] The parameter vector β determines, of course, the processes v_{jt}, w_t, etc. (see Section 2).

[8] Gallant and Tauchen (1989) observe that a M-dimensional SNP expansion $G(x)$ can approximate arbitrarily accurately distributions with fat tails such as those proportional to $(1 + x'x)^{-\delta}$, $\delta > M/2$, and

choosing $K_1(n_1)$ and $K_2(n_1)$ for a given n_1. Secondly, we are free to choose any p.d.f. as the lead term, but the choice may affect the finite-sample properties of the approximation. Following Gallant and Tauchen (1989), we chose $\phi(\cdot)$ so that a special case of interest is nested within the approximation. For v_{jt}, the lead term is chosen to be homoskedastic mean zero normal, while for w_t, U_j and W, the lead terms are chosen to be $\chi^2(m)$, $\chi^2(n_2)$ and $\chi^2(mn_2)$ respectively. Thus, in each case, the lead term of the expansion is the p.d.f. implied by $\{u_t\}$ being Gaussian. Of course, this choice is somewhat arbitrary and will have an incidence on the finite-sample properties of the procedure.[9] However, if n_1 is large enough, the semi-nonparametric procedure will work well, irrespective of the choice of lead term. Thirdly, to use this approximation, we assume that $\{u_t : t = -n_1 + 1, \ldots, n_2\}$ comes from a realization of a stationary time series $\{u_t\}_{t=-\infty}^{\infty}$. Fourth, we use this framework to model the conditional densities of v_{jt}, w_t, U_j and W. Finally, after estimating these conditional densities, the latter are applied to the observed values of $\tilde{v}_{jt}$, $\tilde{w}_t$, $\tilde{U}_j$ and $\tilde{W}$ to obtain p-values. Under appropriate regularity conditions, the conditional densities of v_{jt}, w_t, U_j and W are consistently estimated by the SNP approximations (as $n_1 \to \infty$), and similarly for the corresponding distribution functions.[10] Further, $\tilde{v}_{jt}$, $\tilde{w}_t$, $\tilde{U}_j$ and $\tilde{W}$ are asymptotically equivalent to v_{jt}, w_t, U_j and W respectively as $n_1 \to \infty$.

To calculate the p-values for the predictive tests using a semi-nonparametric approximation, one proceeds as follows.

1) The model is estimated by maximum likelihood from the sample T_1 using the SNP approximate p.d.f. (5.1). Details about such estimation appear in Gallant and Tauchen (1989).
2) Once the SNP density has been estimated over T_1, yielding the parameter estimate $\hat{\theta}(K)$, the predictive test statistics are computed. For the individual predictive statistics $\tilde{v}_{jt}$ and $\tilde{w}_t$, the conditioning set Ψ will usually be $\Psi_{t-1} = \{X_s : s \le t-1\}$ or $\Psi_0 = \{X_s : s \le 0\}$. For the joint predictive tests,

those with thin tails such as those proportional to $\exp[-(x'x)^{\rho}]$, $1 < \rho < \delta - 1$. Off the tails, the SNP expansion is capable of approximating a wide variety of p.d.f.'s. Essentially, the only types of behaviour ruled out are violent oscillations.

[9] As pointed out by a referee, the choice of lead term may be especially influential for the individual test statistics $\tilde{v}_{jt}$ and $\tilde{w}_t$. On the other hand, for the portmanteau statistics ($\tilde{U}_j$ and $\tilde{W}$) and provided n_2 is large, it is plausible that the choice of lead term will be less important, because $(U_j - n_2)/\sqrt{n_2}$ and $(W - mn_2)/\sqrt{mn_2}$ are approximately normal, just like $(\chi^2(n_2) - n_2)/\sqrt{n_2}$ and $(\chi^2(mn_2) - mn_2)/\sqrt{mn_2}$ are approximately normal for n_2 large. Further work would be useful to assess the sensitivity of the method to the choice of lead term.

[10] General results on the consistency of SNP density estimators appear in Gallant and Nychka (1987). For related results, see also Gallant (1982, 1987b), Eastwood and Gallant (1987), Andrews (1988) and the references therein. These results, however, are better adapted to cross-sectional data. Gallant and Tauchen (1989, p. 1096) observe that while there has been much analysis of SNP models for cross-sectional data, "whether or not these results and methods of proof extend to time series data is an open question." In this paper, we adopt their conjecture that the underlying true p.d.f. can be approximated arbitrarily well by a SNP density as the sample size n_1 tends to infinity. It is important to note that further restrictions may be needed for this to hold. An explicit determination of appropriate conditions is left to future research, and we assume here that such conditions apply. Andrews (1989a, 1989b) provides asymptotic distributions for semiparametric estimators in time series contexts, but his results do not cover Gallant's semi-nonparametric estimators (except for special cases).

the natural conditioning set is Ψ_0. Of course, it is also possible to consider unconditional tests. If $\Psi = \Psi_{t-1}$ in (5.1), densities conditional on Ψ_0 and unconditional densities can be obtained by integration. Such calculations may, however, be costly so that joint SNP predictive tests are often more difficult to implement than individual SNP predictive tests.

3) The p-values of the predictive tests are calculated as follows: for $\bar{v}_{jt}$ and $\bar{w}_t$, with $\Psi = \Psi_{t-1}$, we take

$$\bar{p}(\bar{v}_{jt}) = 1 - \int_{-\infty}^{\bar{v}_{jt}} G_{v_{jt}}(x;\, K,\, \hat{\theta}(K)|\Psi_{t-1})\, dx, \tag{5.2}$$

$$\bar{p}(\bar{w}_t) = 1 - \int_0^{\bar{w}_t} G_{w_t}(x;\, K,\, \hat{\theta}(K)|\Psi_{t-1})\, dx,$$

while for $\bar{U}_j$ and $\bar{W}$, with $\Psi = \Psi_0$, we take

$$\bar{p}(\bar{U}_j) = 1 - \int_0^{\bar{U}_j} G_{U_j}(x;\, K,\, \hat{\theta}(K)|\Psi_0)\, dx, \tag{5.3}$$

$$\bar{p}(\bar{W}) = 1 - \int_0^{\bar{W}} G_W(x;\, K,\, \hat{\theta}(K)|\Psi_0)\, dx,$$

where $t \in T_2$. A two-sided critical region based on $\bar{v}_{jt}$ (with critical values yielding equal right and left tail areas) has the form $\bar{p}(\bar{v}_{jt}) \equiv 2 \min\{\bar{p}(\bar{v}_{jt}),\, 1 - \bar{p}(\bar{v}_{jt})\} \leq \alpha$; the critical regions for the other statistics simply are: $\bar{p}(\bar{w}_t) \leq \alpha$, $\bar{p}(\bar{U}_j) \leq \alpha$ and $\bar{p}(\bar{W}) \leq \alpha$. It is clear how different conditioning sets or unconditional tests could be considered. Again, under appropriate regularity conditions, these estimated p-values are asymptotically equivalent to those based on the "true" densities. For example, this will occur if the SNP estimated densities converge uniformly in x to the true densities.

