The Review *of* Economics *and* Statistics

VOL. LXXXI NOVEMBER 1999 NUMBER 4

SYMPOSIUM ON FORECASTING AND EMPIRICAL METHODS IN MACROECONOMICS AND FINANCE

The articles in this symposium present a variety of contributions to empirical macroeconomics and finance, proposing and applying new methods for data analysis. In several cases, these applications yield intriguing—indeed, surprising—substantive conclusions.

Recent advances in computational capabilities and algorithms have permitted investigating increasingly sophisticated econometric models, which in turn yield increasingly subtle insights into macroeconomic and financial relationships. Hence, a common feature of most of these articles is their use of computationally intensive methods and/or very large data sets to study issues that are intractable using conventional methods.

Earlier versions of many of these articles were presented at the 1998 joint meeting of the NBER/NSF Forecasting Seminar and the NBER Working Group on Empirical Methods in Macroeconomics, held in Cambridge, Massachusetts, on July 14–17, 1998. Accordingly, many of these articles are directly or indirectly concerned with forecasting, which is intimately related to central ideas in macroeconomics and finance. All the papers considered for this symposium were subject to the standard peer-review process at the *Review of Economics and Statistics*. This symposium was edited by guest coeditors Professor Francis X. Diebold (NYU's Stern School of Business and the University of Pennsylvania) and Professor Kenneth D. West (University of Wisconsin) in cooperation with James Stock, Chair of the Board of Editors of the *Review*. The Board of Editors of the *Review* thanks Professors Diebold and West for their efforts on its behalf.

STOCHASTIC PERMANENT BREAKS

Robert F. Engle and Aaron D. Smith*

Abstract—This paper bridges the gap between processes where shocks are permanent and those with transitory shocks by formulating a process in which the long-run impact of each innovation is time-varying and stochastic. In the stochastic permanent breaks (STOPBREAK) process, frequent transitory shocks are supplemented by occasional permanent shifts. Consistency and asymptotic normality of quasi-maximum-likelihood estimates is established, and locally best hypothesis tests of the null of a random walk are developed. The model is applied to relative prices of pairs of stocks and significant test statistics result.

I. Introduction

TIME-SERIES analysis tends to draw a sharp line between processes where shocks have a permanent effect and those where they do not. The most notable example of this is the distinction between stationary AR(1) processes, where all shocks are transitory, and the random walk. As the autoregressive root approaches one, the rate at which shocks are expected to decay decreases, but they remain transitory. This paper aims to bridge the gap between transience and permanence by formulating a process in which the long-run impact of each observation is time-varying and stochastic. At one extreme, all innovations are transitory; at the other, all shocks are permanent.

The concept of varying the permanent impact of shocks is linked to the familiar topic of structural change. Whenever a shock or part of a shock has a permanent effect, we can interpret this as a specific type of structural break. Under this definition, a random walk has a break every period, but a stationary ARMA process has no breaks. Processes such as a threshold autoregression (Tong, 1983) have no breaks. The parameters change values, which causes the innovations to decay at a different rate, but nonetheless they remain transitory.

The stochastic permanent breaks (STOPBREAK) process is motivated by a class of processes that incur random structural shifts at random intervals. Analysis of a structural shift when the break point is known a priori generally

Received for publication: September 21, 1998. Revision accepted for publication December 31, 1998.

* University of California at San Diego

This research was supported by NSF Grant SE94-22575 to the first author, and a PEA Fellowship and a Sloan Fellowship to the second author. This work has benefited from useful conversations with Graham Elliott, Clive Granger, Hal White, and Jeff Wooldridge. Thanks also to the participants at the 1997 NBER/NSF Time Series Meeting, the 1998 Winter and Summer Meetings of the Econometric Society, the 1998 NBER Summer Institute Forecasting and Empirical Macroeconomics Workshop, the 1998 Summer Program on Nonlinear Time Series at the Econometrics Laboratory at the University of California, Berkeley, and the UCSD Econometrics Workshop. All remaining errors are our own.

involves standard test statistics and estimators (Chow, 1960). When the break point is unknown, the problem becomes more difficult because the break point must be estimated and, under the null of no break, this parameter is unidentified. Andrews et al. (1996), Andrews (1993), Hansen (1992), Christiano (1992), and others have studied this problem in various contexts. When considering multiple break points, the problem becomes further complicated because it requires specification of the number of breaks and inclusion of enough parameters to account for each regime. This becomes intractable when the number of break points becomes large.

From a forecasting perspective, the errors of finding too many breaks and not enough breaks are very different, resulting in either bias or imprecision. Linear moving-average smoothers often forecast this type of data relatively well, but they lack flexibility. We approach the problem from a different angle, treating the breaks endogenously by inferring their magnitude and frequency from realizations on a single random variable. This allows the model enough flexibility to react to breaks without overloading on parameters.

We apply the STOPBREAK model to relative prices of pairs of stocks, conjecturing that a pair of stock prices may move together for periods of time and jump apart occasionally. They may exhibit a type of temporary cointegration. We expect this relationship to be strongest between stocks within the same industry, as they are likely to have more common components in their stock price determinants.

The paper proceeds as follows. In the next section, we introduce the process and discuss its properties and its relation to other nonlinear time series processes. Sections III and IV treat hypothesis testing, and section V estimation issues. Empirical results follow in section VI.

II. STOPBREAK Process

In its simplest form, the STOPBREAK process is:

$$y_t = m_t + \epsilon_t, \qquad t = 0, 1, \ldots, T, \tag{1}$$

where ϵ_t is a stationary martingale difference sequence, and m_t is a time-varying conditional mean which is updated via

$$\begin{aligned} m_t &\equiv m_{t-1} + q_{t-1}\epsilon_{t-1} \\ &= m_0 + \sum_{i=1}^{t} q_{t-i}\epsilon_{t-i}, \qquad t = 1, 2, \ldots, T. \end{aligned} \tag{2}$$

We assume throughout that m_0 is fixed and known and that ϵ_0 is known. Let $\{F_t\}$ denote an increasing sequence of σ-fields adapted to y_t and let $q_t = q(\epsilon_t)$; then $m_{t-1} = E(y_t|F_{t-1})$, the conditional mean. If the function $q_t = q(\epsilon_t)$ satisfies $E(q_t\epsilon_t|F_{t-1}) = 0$, then m_t is also the multiperiod forecast of y.

In addition, we assume that the function is bounded by zero and one and that $\partial q/\partial|\epsilon|$ is non-negative and finite with probability to 1 (wp 1). This formulation is based on the assumption that the time series is less likely to mean revert after a large shock than after a small one. If $\tilde{q}_t = 1$, then the realized process at time t is a random walk.[1] If $\tilde{q}_t = 0$, the conditional mean does not change and, consequently, neither does the long-run forecast for y_t. Thus, we have a process where the permanence of a shock is determined endogenously. For example, in the stock market, investors may perceive large shocks as containing significant informational content and small shocks as mere noise. Consequently, their valuations and expectations react only to large shocks. Biological systems may fluctuate around some constant level of fitness with occasional mutations having large permanent effects. An economy may be subject to sporadic permanent supply shocks and frequent transitory demand shocks. Technology growth and crime rates are examples of other series that could potentially behave similarly.

In the above discussion, q is formulated as a function only of the current innovation; however, any factor that is part of the information set could be an argument in the function q_t. For example, for modeling stock prices, relevant variables may include macroeconomic announcements, profit announcements, interest rates, exchange rates, and so on. However, it is unlikely that one could account for all potential factors that could cause a permanent shift, and, even if it were possible, it would overload the process with parameters.