6. EMPIRICAL EXAMPLES

The purpose of this section is to show how to apply the generalized predictive tests introduced in Sections 2, 4 and 5. Throughout this section, we use the same data to test the stability of two models involving different assumptions, model specifications and levels of generality. The data and sample employed correspond to those used in the empirical study of Gallant and Tauchen (1989). In a companion discussion paper (Dufour, Ghysels, and Hall 1991), we report other examples as well. It should be pointed out that the model specifications studied do not per se contradict each other. They differ in their degree of specificity about distributional assumptions, moment conditions, etc. Both models explain the comovements of asset returns and consumption growth. The estimation and prediction samples are the same. As we shall see, there is a fairly strong agreement across the specifications about the outcome of the predictive stability analysis.

The first model specification is a vector autoregressive time series model of consumption growth and T-bill returns. Such a model can be viewed as the linear projection indeterministic part of a more complex (possibly nonlinearly predictable) data generating process. As a benchmark, we consider first the case where the errors are assumed normal. Then, to deal with the possibility of nonnormal errors, we consider two approaches: the first one uses Markov inequalities to construct predictive tests, while the second one is based on estimating a semi-nonparametric density for the VAR error process. As noted in Section 5, we use the estimated density to compute p-values for the predicted residuals. While neither the tests based on the Markov inequalities nor those based on SNP density estimation require normality, it should be noted that both represent different types of tests. The former are unconditional tests, while the latter are conditional. With the second model specification, we move to a nonlinear model characterized by a set of implicit equations. Here no closed-form solution is specified and no specific distributional assumptions are made except, of course, for the relatively mild requirements that ensure the consistency of the parameter estimates. The model is a constant relative risk aversion consumption-based asset pricing model introduced by Hansen and Singleton (1982), estimated with a one month T-Bill return. Ghysels and Hall (1990b) tested its stability with Wald, LM-type, LR-type and predictive tests: the first two tests were introduced by Andrews and Fair (1988), the LR-type test was presented in Eichenbaum, Hansen, and Singleton (1988) and adapted to test structural stability by Ghysels and Hall (1990b), while the predictive test was introduced by Ghysels and Hall (1990a). The four tests applied by Ghysels and Hall (1990b) to the Hansen-Singleton (henceforth HS) empirical asset pricing models were based on two crucial assumptions: (1) the sample split is known and (2) the two subsamples grow proportionally, i.e., n_1/n_2 is constant. Both assumptions can be relaxed with the tests presented in Sections 2, 4 and 5. The asset pricing models can now be scrutinized without assuming a priori that, for instance, the October 1979 monetary policy shift is a breakpoint. Equally important is the fact that the tests presented in this paper allow one to test stability with a small prediction subsample, i.e., n_2 does not have to be proportional to the first sample size n_1 (which is assumed to be large).

The models are all estimated from the same set of monthly data covering the period 1959:1 until 1978:12. The variables considered are per capita consumption of nondurable goods and services (c_t) and the return on one-month T-Bills corrected for inflation using the price index corresponding to the consumption measure (r_t). The prediction sample consists of two years of observations covering 1979 and 1980, which amounts to 24 observations. The test results will be reported in "predictive analysis tables" (PAT), as described in Section 2.

The VAR model is based on the variables $\log(r_t)$ and $\log(c_t/c_{t-1})$. In estimating the latter, we followed closely Gallant and Tauchen (1989) with regard to data transformations and normalizations, before and after fitting the VAR model. This will facilitate comparison with the SNP-based tests discussed later. In particular, before fitting the VAR model by least squares, the vector $[\log(r_t), \log(c_t/c_{t-1})]'$ is centered by subtracting the sample mean and is normalized by multiplying the centered data by the inverse of the upper triangular Choleski factorization matrix associated with the empirical covariance matrix of the observations. The autoregressions include two lags of each variable. The least squares residuals from these regressions are also normalized by multiplying them by the inverse of the Choleski matrix associated with their empirical covariance matrix. The predicted residuals are computed in a similar way, using the parameter estimates and the Choleski matrix obtained from the estimation sample.[11]

Let us consider first the results obtained under the assumption that the disturbances in the VAR are normally distributed. The predicted residuals for the 24 months of 1979–80 and their associated p-values appear in Table 6.1. This table has the structure of a PAT discussed in Section 2. Each equation is listed separately first for each observation in the prediction sample, and then for all observations at once (bottom of the table). Finally, the predicted residuals and p-values for the joint set of equations appear in the last pair of columns. Three observations stick out with low p-values (say, less than or equal to 0.05): 79:12 (for the T-bill equation), 80:2 (for both equations, individually and jointly), 80:10 (for the consumption equation) and 80:12 (for the T-Bill equation). The rejection is especially strong for 80:2. p-values less than 0.10 also include 80:10 and 80:12 when the equations are taken jointly. Of course, we know that October 1979 is the month of a major policy announcement of the Federal Reserve. We will discuss this subject further as we review the evidence from the other models to be presented. None of the equations, taken over the entire two-year sample, separately appears to be unstable. Let us now drop the normality assumption and consider unconditional generalized predictive tests based on the second, fourth, sixth and eight moment Markov inequalities. The lowest of these four bounds is reported in Table 6.1. We see that the results are qualitatively similar to those obtained under the normal assumption, although 79:12 and 80:10 are certainly less significant than in the normal case. The observation for 80:2 still stands out as special despite the fact that we do not use the normality assumption.

Next, we consider conditional predictive tests where the conditional distributions of the observations are approximated by SNP expansions based on the work of Gallant and Tauchen (1989), as described in Section 5. The estimation sample and data transformations used to estimate the VAR model were also used in this approach. Instead of considering the linear projection error processes obtained from a VAR, we use the Hermite polynomial described in (5.1) to approximate the conditional joint density of $\log(r_t)$ and $\log(c_t/c_{t-1})$. The information set in (5.1) denoted Ψ was set equal to Ψ_{t-1} which consists of past realizations of T-Bill returns and consumption growth.[12] The SNP density was estimated with $K_1(n_1) = 2$ and $K_2(n_1) = 1$ involving 42 parameters (corresponding to the fourth entry of Gallant and Tauchen (1989, Table III). Via numerical integration of the conditional

[11] Because of the Choleski normalizations the "transformed interest rate variable" can be interpreted (asymptotically) as a linear transformation of the original variables. The interest rate equation and residuals should be interpreted accordingly. Such a reinterpretation is not required for the consumption growth equation, because it is only rescaled and reparameterized.