In this paper, we take an agnostic approach, assuming only that permanent shifts will largely be reflected in an innovation that is larger than the norm. A model specified in this manner may not pick up small shifts in the mean and thus will subsequently make systematic errors until the mean moves sufficiently. It will also be prone to overreacting to large transitory shocks, indicating equal and opposite breaks rather than no break. Nonetheless, we maintain that the simplicity and flexibility of the process more than offsets these negatives if the goal is to obtain conditional forecasts.

The STOPBREAK process can also be considered a type of error correction mechanism in the sense that it is reactionary rather than anticipatory. It makes no attempt to predict when a permanent innovation will occur, but merely reacts to shocks by forecasting their degree of permanence.

A. *Relation to Other Time Series Processes*

The distinguishing feature of the STOPBREAK process is that the permanent effect of shocks is time-varying and stochastic. In some periods, breaks are permanent and in other periods they are not.

DEFINITION: Consider some stochastic process $\{y_t\}$, $t = 0, 1, \ldots, T$. The *permanent effect of innovation t* is

$$\lambda_t \overset{d}{=} \lim_{k\to\infty} \frac{\partial f(y_t, k)}{\partial \epsilon_t},$$

[1] Unless otherwise indicated, the notation $\tilde{x}_t$ will indicate a realization on the random variable x_t.

where $f(y_t, k) \equiv E(y_{t+k}|F_t)$, $\epsilon_t = y_t - E(y_t|F_{t-1})$, F_t incorporates the entire past history of $\{y_t\}$ and $\stackrel{d}{=}$ signifies equality in distribution. This derivative can also be interpreted as the derivative with respect to y_t, holding all past y's fixed.

A martingale has the property that all shocks have a permanent effect; i.e., $\lambda_t = 1$ wp 1 $\forall t$. Conversely, if $\lambda_t = 0$ wp 1 $\forall t$, the process has no permanent breaks. Realizations, $\tilde{\lambda}_t$, between zero and one indicate partial permanent breaks, so that some fraction of a shock is remembered. An example is an integrated process with a negative invertible moving-average component. If $\tilde{\lambda}_t < 0$, we have negative permanent breaking where the process overcorrects for shocks, and, if $\tilde{\lambda}_t > 1$, shocks are magnified; i.e., the permanent effect is greater than the initial effect. A linear process with positively correlated first differences would have $\lambda_t > 1$ wp 1 $\forall t$. A permanent break is deemed to have occurred at time t if the realized permanent impact of that observation is nonzero; i.e., if $\tilde{\lambda}_t \neq 0$.

A k period ahead forecast of the STOPBREAK process in equation (2) is

$$E(y_{t+k}|y^t) = m_{t+1} \equiv m_t + q_t\epsilon_t,$$

since $E(m_{t+k}|y^t) = m_{t+1}$. Differentiating with respect to ϵ_t gives the permanent impact of observation t as

$$\lambda_t = q_t + \left.\frac{\partial q}{\partial \epsilon}\right|_{\epsilon_t} \epsilon_t$$

$$= q_t(1 + \eta_{q,t}),$$

where $\eta_{q,t} \equiv (\partial q/\partial\epsilon|_{\epsilon_t})(\epsilon_t/q_t)$. Since both q_t and $\eta_{q,t}$ are nonnegative wp 1, we have $\lambda_t \geq 0$ wp 1 $\forall t$. In the STOPBREAK process, the long-run impact of shocks varies over time: The limiting value of the impulse responses is stochastic.[2]

Compare this to a stationary autoregressive process in which the coefficients change values; i.e., $y_t = \rho_t y_{t-1} + \epsilon_t$, where ϵ_t is zero mean i.i.d. and ρ_t is a random variable taking the values ρ_1 and ρ_2 each with positive probability. The mechanism that determines ρ_t could, for example, be governed by a Markov chain as in Hamilton (1989) or a threshold (Tong, 1983). As long as the process is stationary in all regimes, it will have $\lambda_t = 0$ wp 1 $\forall t$. The rate at which shocks decay changes across regimes, but the time series remains fully mean-reverting. In fact, even if, say ρ_1 is unity, the permanent effect of all observations will still be zero with probability one. This arises because—assuming that each regime is realized with positive probability—there will eventually be enough periods of stationarity for the effect of a shock to disappear.

[2] At this point, there is no upper bound on λ_t. However, sufficient conditions for invertibility of the moving-average representation of the process will be shown to suggest some logical bounds.

The stochastic unit root process of Granger and Swanson (1997), where ρ_t varies stochastically around one, is also an interesting case. Consider $\rho_t = \exp(v_t)$, where v_t is a stationary Gaussian series with mean μ_v and variance σ_v^2. If v_t contains some positive temporal correlation and $E(\exp(v_t)) = 1$, then the effect of past shocks is magnified, and λ_t is infinite wp 1. If v_t is such that $E(\exp(\Sigma_{k=1}^{\infty} v_{t+k})) < 1$, all shocks eventually die away; i.e., $\lambda_t = 0$ wp 1 $\forall t$. Between these two extremes is a knife edge where the process exhibits stochastic permanent breaks. However, the permanent impact of shocks tends to fluctuate around one in this case, so the process does not bridge the gap between permanence and transience as STOPBREAK does.

Consider the following process:

$$y_t = \mu_t + \epsilon_t, \qquad t = 1, 2, \ldots, T, \tag{3}$$

where $\mu_t = \mu_{t-1} + u_t$, prob $(u_t = 0) = (1 - p_u)$, prob $(u_t \sim N(0, \sigma_u^2)) = p_u$ and $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. We can rewrite equation (3) as

$$\Delta y_t = u_t + \epsilon_t - \epsilon_{t-1}.$$

It is obvious that the shocks u_t have a permanent effect on y_t and that the shocks ϵ_t have a transitory effect. However, it is difficult to formally define the permanent effect of an innovation as above, because it is not clear what form $E(y_{t+k}|y^t)$ takes for this model. The random variable μ_t is not measurable-F_{t-1} and is not the conditional mean.

The best linear representation for the process in equation (3) is an integrated moving-average or, by another name, the exponential smoother. The exponential smoother has constant partial permanent breaks wp 1 in each period. A fixed proportion of each shock remains permanent, where this proportion is determined by the probability of a permanent shock, p_u, and by the relative variances of the two innovation terms. We show in section VI that a STOPBREAK model is able to forecast the process in equation (3) significantly better than the exponential smoother.

B. *Properties of the STOPBREAK Process*

We can rewrite the basic STOPBREAK in the equation (1) and (2) process as

$$\Delta y_t = \epsilon_t - \theta_{t-1}\epsilon_{t-1}, \qquad t = 1, 2, \ldots, T, \tag{4}$$

where $\theta_{t-1} = 1 - q_{t-1}$. To forecast from this process, we require an estimate of ϵ_{t-1}, which ultimately requires a forecast of ϵ_0 which is not observed. That is, we require the process to be invertible, which means that the estimate of ϵ_{t-1} depends on ϵ_0 with a coefficient that vanishes with probability one for large t.

THEOREM 1: *The nonlinear moving-average process in equation (4) is invertible* wp 1 *if* $E(|1 - q_t(1 + \eta_{qt})|$

$F_{t-1}) \leq c_t < 1$, *where* $\eta_{qt} = (\epsilon_t/q_t)(\partial q/\partial \epsilon)|_{\epsilon_t}$ and $\{c_t\}$ *is a deterministic sequence defined such that* $\lim_{T\to\infty} \Pi_{t=1}^{T} c_t = 0$.