[12] The VAR parameters were reestimated jointly with the parameters of the Hermite expansion and hence are different from the VAR estimates obtained by OLS. The latter were used as starting values in the SNP estimation program.

TABLE 6.1
PREDICTIVE ANALYSIS TABLE FOR THE VAR MODEL OF CONSUMPTION AND T-BILL RETURNS: UNCONDITIONAL TESTS BASED ON GAUSSIAN ASSUMPTION AND MARKOV INEQUALITIES ESTIMATION SAMPLE. 1959–78. PREDICTION SAMPLE: 1979–80

	T-Bill Equation			Consumption Equation			Joint Set of Equations		
	Predicted residuals	p-values normal	p-values Markov	Predicted residuals	p-values normal	p-values Markov	Tests ($\bar{U}_j$)	p-values normal	p-values Markov
79:1	−1.014	0.310	0.97(2)	−0.667	0.504	1.00	1.417	0.492	1.00
2	−0.523	0.600	1.00	−0.229	0.819	1.00	0.316	0.854	1.00
3	−0.121	0.904	1.00	−1.094	0.274	0.83(2)	1.201	0.548	1.00
4	−0.220	0.825	1.00	−0.947	0.344	1.00	0.928	0.629	1.00
5	−0.973	0.330	1.00	−0.827	0.408	1.00	1.563	0.458	1.00
6	0.640	0.522	1.00	−0.501	0.616	1.00	0.691	0.708	1.00
7	0.512	0.609	1.00	−0.746	0.455	1.00	0.854	0.652	1.00
8	−0.128	0.899	1.00	−0.249	0.803	1.00	0.076	0.963	1.00
9	0.245	0.807	1.00	1.075	0.282	0.87(2)	1.194	0.550	1.00
10	−0.571	0.568	1.00	−0.296	0.767	1.00	0.399	0.819	1.00
11	0.119	0.905	1.00	−1.332	0.183	0.56(2)	1.806	0.405	1.00
12	−1.957	0.050*	0.25(4)	0.392	0.694	1.00	4.061	0.131	1.00
80:1	−0.581	0.561	1.00	−0.132	0.895	1.00	0.348	0.840	1.00
2	−2.813	0.005*	0.05(6)*	2.020	0.043*	0.23(4)	12.523	0.002*	0.04(4)*
3	1.060	0.289	0.89(2)	−0.704	0.482	1.00	1.555	0.460	1.00
4	1.240	0.215	0.65(2)	0.136	0.892	1.00	1.545	0.462	1.00
5	−0.004	0.997	1.00	−1.425	0.154	0.49(2)	2.034	0.362	1.00
6	0.114	0.909	1.00	−1.605	0.109	0.38(2)	2.610	0.271	1.00
7	−0.727	0.467	1.00	−0.881	0.378	1.00	1.250	0.535	1.00
8	0.648	0.517	1.00	0.528	0.598	1.00	0.669	0.716	1.00
9	−0.745	0.456	1.00	0.135	0.892	1.00	0.583	0.747	1.00
10	0.038	0.969	1.00	2.440	0.015*	0.11(4)	5.959	0.051	0.49(2)
11	−1.131	0.258	0.78(2)	−0.098	0.922	1.00	1.282	0.527	1.00
12	−2.162	0.031*	0.16(4)	0.440	0.660	1.00	4.961	0.084	0.70(2)
79–80	25.612	0.627	1.00	23.77	0.525	1.00	49.826	0.401	1.00

Note: Predicted residuals are normalized as in Gallant and Tauchen (1989) via the Choleski factorization matrix obtained from the estimation sample, using OLS as the estimation procedure. The Markov bound is based on the minimum of the Markov inequalities of orders 2, 4, 6 and 8. The order of Markov inequality appears between parentheses. The p-values are based on the absolute values of the statistics considered (i.e., they correspond to two-sided tests). A star * indicates a p-value not greater than 0.05.

TABLE 6.2
PREDICTIVE ANALYSIS TABLE FOR A VAR MODEL OF CONSUMPTION AND T-BILL RETURNS BASED ON SEMI-NONPARAMETRIC ESTIMATION OF THE CONDITIONAL DENSITIES OF THE PREDICTED RESIDUALS

	T-Bill Equation			Consumption Equation		
	Predicted residuals (centered)	SNP p-values		Predicted residuals (centered)	SNP p-values	
		$\tilde{p}$	$\bar{p}$		$\tilde{p}$	$\bar{p}$
79:1	−1.150	0.910	0.180	0.430	0.530	0.940
2	−0.744	0.531	0.938	−0.635	0.799	0.402
3	0.152	0.538	0.924	−1.209	0.899	0.202
4	−0.161	0.538	0.924	−1.578	0.872	0.256
5	−0.878	0.844	0.312	−1.502	0.850	0.300
6	0.882	0.348	0.696	−1.558	0.931	0.138
7	0.612	0.318	0.636	−0.912	0.738	0.524
8	−0.427	0.691	0.618	−0.805	0.532	0.936
9	0.057	0.594	0.812	0.872	0.195	0.390
10	−0.784	0.840	0.320	0.798	0.411	0.822
11	−0.124	0.539	0.922	−1.607	0.968	0.064*
12	−1.998	0.989*	0.022*	0.370	0.310	0.620
80:1	−0.278	0.555	0.890	−0.378	0.611	0.778
2	−2.672	0.992*	0.016*	3.855	0.000*	0.000*
3	−0.703	0.751	0.482	0.526	0.629	0.742
4	1.584	0.304	0.608	0.691	0.505	0.990
5	−0.021	0.452	0.904	−0.934	0.764	0.472
6	−0.238	0.584	0.832	−3.153	0.993*	0.014*
7	−0.628	0.795	0.410	−2.135	0.933	0.134
8	0.839	0.482	0.964	−0.568	0.692	0.616
9	−0.800	0.862	0.276	1.115	0.177	0.354
10	−0.341	0.522	0.956	2.955	0.011*	0.022*
11	−1.314	0.923	0.154	2.399	0.035	0.070
12	−2.601	0.897	0.206	1.514	0.219	0.438

Notes: Predicted residuals are normalized as in Gallant and Tauchen (1989) via the Choleski factorization matrix obtained from the estimation sample, using the SNP estimation procedure, and then centered (conditional on Ψ_{t-1}). The p-value $\tilde{p}$ is the probability that the residual is larger than the observed value (conditional on Ψ_{t-1}). Therefore, either large or small p-values may be indicative of structural instability. A star * indicates a value of $\tilde{p}(\tilde{v}_{jt})$ less than 0.025 or greater than 0.975, while $\bar{p} = 2 \min\{\tilde{p}, 1 - \tilde{p}\}$ is the marginal significance level associated with a two-sided test based on $\tilde{v}_{jt}$ (with critical values yielding equal right and left tails).

SNP density we obtained an estimate, denoted $\tilde{u}_{1t}(n_1)$, of the prediction error process for T-Bills $u_{1t} = \log(r_t) - E[\log(r_t)|\Psi_{t-1}]$. Likewise, an estimate $\tilde{u}_{2t}(n_1)$ of $u_{2t} = \log(c_t/c_{t-1}) - E[\log(c_t/c_{t-1})|\Psi_{t-1}]$ was computed from the estimated conditional density. Subtracting the conditional means is important to ensure that the conditional mean of the disturbances be zero. To calculate p-values for both series of predicted residuals, we computed, again by numerical integration, the (estimated) conditional densities of $\tilde{u}_{1t}(n_1)$ and $\tilde{u}_{2t}(n_1)$.

In Table 6.2, we report the predicted residuals for T-Bills and consumption growth as well as their p-values computed from the SNP densities for u_{1t} and u_{2t}. The residuals in Table 6.2 are numerically different from the VAR residuals because the latter are from a linear projection of $\log(r_t)$ and $\log(c_t/c_{t-1})$ on Ψ_{t-1}, while the former are computed from the conditional expectations implied by the estimated SNP densities. The results in Table 6.2 generally agree with those in Table

6.1.[13] We observe again p-values less than 0.05 for 79:12 (for the T-Bill equation), 80:2 (for both equations, very strongly) and 80:10 (for the consumption equation). But some other predicted residuals exhibit low conditional p-values, especially in the consumption equation: 79:11, 80:6 and 80:11. These results are not surprising since: (1) we dropped the normality assumption, which Gallant and Tauchen (1989) found to be inappropriate for the data being considered, and (2) we now consider conditional instead of unconditional distributions. It was noted in Section 2 that conditional tests are likely to be more powerful than unconditional tests.

Unlike the VAR, the second model we consider does not have a closed-form solution. We reestimated the Hansen and Singleton (1982, henceforth HS) model with an instrument set slightly different from the original set. Besides a constant, lagged consumption growth and lagged T-Bill return, we introduced lagged detrended money growth as an instrument. Money growth was detrended by subtracting a linear trend as suggested by Stock and Watson (1989). This trend was estimated with data until 78:12. The moment restrictions implied by the HS model are described as follows:

$$f_t[X_t, \beta] \equiv \left(b \left(\frac{c_t}{c_{t-1}} \right)^{\alpha} r_t - 1 \right) \otimes Z_{t-1} \tag{6.1}$$

where c_t/c_{t-1} is the consumption growth (nondurables plus services, denoted NDS) and r_t is the asset return, i.e., the one month T-Bill return corrected for inflation (TB1).[14] The parameter b measures a subjective discount rate while α measures constant relative risk aversion. The vector Z_t is the instrument set

$$Z_t \equiv (1, c_t/c_{t-1}, r_t, \tilde{m}_t)', \tag{6.2}$$

i.e., a constant, growth in NDS, TB1 and linearly detrended nominal M1 growth (denoted M1G).

Equations (6.1) and (6.2) yield a vector function of dimension $m = 4$. The unconditional statistics $\bar{v}_{jt}(n_1)$, $\bar{w}_t(n_1)$, $\bar{U}_j(n_1)$ and $\bar{W}(n_1)$, for $j = 1, \dots, 4$ and $t \in T_2$, are considered once again. For the HS model, the process $\{f_t\}$ defined by equation (6.1) is a martingale difference sequence. This property greatly simplifies the joint predictive tests $\bar{U}_j, j = 1, \dots, 4$, and $\bar{W}$. The parameter estimates for the HS model appear in Table 6.3. The results are similar to previously reported estimates (see Hansen and Singleton 1982, and Ghysels and Hall 1990b). According to the overidentifying restrictions test, the T-Bill model is rejected.

While the overidentifying restrictions do not support the model empirically, it is still useful to conduct the type of tests we propose as part of the set of diagnostic tests one would like to consider. In particular, this may provide us more information about the nature of the problem: which months and equations are likely to be most related to a breakdown of the model. Here, we do not assume normality

[13] These results should be interpreted in view of the caveats of Section 5 (footnote 9) on the consistency of SNP density estimator in time series contexts.

[14] It should be noted that there are some differences in the data transformations between (6.1) and the VAR model. In the VAR the logarithms of c_t/c_{t-1} and r_t were taken and normalized as previously described. In contrast, c_t/c_{t-1} and r_t enter directly into the moment condition (6.1).

TABLE 6.3
PARAMETER ESTIMATES OF THE HS MODEL WITH T-BILL RETURNS: 59:02–78:12

b	1.000
	(0.0003)
α	−0.1594
	(0.1083)
SDMR1	0.0385
SDMR2	0.0387
SDMR3	0.0385
SDMR4	0.0001
HORT	5.7080

Notes: b and α are the parameters defined in equation (6.1). SDMRi is the standard deviation of the ith moment restriction $i = 1, \ldots, 4$. HORT = Hansen's overidentifying restrictions test.

and apply the unconditional predictive tests based on the Markov inequalities. The PAT table and the Markov upper bounds on their p-values are reported in Table 6.4.[15] According to the individual predictive tests, it is interesting to observe that the fourth equation, which involves money growth, does not exhibit the same out-of-sample pattern as the other three equations. Firstly, it should be noted that the null hypothesis of structural stability is rejected at the 10 percent level for the moment conditions involving money over the entire two-year sample (bottom of Table 6.4) on the basis of a second order Markov inequality. Looking then at individual observations, we see that the predicted residuals of the money equation exhibit several relatively low p-values (79:9, 80:2, 80:5, 80:6, 80:7, 80:10) just before and after October 1979 which is the timing of the announcement of the Federal Reserve monetary policy regimes shift; the marginal significance level is especially small for 80:2 (below 0.002). In addition, as with the VAR model, 80:2 appears with low p-values in the three other equations as well as for the equations taken jointly. The fact that the moment condition involving money growth appears so different during a period of monetary policy transition suggests questions about the choice of the instruments used in the estimation. Such questions require more investigation, which is beyond the scope of the present paper.