Proof: See appendix.

Invoking the conditions of theorem 1, we have $E(\lambda_t|F_{t-1}) < 2$, although intuition suggests that the majority of the probability mass for λ_t would lie in the [0, 1] interval. Values greater than one may arise from second-order effects through the function q_t. An increase in ϵ_t raises the long-run forecast of y_{t+k} firstly by proportion q_t and secondly by raising the value of q_t. Thus, depending on the properties of the function q_t, the final effect of a change in ϵ_t may exceed the value of the change; i.e., the permanent effect of shocks may be greater than one.

The simplest example of a functional form for q_t that satisfies the conditions in theorem 1 is a threshold function, where q_t takes the value one for values of ϵ_t greater in absolute value than some threshold, and zero otherwise. Although the elasticity of q_t with respect to ϵ_t is infinite at the threshold, we can still claim invertibility wp 1 since this event occurs on a set of measure zero.

We choose to specify q_t as a continuous function. This allows for partial permanent breaks in the process, so that shocks in the gray area between large and small have only some proportion that is permanent. Intuition suggests that this less rigid specification may better represent many empirical data.

Suppose that a correctly specified model for q_t is

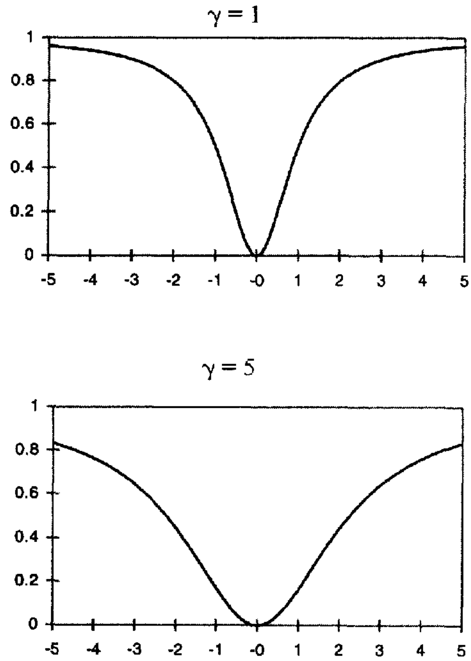
$$q_t(\gamma) = \frac{\epsilon_t^2}{\gamma + \epsilon_t^2}, \qquad \gamma > 0. \tag{5}$$

Under this specification, there is a smooth transition between the two extremes as shown in figure 1. A more rapid transition is obtained by increasing the exponent on ϵ_t to four. It turns out that most smooth functions of the shape in model (5) will satisfy the conditions of theorem 1. To violate these conditions, one needs a function with a steep section that is realized with a high-enough probability to drive $E(\lambda_t|F_{t-1})$ above two.

Along with its simplicity, an advantage of this specification is that it collapses to one as γ goes to zero. This enables parametric testing of the random-walk null hypothesis. A number of other specifications, such as the logistic, can only collapse in this fashion if an extra parameter is added. Then, under the null, there is an unidentified parameter, which introduces further complications to the testing problem.

Another useful property of the specification in (5) is that, for all $\gamma > 0$, $q_t = 0$ if and only if $\epsilon_t = 0$; i.e., no nonzero shocks are completely transitory. Thus, in periods of small errors, the process is only approximately stationary. This approximation proves beneficial for hypothesis testing because it yields test statistics with the standard distributions. We return to this topic in section III.

FIGURE 1.—$q_t(\gamma)$ PLOTTED AGAINST ϵ_t



C. *Generalizations*

The STOPBREAK process is a special case of the following general breaking process:

$$\begin{aligned} A(L)B(L)(y_t - x_t'\delta) &= A(L)z_t\epsilon_t + B(L)(1 - z_t)\epsilon_t, \\ t &= 1, 2, \ldots, T, \\ A(L) &= 1 - \alpha_1 L \\ &\quad - \alpha_2 L^2 - \cdots - \alpha_p L^p, \\ B(L) &= 1 - \beta_1 L \\ &\quad - \beta_2 L^2 - \cdots - \beta_s L^s, \end{aligned} \tag{6}$$

where x_t denotes a vector of explanatory variables, ϵ_t an innovation term, z_{t-1} some measurable function of information up to $t - 1$, and L the lag operator.

When $z_{t-1} = 0$ wp 1 $\forall t$, $B(L)$ is a common autoregressive and moving-average factor that cancels out to leave $A(L)y_t = \epsilon_t$. Similarly, the process reduces to $B(L)y_t = \epsilon_t$ when $z_{t-1} = 1$ wp 1 $\forall t$. Setting $\delta = 0$, $B(L) = 1 - L$, $A(L) = 1$, and $z_{t-1} = q_{t-1}(\gamma_0)$ obtains the basic STOPBREAK process.

TABLE 1.—POWER OF DICKEY-FULLER TEST AGAINST STOPBREAK

γ/σ^2	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
Power	0.05	0.15	0.23	0.30	0.41	0.45	0.56	0.62	0.65	0.68	0.72

Monte Carlo experiment conducted at nominal size of 5% using 5,000 repetitions on samples of 1,000 observations.

The general process in equation (6) will exhibit a changing permanent effect of shocks only if one of the lag polynomials, say $B(L)$, contains a unit root and if the other one has all roots outside the unit circle. This causes the effect of innovations to range between permanent and transitory. For example, consider $\delta = 0$, $B(L) = 1 - L$, $A(L) = 1$ and suppose that $z_{t-1} = 1$ wp 1 if $t = t^*$ and $z_{t-1} = 0$ wp 1 if $t \neq t^*$, where $1 < t^* < T$. This process has a break in its mean at $t = t^*$; i.e., $\lambda_t = 1$ wp 1 at $t = t^*$, and $\lambda_t = 0$ wp 1 in all other periods.

The general formulation in equation (6) also reveals a number of possible generalizations to the simple STOPBREAK process. For example, the process could have some temporal correlation when in "non-breaking" periods. This corresponds to $\delta = 0$, $B(L) = 1 - L$, $A(L) = 1 - \alpha_0 L$, and $z_{t-1} = q_{t-1}(\gamma_0)$, and implies the moving-average representation:

$$\Delta y_t = \alpha_0 \Delta y_{t-1} + \epsilon_t - \theta_{t-1}\epsilon_{t-1}, \qquad t = 1, 2, \ldots, T, \tag{7}$$

where $\theta_{t-1} = 1 - (1 - \alpha_0)q_{t-1}$ and $0 \leq \alpha_0 < 1$. Now y_t has an AR(1) and a random walk as its two extremes.

Including explanatory variables implies a type of temporary cointegration because it implies that a linear combination of variables follows a STOPBREAK process. This linear combination is approximately stationary for periods of time before moving and then remaining nearly stationary at a new level for a period of time. The cointegrating coefficients do not change. This parallels the practice of intercept correction, which is often used in forecasting. (See Hendry and Clements (1996).) The intercept is allowed to shift to correct for mean shifts while keeping the fundamental relationship between the variables constant.