7. CONCLUDING REMARKS

In this paper, we have proposed predictive tests for analyzing the structural stability of a nonlinear dynamic simultaneous equations model. The tests can be applied when the model is structurally stable during an asymptotically large

[15] In the context of model (6.1), the normality assumption seems difficult to maintain and so we do not report p-values based on this distribution in the main text. It would imply that the Euler equation disturbance is normally distributed, as well as the products of the latter with lagged consumption growth, interest and money growth. However, it is straightforward to perform significance tests under the assumption that $f_t[X_t, \beta_0]$ is normal. In this case, the global test $\bar{W} = 204.0$ and the joint test $\bar{U}_4 = 151.8$ (for the money equation) are strongly significant (at levels much lower than 0.05), while the joint tests for the other equations do not come out significant (at levels much lower than 0.05), while the joint tests for the other equations do not come out significant. The money equation has several statistically large predicted residuals (79:9, 80:2, 80:5, 80:6, 80:7, 80:10), while 80:2 and (to a lesser extent) 80:10 appear to be outlying in all the equations.

TABLE 6.4
PREDICTIVE ANALYSIS TABLE FOR THE HANSEN-SINGLETON MODEL WITH T-BILL RETURNS
ESTIMATION SAMPLE: 1959–78. PREDICTION SAMPLE: 1979–80

	Euler eq.		Euler eq. * NDS (−1)		Euler eq. * TB1 (−1)		Euler eq. * M1G (−1)		System	
	Predicted residuals	p-values Markov	Predicted residuals	p-values Markov	Predicted residuals	p-values Markov	Predicted residuals	p-values Markov	Tests ($\bar{U}_j$)	p-values Markov
79:1	−0.2263	1.0000	−0.2261	1.0000	−0.2270	1.0000	0.1023	1.0000	0.0621	1.0000
2	−0.2851	1.0000	−0.2830	1.0000	−0.2853	1.0000	0.3261	1.0000	0.2827	1.0000
3	−0.9064	1.0000	−0.9004	1.0000	−0.9049	1.0000	−0.6022	1.0000	1.8091	1.0000
4	−0.7910	1.0000	−0.7871	1.0000	−0.7898	1.0000	0.7371	1.0000	1.5490	1.0000
5	−0.2637	1.0000	−0.2684	1.0000	−0.2633	1.0000	−0.6181	1.0000	0.5091	1.0000
6	−0.7577	1.0000	−0.7531	1.0000	−0.7561	1.0000	−1.2830	0.6077(2)	3.3864	1.0000
7	−0.7204	1.0000	−0.7211	1.0000	−0.7185	1.0000	−0.8724	1.0000	1.5125	1.0000
8	−0.1569	1.0000	−0.1568	1.0000	−0.1559	1.0000	−0.2180	1.0000	0.0833	1.0000
9	0.9480	1.0000	0.9440	1.0000	0.9454	1.0000	−3.1612	0.1001(2)	11.2391	1.0000
10	0.3895	1.0000	0.3890	1.0000	0.3885	1.0000	0.0014	1.0000	0.2695	1.0000
11	−1.3655	0.5364(2)	−1.3610	0.5397(2)	−1.3661	0.5356(2)	0.4983	0.9438(2)	3.4927	1.0000
12	0.6228	1.0000	0.6228	1.0000	0.6215	1.0000	−1.0291	1.0000	2.1396	1.0000
80:1	0.2318	1.0000	0.2303	1.0000	0.2322	1.0000	0.3050	1.0000	0.4882	1.0000
2	2.5971	0.0849(6)	2.5878	0.0870(4)	2.5979	0.0854(4)	−9.5816	0.0014(8)	101.6471	0.0047(6)
3	−0.0580	1.0000	−0.0575	1.0000	−0.0581	1.0000	0.4386	1.0000	0.2202	1.0000
4	0.0761	1.0000	0.0768	1.0000	0.0774	1.0000	−0.0615	1.0000	0.0187	1.0000
5	−1.1194	0.7976(2)	−1.1251	0.7895(2)	−1.1221	0.7955(2)	−2.7022	0.1369(2)	11.2536	0.9497(2)
6	−1.4701	0.4626(2)	−1.4673	0.4646(2)	−1.4655	0.4653(2)	−3.2414	0.0952(2)	13.9944	1.0000
7	−0.5940	1.0000	−0.5924	1.0000	−0.5919	1.0000	−2.8041	0.1271(2)	8.7551	1.0000
8	0.3210	1.0000	0.3181	1.0000	0.3213	1.0000	1.0955	0.8330(2)	1.5511	1.0000
9	0.7061	1.0000	0.7073	1.0000	0.7072	1.0000	−0.6343	1.0000	0.8862	1.0000
10	2.3571	0.1254(4)	2.3511	0.1279(4)	2.3617	0.1254(4)	3.8261	0.0598(4)	36.4121	0.1403(2)
11	0.3004	1.0000	0.3019	1.0000	0.3032	1.0000	−1.1866	0.7098(2)	1.5137	1.0000
12	0.2831	1.0000	0.9313	1.0000	0.8321	1.0000	1.1213	1.0000	2.2214	1.0000
79–80:	24.8140	1.0000	23.8243	1.0000	23.9001	1.0000	151.8156	0.0840(2)	204.0144	1.0000

Notes: Predicted residuals are normalized (see Table 6.3 for standard error estimates). The order of the Markov bounds appear between parentheses. TB1(−1), VWR(−1) and M1G(−1) represent $TB1_{t-1}$, VWR_{t-1} and $M1G_{t-1}$ respectively.

estimation subsample, but the form and timing of possible structural changes during the prediction subsample are unknown. An important advantage of our tests is that they can be used to test for structural stability at the end of the sample. Further, the tests are applicable even if the asymptotic distribution of the parameter estimates is unknown: only a consistency assumption is needed. We demonstrated that our tests can be conveniently summarized using the predictive analysis table and this provided a simple exploratory technique for analyzing the timing of structural change. We illustrated our tests and the predictive analysis table by examining the structural stability of some recent models for the comovements of asset prices and consumption. The empirical results reported in Section 6 mainly indicated some structural instability near or after the October 1979 policy shift. Further, 80:2 appears to be an outlying observation.