There are other special cases of process (6) where λ_t is equal to a constant wp 1. For example, if $\delta = 0$, $A(L) = 1 - \alpha L$, $B(L) = 1 - \beta L$, and if $z_{t-1} = 0$ wp 1 $\forall t < t^*$ and $z_{t-1} = 1$ wp 1 $\forall t \geq t^*$, we have an AR(1) process where the autoregressive parameter transforms from α to β at $t = t^*$. This formulation is slightly different from a conventional parameter shift, since the break in this case is not clean. Unless $\tilde{y}_t$ is equal to its unconditional mean at the change point, the influence of that point decays away exponentially.

III. Hypothesis Testing

In the long run, a STOPBREAK random variable has no tendency to return to any previous point. With probability one, there will be a period with a nonzero permanent break, after which point the past has no predictability. Thus, the process is not covariance-stationary, and its spectral density at frequency zero is infinite. The long-run properties of the series are like those of a random walk.

As shown in equation (4), the STOPBREAK process can be written as a unit root process with a specific type of serial correlation in differences. Thus, when data are generated from a STOPBREAK process, Dickey-Fuller type tests do not reject the unit root null with probability one in large samples. However, there is some power against STOPBREAK because the downward bias in the estimate of the autoregressive root is increasing in γ_0. Alternatively, one could think of this as size distortion caused by inadequate augmentation of the Dickey-Fuller statistic. Table 1 lists the simulated power of a simple Dickey-Fuller test against γ_0/σ_0^2 for a sample size of 1,000 using the STOPBREAK specification in equation (7).

We formulate a test that is designed to distinguish between a random walk and a STOPBREAK process. Essentially, this test will search for plateaus in y_t or, more specifically, periods where the permanent effect of shocks is low. Using the parameterization in equation (5), we can write a model for the process in (1) and (2) as

$$\Delta y_t = \frac{-\gamma\epsilon_{t-1}}{\gamma + \epsilon_{t-1}^2} + \epsilon_t.$$

From this, we see that the random-walk null can be formulated parametrically as a test of H_0: $\gamma = 0$. Consider testing against the point alternative $\gamma = \bar{\gamma}$. From the Neyman-Pearson lemma, the most powerful test rejects for large values of the likelihood ratio. Given theorem 1, we can form the Gaussian conditional log likelihood functions under the null and under the alternative, and write their difference as

$$\begin{aligned} L_1 - L_0 &= -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\left(\Delta y_t + \frac{\bar{\gamma}\epsilon_{t-1}}{\bar{\gamma} + \epsilon_{t-1}^2}\right)^2 - (\Delta y_t)^2\right) \\ &= -\frac{\bar{\gamma}}{\sigma^2}\sum_{t=1}^{T}\left(\Delta y_t \frac{\epsilon_{t-1}}{\bar{\gamma} + \epsilon_{t-1}^2}\right) \\ &\quad - \frac{\bar{\gamma}^2}{2\sigma^2}\sum_{t=1}^{T}\left(\frac{\epsilon_{t-1}}{\bar{\gamma} + \epsilon_{t-1}^2}\right)^2. \end{aligned} \tag{8}$$

Thus, the most powerful test of the random-walk null is a linear combination of two statistics, each of which depend on the values of the parameters under the alternative. It follows that there is no uniformly most powerful test of a random walk against a STOPBREAK model. Consequently, we consider $\bar{\gamma}$ local to zero; i.e., $\bar{\gamma} = \bar{c}/\sqrt{T}$.

THEOREM 2: *Suppose that ϵ_t is a strictly stationary random variable on a complete probability space $(\Omega, \mathcal{F}, P)$ and is α-mixing such that $m^v\alpha(m) \to 0$ for some $v > 32$. Then, for any $c > 0$,*

$$\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\epsilon_t^2}{(c/\sqrt{T} + \epsilon_t^2)^2}\right) - E\left(\frac{\epsilon_t^2}{(c/\sqrt{T} + \epsilon_t^2)^2}\right)\right] \xrightarrow{p} 0.$$

Proof: See appendix.

From theorem 2, the second term in equation (8) is asymptotically equivalent to a constant. We show in theorem 3 below that the first term in equation (8) has a nondegenerate asymptotic distribution, which implies that it is a sufficient statistic for the locally best test. This suggests using a t-test of H_0: $\phi = 0$ against a negative alternative in the following regression:

$$\Delta y_t = \phi \frac{\Delta y_{t-1}}{\bar{\gamma} + \Delta y_{t-1}^2} + \mu_t. \tag{9}$$

The standard distribution theory applies to this t-statistic. This arises because the form of q_{t-1} means that the process is never exactly stationary, implying that—under both the null and the alternative—Δy_t contains no unit moving-average roots. The result is formalized below under general assumptions on ϵ_t.

THEOREM 3: *Suppose that $\{y_t\}$, $t = 0, 1, \ldots, T$, is a stochastic process represented by equation (1) and (2) with $q_t(c_0) = \epsilon_t^2/(c_0/\sqrt{T} + \epsilon_t^2)$, $c_0 \geq 0$. Let $\{\epsilon_t, F_t\}$ be an α-mixing strictly stationary martingale difference sequence with $E|\epsilon_t|^{4pq} < \infty$ and $m^v\alpha(m) \to 0$ where $p, q > 1$ and*

$$v > \frac{8p^2q}{(q-1)(p-1)}.$$

Suppose also that $\epsilon_t|F_{t-1}$ is distributed symmetrically about zero. Then,

$$t_{\bar{\gamma}} - \mu(T, c_0, \bar{c}) \xrightarrow{d} N(0, 1)$$

where

$$t_{\bar{\gamma}} = \frac{\sum_{t=1}^{T} \Delta y_t \frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}}{\left(\sum_{t=1}^{T}\left(\frac{\hat{u}_t \Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right)^2\right)^{1/2}},$$

$$\mu(T, c_0, \bar{c}) = -\frac{c_0\omega_{T\gamma\bar{\gamma}}}{\omega_{T\sigma\bar{\gamma}}},$$

$$\omega_{T\gamma\bar{\gamma}} = E\left(\frac{\epsilon_t^2}{(c_0/\sqrt{T} + \epsilon_t^2)(\bar{c}/\sqrt{T} + \epsilon_t^2)}\right),$$

$$\omega_{T\sigma\bar{\gamma}}^2 = E\left(\frac{\epsilon_t^2\epsilon_{t-1}^2}{(\bar{c}/\sqrt{T} + \epsilon_{t-1}^2)^2}\right),$$

$\bar{c} > 0$ and $\hat{u}_t$ denotes the least-squares residuals from the regression in equation (9).

Proof: See appendix.

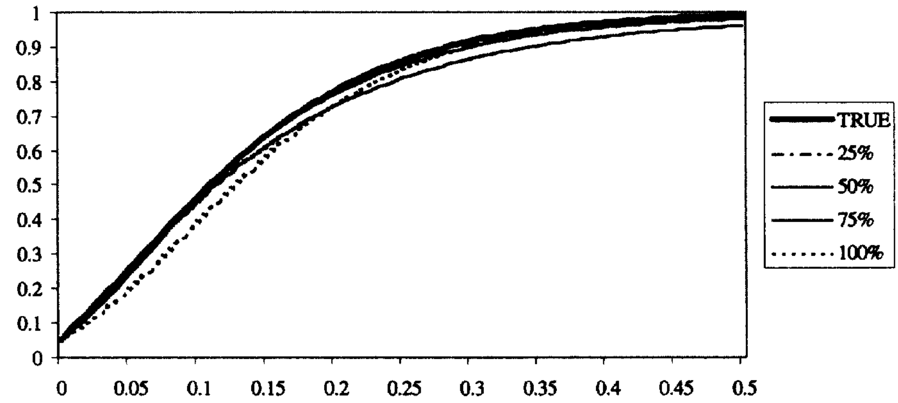
Since $\{\epsilon_t, F_t\}$ is a martingale difference sequence, the mixing assumption in theorem 3 governs dependence in the higher conditional moments of ϵ_t. The strength of these mixing conditions can be reduced by increasing the moment restrictions. For example, if all moments of ϵ_t are assumed to exist, then the size of the mixing coefficients can be as low as 32. Alternatively, should we allow the size of $\alpha(m)$ to be arbitrarily large, we only require that ϵ_t have more than four finite moments. It seems that the latter of these two would be more likely to be binding in practice. Switching from mixing of size 32 to size ∞ is unlikely to exclude many processes and mixing of an arbitrarily large size still covers m-dependent processes (those with a finite memory) and processes in which the mixing coefficients decay exponentially.[3]

From theorem 3, comparing $t_{\bar{\gamma}}$ to a standard normal is asymptotically equivalent to the locally best test of a random walk against the alternative $c = \bar{c}$. However, in practice, researchers are unlikely to know the correct value for $\bar{c}$. Further, they will not generally be interested in testing against this specific alternative. They are interested in determining whether there exists a value that provides a significantly better fit than a random walk.

Setting $c = \bar{c}$ and assuming ϵ_t i.i.d. normal, we use theorem 3 to compute the envelope of maximum power. Then, by comparing the power of the test against c_0 for a given $\bar{c}$, we search for values of $\bar{c}$ that yield power close to the envelope. From figure 2, we see that choosing $\bar{c}$ such that power is optimized somewhere between 50% and 75% causes little loss in overall power. This translates to a choice of $\mu(T, \bar{c}, \bar{c})$ between -1.5 and -2.5. Given the Gaussianity

[3] The main problem with the concept of mixing is that it is difficult to verify theoretically and cannot be verified empirically. With considerable extra work, one can replace the assumption that ϵ_t is mixing in theorem two and three with the assumption that it is near-epoch dependent (NED) on a mixing process (Davidson, 1994, Ch. 17). For example, if ϵ_t is a GARCH(1, 1) process with more than four finite moments and i.i.d. innovations, it can be shown that ϵ_t is NED on its innovation sequence. The results in theorem two and three still hold in this case, even though it is unclear whether or not ϵ_t is mixing. Assuming that ϵ_t is NED also adds significantly to the proof of corollary (4) (below) as we would then be dealing with a process that is near-epoch dependent on an NED process.

FIGURE 2.—ASYMPTOTIC POWER CURVES



Note: Curves plot asymptotic power against γ^* for $T = 200$. The "true" curve is optimal everywhere, and the other curves are derived for fixed $\bar{c}$ such that power is optimized at the indicated level.

assumption, we can integrate to find $\omega_{T\sigma\bar{\gamma}}$, yielding

$$\mu(T, \bar{c}, \bar{c}) = -\frac{\bar{c}\omega_{T\sigma\bar{\gamma}}}{\sigma_0}$$

$$= -y^* \sqrt{T}\left(\frac{1 + y^*}{\sqrt{y^*}} \sqrt{\pi/2} \exp(y^*/2) \times (1 - \Phi(\sqrt{\gamma^*})) - \frac{1}{2}\right)^{1/2}$$

where $\gamma^* = \bar{c}/\sqrt{\sigma_0 T}$ and $\Phi(z)$ is the standard normal CDF. Table 2 lists recommended choices of γ^* for various sample sizes.

Since σ_0^2 is generally unknown, we recommend approximating it with $T^{-1}\Sigma_{t=1}^{T}(\Delta y_t)^2$. This use of the data in determining γ^* will affect the distribution of the statistic, but we speculate that the effect is small when compared to the benefit of choosing the right order of magnitude for γ^*.

When $\bar{c}$ is set to zero, the $t_{\bar{\gamma}}$ statistic is a variant of the Lagrange multiplier (LM) statistic for a test of H_0: $\gamma_0 = 0$. From theorem 3, we see that $\omega_{T\sigma\bar{\gamma}}$ is infinite and $\omega_{T\gamma\bar{\gamma}}$ is finite, which implies that $\mu(T, c_0, 0) = 0$ for all c_0; therefore, the test has no power. The statistic is asymptotically distributed as standard normal under both the null and the alternative. This failure of the LM test results from a discontinuity in the function $q_t(\gamma)$ at the null, which causes the likelihood function to degenerate. In particular, the derivatives of the likelihood have no finite moments.

All of the above analysis is performed under the assumption of Gaussianity, although it is unlikely that the STOPBREAK disturbances will be Gaussian. However, Monte Carlo simulations of the asymptotic power curves under both a mixture of normals and a GARCH(1, 1) reveal a similar picture to that in figure 2. In fact, the power of the test in these cases is slightly higher due to the excess kurtosis, which helps by driving a wedge between small and large shocks. The size remains correct.

IV. Testing in a More General Context

In many cases, empirical data will exhibit some temporal correlation in all periods. In the remainder of the paper, we analyze a more general process of the form in equation (7). This introduces a complexity to the hypothesis-testing problem, as α_0 is unidentified under the random-walk null, yet knowledge of it is required to compute the test statistic. Thus, we must choose a value for α_0, despite the fact that it does not exist when the null is true. Suppose we choose the value $\bar{\alpha}$.

We compute the Neyman-Pearson test statistic as the t-statistic on ϕ in the following regression:

$$\Delta y_t = \varphi \sum_{i=1}^{t} \bar{\alpha}^{-i-1} \frac{\Delta y_{t-i}}{\bar{\gamma} + \Delta y_{t-i}^2} + u_t. \tag{10}$$

As before, the test statistic has an asymptotic normal distribution. This result is given below in corollary 4.

TABLE 2.—CHOOSING LOCAL POINT ALTERNATIVE

γ^*	0.79	0.52	0.40	0.18	0.10	0.074	0.060	0.035	0.018	0.011
T	50	75	100	250	500	750	1000	2000	5000	10000

COROLLARY 4: *Suppose that* $\{y_t\}$, $t = 0, 1, \ldots, T$, *is a stochastic process represented by equation (7) under the assumptions of theorem 3. Then*

$$t_{\bar\gamma} - \mu(T, c_0, \bar c, \alpha_0, \bar\alpha) \xrightarrow{d} N(0, 1),$$

where

$$t_{\bar\gamma} = \frac{\sum_{t=1}^{T} \Delta y_t \sum_{i=1}^{t} \bar\alpha^{i-1} \frac{\Delta y_{t-i}}{\bar c/\sqrt{T} + \Delta y_{t-i}^2}}{\left(\sum_{t=1}^{T}\left(\hat u_t \sum_{i=1}^{t} \bar\alpha^{i-1} \frac{\Delta y_{t-i}}{\bar c/\sqrt{T} + \Delta y_{t-i}^2}\right)^2\right)^{1/2}},$$

$$\mu(T, c_0, \bar c, \alpha_0, \bar\alpha) = c_0(\alpha_0 - 1)\frac{\omega_{T\bar\alpha\gamma\bar\gamma}}{\omega_{T\bar\alpha\bar\gamma}},$$

$$\omega^2_{T\bar\alpha\bar\gamma} = E\left(\epsilon_t^2 \sum_{i=1}^{t} \bar\alpha^{2(i-1)} \frac{\epsilon_{t-i}^2}{(\bar c/\sqrt{T} + \epsilon_{t-i}^2)^2}\right),$$

$$\omega_{T\bar\alpha\gamma\bar\gamma} = E\left(\left(\sum_{i=0}^{\infty} \frac{\alpha_0^i \epsilon_{t-i}}{(c_0/\sqrt{T} + \epsilon_{t-i}^2)}\right)\left(\sum_{i=0}^{\infty} \frac{\bar\alpha^i \epsilon_{t-i}}{(\bar c/\sqrt{T} + \epsilon_{t-i}^2)}\right)\right),$$

and $\hat u_t$ *denotes least squares residuals from the regression in (10).*

Proof: See appendix.