Each of the three methods we examined for calculating the probability of the prediction subsample involved a different strength of the distributional assumptions. In the most restrictive case, it is assumed that the disturbances follow a normal distribution. If the distribution is unknown, one can either try to estimate the underlying distribution using a flexible functional form and calculate the probability, or one can use moment-based inequalities which provide bounds on the probabilities of interest. There are clearly advantages and disadvantages to all three. If the normality assumption is correct, then this method should lead to sharper inference. However, if it is incorrect, then the resulting tests may be unreliable. One can avoid this problem by adopting the approach of estimating the probability distribution using a flexible functional form. However, the validity of the resulting inference still depends on the accuracy of the density approximation. Further, this approach can be complicated and computationally expensive. The simplest and most robust method is to use Markov inequalities to calculate bounds on the appropriate probabilities. This results in conservative inference. Further research is clearly needed to improve the applicability of the SNP approach, to explore how the method of calculating the probability of the prediction subsample affects the power of the tests, and also to study the finite sample properties of the procedures proposed above. This goes beyond the scope of this paper and is left to future research.

Centre de recherche et développement en économique and Départment de sciences économiques, Université de Montréal, Canada

Centre de recherche et développement en économique and Départment de sciences économiques, Université de Montréal, Canada

North Carolina State University, U.S.A.

APPENDIX

PROOF OF PROPOSITION 3.1. Let $l_0 \equiv l(k_0)$. Consider first the case where B1, B2 and B3 hold. For any $0 < \varepsilon_0 < 1$, we can find a compact subset $C(\varepsilon_0) \subseteq \mathbb{R}^q$ such that $P[X \in C(\varepsilon_0)] \geq \varepsilon_0$, e.g., a sufficiently large closed hypercube in $\mathbb{R}^q$; since $\mathbb{R}^q$ is the union of a countable collection of such hypercubes, the set $C(\varepsilon_0)$ does exist (see Rudin 1987, Theorem 1.19). Since $g_t[k_0, x, \beta]$ is a continuous function of $(x,$

β) for $\|\beta - \beta_0\|_{l_0} < \tau(\beta_0)$, where $\beta \in \mathbb{R}^{l_0}$, it is uniformly continuous on the compact set $C(\varepsilon_0) \times \{\beta \in \mathbb{R}^{l(k_0)}: \|\beta - \beta_0\|_{l_0} \leq \tau(\beta_0)\}$; see Royden (1968, p. 164). Thus, for any $\varepsilon > 0$, we can find $0 < \delta(\varepsilon, \varepsilon_0) \leq \tau(\beta_0)$ such that

$$\|\beta - \beta_0\|_{l_0} < \delta(\varepsilon, \varepsilon_0) \Rightarrow \|g_t[k_0, x, \beta] - g_t[k_0, x, \beta_0]\|_{l_0} < \varepsilon$$

for any $x \in C(\varepsilon_0)$, where $\delta(\varepsilon, \varepsilon_0)$ does not depend on x. Hence,

$$\begin{aligned} P[\|g_t[\hat{k}_n, X, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X, \beta_0]\|_m < \varepsilon] &\geq P[\hat{k}_n = k_0 \\ &\quad \text{and } \|g_t[k_0, X, \tilde{\beta}_n(k_0)] - g_t[k_0, X, \beta_0]\|_m < \varepsilon] \\ &\geq P[X \in C(\varepsilon_0), \hat{k}_n = k_0 \text{ and } \|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} < \delta(\varepsilon, \varepsilon_0)] \\ &\geq 1 - \{P[X \notin C(\varepsilon_0)] + P[\hat{k}_n \neq k_0] + P[\|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} \geq \delta(\varepsilon, \varepsilon_0)]\}. \end{aligned}$$

By construction $P[X \notin C(\varepsilon_0)] \leq 1 - \varepsilon_0$. From Assumption B1, $\lim_{n\to\infty} P[\hat{k}_n \neq k_0] = 0$ and $\lim_{n\to\infty} P[\|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} \geq \delta(\varepsilon, \varepsilon_0)] = 0$. Hence, $\lim_{n\to\infty} P[\|g_t[\hat{k}_n, X, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X, \beta_0]\|_m < \varepsilon] \geq \varepsilon_0$ for any $0 < \varepsilon_0 < 1$ and $\varepsilon > 0$. Consequently, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P[\|g_t[\hat{k}_n, X, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X, \beta_0]\|_m < \varepsilon] = 1,$$

from which (3.1) follows. Consider now the case where B4 holds instead of B3. For any $\varepsilon_1 > 0$, we can choose $0 < \delta(\varepsilon_1) < \tau(\beta_0)$ such that $\|\beta - \beta_0\|_{l_0} < \delta(\varepsilon_1)$ implies $h[d(\beta, \beta_0), \beta_0] < \varepsilon_1$. Then, for any $\varepsilon > 0$,

$$\begin{aligned} &P[\|g_t[\hat{k}_n, X, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X, \beta_0]\|_m < \varepsilon] \\ &\quad\geq P[\hat{k}_n = k_0, \|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} < \delta(\varepsilon_1) \\ &\qquad\qquad \text{and } \|g_t[k_0, X, \tilde{\beta}_n(k_0)] - g_t[k_0, X, \beta_0]\|_m < \varepsilon] \\ &\quad\geq P[\hat{k}_n = k_0, \|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} < \delta(\varepsilon_1) \text{ and } B_t(k_0, X, \beta_0)\varepsilon_1 < \varepsilon] \\ &\quad\geq 1 - \{P[\hat{k}_n \neq k_0] + P[\|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} \geq \delta(\varepsilon_1)] \\ &\qquad\qquad + P[B_t(k_0, X, \beta_0) \geq \varepsilon/\varepsilon_1]\}. \end{aligned}$$

For any $0 < \varepsilon_0 < 1$, we can choose ε_1 small enough that $P[B_t(k_0, X, \beta_0) \geq \varepsilon/\varepsilon_1] \leq \varepsilon_0$. Hence, using B1,

$$\lim_{n\to\infty} P[\|g_t[\hat{k}_n, X, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X, \beta_0]\|_m < \varepsilon] \geq 1 - \varepsilon_0$$

for all $0 < \varepsilon_0 < 1$, from which (3.1) follows. Q.E.D.