We see that, if $\alpha_0 = 1$, the asymptotic power of the test is equal to the size for all c_0. This also occurs when $\bar\alpha = 1$, because in this case $\omega^2_{T\bar\alpha\bar\gamma}$ is infinite. This lack of power arises because, when $\alpha_0 = 1$, the null of a random walk is true. As α_0 goes to one, the correlation within the flat spots increases until they are no longer flat; i.e., the rate at which shocks are expected to decay goes to zero. In contrast, when γ goes to zero, the plateaus shrink in size, keeping the correlation within them at a given level; i.e., we decrease the probability that an observation will have a low permanent impact. Thus, there are two distinct ways of parametrically representing the random-walk null.

In order to maximize power, we should choose $\bar\alpha$ to be as close as possible to α_0 as often as possible.[4] One strategy could be to choose it arbitrarily. This avoids using the data and thus distorting the distribution of the test statistic. Since the power of the test decreases as α goes to one, it would be advisable to weight this choice of $\bar\alpha$ towards one. However, if we choose it too close to one, the distribution of the test statistic will be a function of Brownian motions and no longer normal. This arises because the regressor in the test equation becomes nearly integrated as $\bar\alpha$ approaches one. A possible choice is $\bar\alpha = 0.8$.

A second strategy is to take the infimum of $t_{\bar\gamma}$ over feasible values of $\bar\alpha$ as suggested by Davies (1977) and elaborated on by Hansen (1996), Andrews (1993), and others. The null distribution of this statistic is well defined but will depend on the correlation of $t_{\bar\gamma}$ across various different values of $\bar\alpha$ that will, in general, depend on the distribution of ϵ_t. In this case, the simulated values prove reasonably robust to GARCH(1, 1) and excess kurtosis. We conduct the test using $\bar\alpha \in [0, 0.9]$.

Finally, we could approximate the test by regressing Δy_t on $\Delta y_{t-i}/(\bar\gamma + \Delta y_{t-i}^2)$, $i = 1, 2, \ldots, p$, where p is some predetermined number. Under the null hypothesis, TR^2 from this regression will be distributed as $\chi^2_{(p)}$. This procedure will lose power because of the approximation and also because it is unable to test against a one-sided alternative, but it is relatively simple to perform.

V. Quasi-Maximum-Likelihood Estimation

Given invertibility, we can specify the Gaussian conditional log likelihood function for the process in equation (7) as

$$L(y^T, \varphi) = -\frac{1}{2\sigma^2}\sum_{t=1}^{T} \times (\Delta y_t - \alpha\Delta y_{t-1} + \theta_{t-1}\epsilon_{t-1})^2 - \frac{T}{2}\log(2\pi\sigma^2), \tag{11}$$

where $\varphi = (\gamma \quad \alpha \quad \sigma)'$ and $y^T = (y_1, y_2, \ldots, y_T)$. The scores are given by

$$\frac{\partial L}{\partial \gamma} = -\frac{1}{\sigma^2}\sum_{t=1}^{T} \epsilon_t w_t, \tag{12}$$

$$\frac{\partial L}{\partial \alpha} = -\frac{1}{\sigma^2}\sum_{t=1}^{T} \epsilon_t v_t, \quad \text{and} \tag{13}$$

$$\frac{\partial L}{\partial \alpha} = \frac{1}{\sigma^3}\sum_{i=1}^{T} \epsilon_t^2 - T\sigma^{-1}, \tag{14}$$

where

$$w_t = b_{t-1}w_{t-1} - (1 - \alpha)\frac{\partial q_{t-1}}{\partial \gamma}\epsilon_{t-1}, \tag{15}$$

[4] If ϵ_t is distributed independently, the value that we choose for $\bar\alpha$ has no relation to the optimal choice of $\bar c$, since changing $\bar\alpha$ serves only to scale the power against a nonzero c_0 value proportionally. Otherwise, we speculate that changing $\bar\alpha$ will have little effect on the optimal $\bar c$.

$$v_t = b_{t-1}v_{t-1} + q_{t-1}\epsilon_{t-1} - \Delta y_{t-1}, \quad (16)$$

$$b_{t-1} = 1 - (1-\alpha)(1+\eta_{qt-1})q_{t-1}, \quad (17)$$

and, since ϵ_0 is fixed and known, $b_0 = w_0 = v_0 = 0$ wp 1.

Though the likelihood is constructed to be Gaussian, it is not necessary that this be the true distribution. In fact, the presence of large permanent shocks would indicate that leptokurtic errors are likely. We derive consistency and asymptotic normality results below under minimal assumptions on the errors.

THEOREM 5: *Suppose that* $\{y_t\}$ *is a stochastic process represented by equation (7) with* $\{\epsilon_t, F_t\}$ *a strictly stationary* α*-mixing martingale difference sequence and* $E|\epsilon_t|^{2p} < \infty$ *for some* $p > 1$. *Assume that* $q_t(\gamma)$ *is as defined in equation (5) and let* Ψ *be a compact subset of* $(0, \infty) \times [0, 1) \times (0, \infty)$. *Define* $\hat{\varphi} = \arg\max_\Psi L(y^T, \varphi)$ *and* $\varphi_0 = \arg\max_\Psi E(L(y^T, \varphi))$, *where* $\varphi \in \Psi$ *and* φ_0 *is unique. Then* $\hat{\varphi} - \varphi_0 \xrightarrow{p} 0$.

Proof: See appendix.

Given the regularity conditions of theorem 5, asymptotic normality of QMLE requires further restrictions on the moments and dependence of ϵ_t. It is worth noting that, as in theorem 3, if it is assumed that ϵ_t has marginally more than four finite moments, then the α-mixing coefficients must decay at an arbitrarily large rate. This allows some conditional heteroskedasticity, although it may not cover some well-known processes such as GARCH. In the case of GARCH, we note that the result in theorem 6 still holds if the concept of near-epoch dependence is utilized (Davidson, 1994, Ch. 17). However, since this adds considerably to the mathematics, we do not incorporate it into the result.[5]

THEOREM 6: *Suppose that the conditions of theorem 5 hold and also that the* α*-mixing coefficients on* ϵ_t *are of size* $p/(p-1)$ *and that* $E|\epsilon_t|^{4p} < \infty$ *for some* $p > 1$. *Then*

$$V_0^{-1/2}H_0T^{1/2}(\hat{\varphi} - \varphi_0) \xrightarrow{d} N(0, I),$$

where $V_0 = \lim_{T\to\infty} \text{cov}\,(T^{-1/2}\nabla_\varphi L(y^T, \varphi_0))$, $\nabla_\varphi L(y^T, \varphi)$ *represents the vector of first derivatives of* $L(y^T, \varphi)$, *and*

$$H_0 = \begin{bmatrix} H_{01} & 0 \\ 0' & H_{02} \end{bmatrix}$$

with $H_{01} = T^{-1}\sigma_0^{-2}\Sigma_{t=1}^T E(s_t s_t')$, $H_{02} = 2/\sigma_0^2$, *and* $s_t = (w_t, v_t)'$. *Assume that* H_0 *and* V_0 *are finite, nonsingular, and positive definite.*

[5] As in corollary (4), we use near-epoch dependence (NED) to prove the result in its current form. Thus, assuming that ϵ_t is also NED on a mixing process would require dealing with NED functions of NED processes in the proof. This requires significant extra work.

Proof: See appendix.

The result in theorem 6 may be of limited use in finite samples, especially if the parameters are close to their bounds. For example, the gradient of the likelihood function approaches infinity as γ goes to zero, causing confidence intervals to become condensed on the lower side. The linear approximation that yields the asymptotic normality result may thus give misleading confidence regions. A similar scenario arises when α is close to one.

Alternatively, confidence intervals can be formed by inverting a likelihood-ratio statistic as in Schoenberg (1997) and Cook and Weisburg (1990). This approach requires finding the value of the parameter such that the likelihood is significantly different from the unconstrained likelihood, where significance is determined by the usual *chi*-square distribution. Since the likelihood is explicitly used, this method better accommodates the nonlinearity and asymmetry in the model.

VI. Applications

A. *Two-Shock Process*

Consider the process in equation (3); i.e., $y_t = \mu_t + \epsilon_t$ where $\mu_t = \mu_{t-1} + u_t$, prob $(u_t = 0) = (1 - p_u)$, prob $(u_t \sim N(0, \sigma_u^2)) = p_u$, and $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. Suppose $\sigma_\epsilon^2 = 1$. Using simulated data and a variety of values for the parameters p_u and σ_u^2, we compare the mean square forecast errors for a number of potential modeling approaches.

Forming conditional forecasts for this process is difficult because it is not invertible. The parameters can be trivially estimated via method of moments, which enables unconditional inference. Shephard (1994) and others have proposed computationally intensive simulation techniques for computing the maximum-likelihood estimates. Although a STOPBREAK model is only an approximation, we contend that it is very useful for producing conditional forecasts.

We consider three models: a random walk, an exponential smoother, and STOPBREAK. Each is compared to the result from an omniscient modeler who is able to recognize both when a permanent shock has occurred and its magnitude.

The experiment is conducted over 100 trials of 6,000 observations each. The STOPBREAK model and the exponential smoother are estimated over the first 5,000 observations. One-step-ahead forecasts are then computed for the next 1,000 observations without reestimation of the model parameters. Mean square forecast errors (MSFE) are computed and compared with analytically calculated MSFEs from a random-walk forecast. The results are shown in table 3.

The far right column in Table 3 contains the difference between the average MSFEs for the exponential smoother and STOPBREAK, with associated standard errors. We see that the relative performance of the STOPBREAK improves

TABLE 3.—MODELING THE TWO-SHOCK PROCESS: AVERAGE MEAN SQUARE FORECAST ERRORS

p	σ_u^2	Omniscient	Random Walk	Exp. Smoother	STOPBREAK	(Exp. Smoother − STOPBREAK)
0.01	2	1.02	2.02	1.16	1.16	0.00
				(0.07)	(0.06)	(0.02)
0.05	2	1.10	2.10	1.38	1.38	0.00
				(0.09)	(0.08)	(0.02)
0.1	2	1.20	2.20	1.58	1.58	0.00
				(0.09)	(0.09)	(0.03)
0.01	5	1.05	2.05	1.26	1.22	0.04
				(0.10)	(0.07)	(0.04)
0.05	5	1.25	2.25	1.66	1.59	0.07
				(0.13)	(0.11)	(0.03)
0.1	5	1.50	2.50	2.03	1.95	0.08
				(0.14)	(0.13)	(0.04)
0.01	10	1.10	2.10	1.38	1.28	0.09
				(0.14)	(0.09)	(0.06)
0.05	10	1.50	2.50	2.02	1.86	0.15
				(0.19)	(0.16)	(0.05)
0.1	10	2.00	3.00	2.66	2.49	0.16
				(0.23)	(0.21)	(0.05)
0.01	20	1.20	2.20	1.57	1.39	0.18
				(0.21)	(0.14)	(0.08)
0.05	20	2.00	3.00	2.64	2.38	0.26
				(0.31)	(0.28)	(0.06)
0.1	20	3.00	4.00	3.79	3.52	0.26
				(0.40)	(0.38)	(0.06)

with the variance of the permanent shocks. When the variance of the permanent shocks is high, the exponential smoother is penalized for its slow adjustment. In other words, this scenario highlights the ability of the STOPBREAK model to react quickly to large permanent innovations. As the permanent shocks become smaller on average, the performance of the two models becomes insignificantly different. These results are largely independent of the frequency of the breaks, although, in some cases, the superiority of the STOPBREAK model is less significant at low break frequencies. This likely arises because the more permanent breaks there are, the more opportunity the STOPBREAK model has to exercise its comparative advantage.

B. *Relative Stock Prices*

Individual stock prices have a tendency to move together, by virtue of their existence in a common market. This empirical observation is supported by a number of asset-pricing models, the most well known being the CAPM, in which individual stock returns are proportional to the return on the market portfolio. However, these models tell us little about the dynamic features of relative stock prices. The goal of this analysis is to gain insight into the dynamics of relations between stock prices over time.

Many researchers have conducted cointegration tests amongst asset prices, with mixed results. Chelley-Steeley and Pentecost (1994) and Cerchi and Havenner (1988) both find evidence of cointegration among stock prices, while Stengos and Panas (1992) do not. Granger (1986) finds that the predictability inherent in cointegrating relationships precludes them from existing in an efficient market.

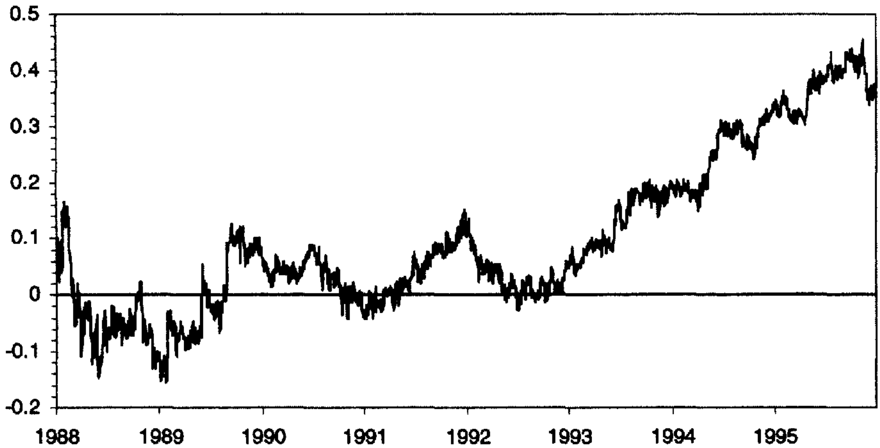
However, it may be that such relationships could exist over short periods of time in a market with imperfect information. Consider a market in which two stocks, A and B, are traded. Suppose that firms A and B are in the same industry. Various types of information will lead to a change in the value of just one stock or of both. Whenever the price of one stock changes, investors must ascertain whether this information is relevant in determining the value of the other stock. For example, if firm A announces better-than-expected profits, it may indicate higher profits for the whole industry. The price of stock B will initially rise to reflect this possibility and, then, as more information becomes available, it may go to a new "high profit" level, in which case the ratio returns to its original value. Otherwise, stock B may return to its initial point leaving the ratio permanently at a new level. Thus, we observe serial correlation in the relative price of the stocks with a time-varying permanent impact of shocks.