PROOF OF PROPOSITION 3.2. For any $\varepsilon > 0$ and $\delta_0 > 0$, we have

$$P\Big[\Big\|c_n(\beta_0)\sum_{t=1}^{n}\{g_t[\hat{k}_n, X_t, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X_t, \beta_0]\}\Big\|_m < \varepsilon\Big]$$

$$\geq P\Big[\hat{k}_n = k_0,\ \|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} < \delta_0 \text{ and } \Big\|c_n(\beta_0)\sum_{t=1}^{n}\{g_t[k_0, X_t, \tilde{\beta}_n(k_0)] - g_t[k_0, X_t, \beta_0]\}\Big\|_m < \varepsilon\Big]$$

$$\geq P\Big[\hat{k}_n = k_0,\ \|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} < \delta_0 \text{ and } \sup_{\beta\in B_0(\delta_0)}\Big\|c_n(\beta_0)\sum_{t=1}^{n}\{g_t[k_0, X_t, \beta] - g_t[k_0, X_t, \beta_0]\}\Big\|_m < \varepsilon\Big]$$

$$\geq 1 - \Big\{P[\hat{k}_n \neq k_0] + P[\|\tilde{\beta}_n(k_0) - \beta_0\|_{l_0} \geq \delta_0] + P\Big[\sup_{\beta\in B_0(\delta_0)}\Big\|c_n(\beta_0)\sum_{t=1}^{n}[g_t[k_0, X_t, \beta] - g_t[k_0, X_t, \beta_0]]\Big\|_m \geq \varepsilon\Big]\Big\}$$

where $B_0(\delta_0) = \{\beta \in \mathbb{R}^{l_0}: \|\beta - \beta_0\|_{l_0} < \delta_0\}$. Consider now the case where B5 holds. We have

$$\sup_{\beta\in B_0(\delta_0)}\Big\|c_n(\beta_0)\sum_{t=1}^{n}\{g_t[k_0, X_t, \beta] - g_t[k_0, X_t, \beta_0]\}\Big\|_m \leq A_n(\beta_0) + \bar{A}_n(\beta_0, \delta_0) + B_n(\beta_0, \delta_0)$$

where $A_n(\beta) = \|c_n(\beta_0)\sum_{t=1}^{n}\{g_t[k_0, X_t, \beta] - Eg_t[k_0, X_t, \beta]\}\|_m$, $\bar{A}_n(\beta_0, \delta_0) = \sup\{A_n(\beta): \beta \in B_0(\delta_0)\}$, and

$$B_n(\beta_0, \delta_0) = \sup_{\beta\in B_0(\delta_0)}\Big\|c_n(\beta_0)\sum_{t=1}^{n}\{Eg_t[k_0, X_t, \beta] - Eg_t[k_0, X_t, \beta_0]\}\Big\|_m.$$

By B5(b), for any $\varepsilon_0 > 0$, we can choose δ_0 such that $0 < \delta_0 \leq \tau(\beta_0)$ and $B_n(\beta_0, \delta_0) < \varepsilon_0$ for all n. Further, by B5(a), $\operatorname{plim}_{n\to\infty} A_n(\beta_0) = \operatorname{plim}_{n\to\infty} \bar{A}_n(\beta_0, \delta_0) = 0$. Thus

$$\operatorname*{plim}_{n\to\infty}\ \sup_{\beta\in B_0(\delta_0)}\Big\|c_n(\beta_0)\sum_{t=1}^{n}\{g_t[k_0, X_t, \beta] - g_t[k_0, X_t, \beta_0]\}\Big\|_m = 0$$

and, using B1,

$$\lim_{n\to\infty} P[\|c_n(\beta_0) \sum_{t=1}^{n} \{g_t[\hat{k}_n, X_t, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X_t, \beta_0]\}\|_m < \varepsilon] = 1$$

from which (3.3) follows. When B4 and B6 hold, instead of B5, we have

$$\|c_n(\beta_0) \sum_{t=1}^{n} \{g_t[k_0, X_t, \beta] - g_t[k_0, X_t, \beta_0]\}\|_m$$

$$\le |c_n(\beta_0)| \sum_{t=1}^{n} \|g_t[k_0, X_t, \beta] - g_t[k_0, X_t, \beta_0]\|_m$$

$$\le \{|c_n(\beta_0)| \sum_{t=1}^{n} B_t(k_0, X_t, \beta_0)\} h[d(\beta, \beta_0), \beta_0]$$

for $\|\beta - \beta_0\|_{l_0} < \tau(\beta_0)$; hence, for $0 < \delta_0 \le \tau(\beta_0)$,

$$\sup_{\beta \in B_0(\delta_0)} \|c_n(\beta_0) \sum_{t=1}^{n} \{g_t[k_0, X_t, \beta] - g_t[k_0, X_t, \beta_0]\}\|_m$$

$$\le \{|c_n(\beta_0)| \sum_{t=1}^{n} B_t(k_0, X_t, \beta_0)\} \bar{h}(\beta_0, \delta_0)$$

where $\bar{h}(\beta_0, \delta_0) = \sup_{\beta\in B_0(\delta_0)} h[d(\beta, \beta_0), \beta_0]$ and $\lim_{\delta_0 \downarrow 0} \bar{h}(\beta_0, \delta_0) = 0$. Using B6, we then have

$$\overline{\lim_{n\to\infty}}\, P\left[\sup_{\beta \in B_0(\delta_0)} \|c_n(\beta_0) \sum_{t=1}^{n} \{g_t[k_0, X_t, \beta] - g_t[k_0, X_t, \beta_0]\}\|_m \ge \varepsilon\right]$$

$$\le \overline{\lim_{n\to\infty}}\, P\left[\left\{|c_n(\beta_0)| \sum_{t=1}^{n} B_t(k_0, X_t, \beta_0)\right\} \bar{h}(\beta_0, \delta_0) \ge \varepsilon\right]$$

$$\le G[\varepsilon/\bar{h}(\beta_0, \delta_0), \beta_0].$$

Choose δ_0 such that $0 < \delta_0 \le \tau(\beta_0)$ and $G[\varepsilon/\bar{h}(\beta_0, \delta_0), \beta_0] \le \varepsilon_0$, where $0 < \varepsilon_0 < 1$ is arbitrary. Then, for all $0 < \varepsilon_0 < 1$,

$$\overline{\lim_{n\to\infty}}\, P\left[\|c_n(\beta_0) \sum_{t=1}^{n} \{g_t[\hat{k}_n, X_t, \tilde{\beta}_n(\hat{k}_n)] - g_t[k_0, X_t, \beta_0]\}\|_m < \varepsilon\right] \ge 1 - \varepsilon_0$$

from which (3.3) follows. Q.E.D.

REFERENCES

ANDREWS, D. W. K., "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica* 55 (1987), 1465–1471.

———, "Asymptotic Normality of Series Estimators for Various Nonparametric and Semiparametric Models." Discussion Paper No. 874, Cowles Foundation for Research in Economics, Yale University, 1988.

———, " Semiparametric Econometric Models: I. Estimation," discussion paper, Cowles Foundation for Research in Economics, Yale University, 1989a.

———, "Semiparametric Econometric Models: II. Estimation," discussion paper, Cowles Foundation for Research in Economics, Yale University, 1989b.

———, "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61 (1993), 821–856.

——— AND R. C. FAIR, "Inference in Econometric Models with Structural Change," *Review of Economic Studies* 55 (1988), 615–640.

BIANCHI, C. AND G. CALZOLARI, "The One-Period Forecast Errors in Nonlinear Econometric Models," *International Economic Review* 21 (1980), 201–208.

BROWN, B. W. AND R. S. MARIANO, "Residual-Based Procedures for Production and Estimation in a Nonlinear Simultaneous System," *Econometrica* 52 (1984), 321–343.

CHESHER, A. AND R. SPADY, "Asymptotic Expansions of the Information Matrix Test Statistic," *Econometrica* 59 (1991), 787–816.

CHOW, G. C., "Tests of Equality between Sets of Coefficients in Two Linear Regressions," *Econometrica* 28 (1960), 591–605.

DAVIS, T. E., "The Consumption Function as a Tool of Prediction," *Review of Economics and Statistics* 34 (1952), 270–277.

DUFOUR, J.-M., "Dummy Variables and Predictive Tests for Structural Change," *Economics Letters* 6 (1980), 241–247.

———, "Recursive Stability Analysis of Linear Regression Relationships: An Exploratory Methodology," *Journal of Econometrics* 19 (1982a), 31–76.

———, "Generalized Chow Tests for Structural Change: A Coordinate-Free Approach," *International Economic Review* 23 (1982b), 565–575.

———, E. GHYSELS, AND A. HALL, "Generalized Predictive Tests and Structural Change Analysis in Econometrics," discussion paper, Centre de recherche et développement en économique (C.R.D.E.), Université de Montréal, 1991.

EASTWOOD, B. J. AND A. R. GALLANT, "Adaptive Truncation Rules for Semi-nonparametric Estimators that Achieve Asymptotic Normality," discussion paper, Department of Statistics, North Carolina University, 1988.

EICHENBAUM, M., L. P. HANSEN, AND K. J. SINGLETON, "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *The Quarterly Journal of Economics* (1988), 51–78.

EPSTEIN L. AND S. E. ZIN, "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns II: An Empirical Analysis," *Journal of Political Economy* 99 (1991), 263–286.

ERLAT, H., "A Note on Testing Structural Change in a Single Equation Belonging to a Simultaneous System," *Economics Letters* 13 (1983), 185–189.

GALLANT, A. R., "Unbiased Determination of Production Technologies," *Journal of Econometrics* 20 (1982), 285–323.

———, *Nonlinear Statistical Models* (New York: Wiley, 1987a).

———, "Identification and Consistency in Semi-Nonparametric Regression," in T. F. Bewley, ed., *Advances in Econometrics*, Fifth World Congress of the Econometric Society, Volume 1 (Cambridge: Cambridge University Press, 1987b), 145–169.

——— AND D. NYCHKA, "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* 55 (1987), 363–390.

——— AND G. TAUCHEN, "Semi-Nonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica* 57 (1989), 1091–1120.

——— AND H. WHITE, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models* (New York: Basil Blackwell, 1988).

GHYSELS, E. AND A. HALL, "A Test for Structural Stability of Euler Conditions Parameters Estimated via the GMM Estimator," *International Economic Review* 31 (1990a), 355–364.

——— AND ———, "Are Consumption-Based Asset Pricing Models Structural?" *Journal of Econometrics* 45 (1990b), 121–139.

HANSEN, L. P. AND K. J. SINGLETON, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50 (1982), 1269–1286.

HODOSHIMA, J., "The LM and LR Tests for Structural Change of a Single Structural Equation," working paper, Faculty of Economics, Nanzan University, Nagoya, Japan, 1986.

HOFFMAN, D. AND A. PAGAN, "Post-Sample Prediction Tests for Generalized Method of Moments Estimators," *Oxford Bulletin of Economics and Statistics* 51 (1989), 333–344.

HONDA, Y., "Chow Tests in the Linear Simultaneous Equation: The Case of Deficient Observations," *Journal of Structural Change and Economic Dynamics* (forthcoming 1990).

KENNAN, J. AND G. NEUMANN, "A Monte Carlo Study of the Size of the Information Matrix Test," unpublished paper, Department of Economics, University of Iowa, 1988.

KULLBACK, S. AND H. M. ROSENBLATT, "On the Analysis of Multiple Regression in k Categories," *Biometrika* 44 (1957), 67–83.

LO, A. W. AND W. K. NEWEY, "A Large-Sample Chow Test for the Linear Simultaneous Equation," *Economic Letters* 18 (1985), 351–353.

LOÈVE, M., *Probability Theory I*, Fourth Edition (New York: Springer-Verlag, 1977).

MARIANO, R. S. AND B. W. BROWN, "Asymptotic Behavior of Predictors in a Nonlinear System," *International Economic Review* 24 (1983a), 523–536.

——— AND ——— "Prediction-Based Tests for Misspecification in Nonlinear Simultaneous Systems," in S. Karlin, T. Amemiya, and L. Goodman, eds., *Studies in Econometrics, Time Series and Multivariate Statistics* (New York: Academic Press, 1983b), 131–151.

——— AND ———, "Stochastic Predictions in Dynamic Nonlinear Econometric Systems," *Annales de l'INSEE* 59/60 (1985), 267–278.

NEWEY, W., "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica* 53 (1985), 1047–1070.

ORME, C., "The Small Sample Performance of the Information Matrix Test," *Journal of Econometrics* 46 (1990), 309–331.

PHILLIPS, P. C. B., "ERA's: A New Approach to Small Sample Theory," *Econometrica* 51 (1983), 1505–1527.

PINDYCK, R. S. AND J. J. ROTEMBERG, "Dynamic Factor Demands and the Effects of Energy Price Shocks," *American Economic Review* 73 (1983), 1066–1079.

PÖTSCHER, B. M. AND I. R. PRUCHA, "Consistency in Nonlinear Econometrics: A Generic Uniform Law of Large Numbers and Some Comments on Recent Results," Working Paper No. 86-9, Department of Economics, University of Maryland, 1986a.

——— AND ———, "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes," *Econometrica* 57 (1986b), 675–683.

STOCK, J. H. AND M. W. WATSON, "Interpreting the Evidence on Money-Income Causality," *Journal of Econometrics* 40 (1989), 161–181.

TAUCHEN, G., "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics* 30 (1985), 415–443.

WILSON, A. L., "When is the Chow Test UMP?" *The American Statistician* 32 (1978), 60–68.