As an example, the logarithm of the Mobil-to-Texaco relative share price is shown in Figure 3. The presence of apparently stationary sections indicates periods where shocks have a low permanent impact punctuated by episodes of a high impact. This is most evident in the last five years of the sample where there are a number of three- to six-month periods where the first-order autocorrelation in the series is approximately 0.7.

We apply the STOPBREAK model to daily data on a number of stocks over a sample period of January, 1988, through December, 1995. The data were obtained from Datastream International.

The results of tests of the random-walk null are presented in table 4, where each column in the table contains the

FIGURE 3.—LOG SHARE PRICE RATIO—MOBILE/TEXACO



statistic computed using different methods of accommodating the unidentified parameter, as outlined in section V. We see that the null is rejected in almost all cases, with the conclusion generally invariant to the method of setting the unidentified parameter.

QMLE estimates of the model over the full sample are given in table 5 for four stock pairs. The estimates of α and the standardized γ are broadly similar across the four pairs, although confidence bands are wider for Coke/Pepsico and J&J/Merck. In both of these cases, there is a high positive correlation between $\hat{\alpha}$ and $\hat{\gamma}$. Intuitively, the model is having difficulty distinguishing between high temporal correlation with few breaks and lower correlation with more breaks.

In all but one case, the estimates of γ are within two standard deviations of zero, when heteroskedasticity-consistent standard errors are used. However, 95 percent-confidence intervals computed through inverting the appropriate likelihood-ratio statistic reveal a much shorter interval on the lower end.

TABLE 4.—TESTING THE NULL OF A SIMPLE RANDOM WALK AGAINST STOPBREAK

	$\sup_\alpha t_{\hat{\gamma}}$	Approx. ($p = 5$)	Approx. ($p = 10$)	$\alpha = 0.8$
Coke/Pepsi	−2.98**	15.08**	21.94**	−2.88**
J&J/Merck	−4.02**	18.00**	28.65**	−4.01**
GM/Ford	−3.23**	17.91**	20.78**	−3.08**
Mobil/Texaco	−5.11**	22.55**	33.20**	−5.11**
ITT/Hilton	−2.12**	16.25**	19.22**	−1.96**
AT&T/MCI	−2.24**	8.93	13.47	−2.24**
McDD/Boeing	−0.03	5.29	11.62	0.28
IBM/Microsoft	−2.60**	7.37	13.46	−2.60**
Coors/Anheuser-Busch	−3.70**	13.63**	21.48**	−3.66**
Critical Values: 10%	−1.72	9.23	15.99	−1.28
Critical Values: 5%	−2.07	11.07	18.31	−1.65

* indicates significance at 10%.
** indicates significance at 5%.
Critical values for the $\sup_\alpha t_{\hat{\gamma}}$ test were simulated using 1,000 repetitions on 1,000 observations and $\bar{\alpha} \in [0, 0.9]$.

Comparing the estimated variance of the STOPBREAK errors with the estimated variance of Δy_t reveals a low fit, with R^2 often less than 1%. The low fit is not unexpected, given the nature of financial data. However, it may possibly be improved by a richer specification of the model. For example, we could estimate some equilibrium relationship between the stock prices rather than using the ratio. Incorporating features such as conditional heteroskedasticity and more memory in q_t could also prove fruitful.

A further indicator of the presence of stochastic permanent breaks in stock price ratios is to compute potential profits from a trading strategy. The STOPBREAK model predicts the direction that the price ratio will move; that is, it forecasts whether the prices will move towards or away from each other. If they are predicted to move apart, the investor will buy the higher-valued stock and sell the lower-valued stock short. In a STOPBREAK framework, such an investor is expected to make small gains regularly and then to make either large gains or large losses when the unexpected permanent shocks occur. On average, these large profits and losses will cancel each other out, leaving an accumulated wealth with no money down.

We compute the profits gained from enacting the above pairs trading strategy for January, 1996, through June, 1997. The model is not reestimated during the forecast period, and standard errors are computed assuming that daily profits are i.i.d. The results are given in table 6.

As a benchmark, we compare the STOPBREAK profits with those from using the same strategy, but forecasting using an exponential smoother and a twenty-day moving average. For J&J/Merck and Mobil/Texaco, STOPBREAK outperforms the other models both in terms of mean return

TABLE 5.—QMLE ESTIMATION

	Coke/Pepsico	Mobil/Texaco	J&J/Merck	Coors/Anheuser
$\hat{\alpha}$	0.711	0.746	0.739	0.830
95% C.I.	(0.476, 0.966)	(0.623, 0.842)	(0.495, 0.970)	(0.681, 0.925)
Usual s.e.	(0.089)	(0.037)	(0.088)	(0.041)
Het. Cons. s.e.	(0.172)	(0.044)	(0.200)	(0.060)
$\hat{\gamma}$ ($\times 10^3$)	0.065	0.101	0.090	0.308
95% C.I.	(0.019, 0.323)	(0.044, 0.195)	(0.025, 1.210)	(0.088, 0.782)
Usual s.e.	(0.023)	(0.025)	(0.038)	(0.114)
Het. Cons. s.e.	(0.040)	(0.038)	(0.082)	(0.184)
$\hat{\gamma}/\hat{\sigma}^2$	0.384	0.766	0.517	0.617
$\hat{\sigma}^2$	0.169×10^{-3}	0.131×10^{-3}	0.175×10^{-3}	0.500×10^{-3}
$T^{-1}\Sigma_{t=1}^{T}(\Delta y_t)^2$	0.170×10^{-3}	0.133×10^{-3}	0.177×10^{-3}	0.504×10^{-3}

Note: Optimization was performed using the BFGS algorithm in GAUSS.

and the Sharpe ratio.[6] For the other two cases, the naïve moving-average model performs better. In no cases does the exponential smoother beat the STOPBREAK model.[7]

VII. Conclusion

We have proposed a new approach to modeling processes in which the effect of shocks fluctuates between permanent and transient. Typically, such data exhibits periods of apparent stationarity punctuated by occasional permanent mean shifts. Rather than attempting to individually characterize a number of different regimes, we allow the process to predict whether part or all of a shock will be permanent or transitory. The stochastic permanent breaks (STOPBREAK) process assumes that permanent shocks can be identified by their larger magnitude, but this assumption should be viewed as a special case rather than a necessity.

We considered two applications of a STOPBREAK model. First, we analyzed simulated i.i.d. data with random mean shifts of random amounts. The STOPBREAK model can be viewed as an approximation to this type of process. Out-of-sample mean square forecast errors reveal STOPBREAK to be on a par with the exponential smoother when the variance of the permanent shifts is small. As the size of the permanent shock increases, the relative performance of STOPBREAK improves, and it becomes significantly better.

The second application involved pairs of stock prices. We posit that the relative price of two stocks follows a STOPBREAK process. The two prices tend to move together for periods of time and occasionally jump apart. Hypothesis tests reveal evidence of stochastic permanent breaks, and maximum-likelihood estimates are used to form a profitable out-of-sample trading strategy.

[6] The Sharpe ratio is usually computed as the ratio of annual return to annual standard deviation. In this case, since there is no initial investment, we do not have a return but rather an accumulated wealth. Thus, the use of the term *Sharpe ratio* here is not strictly correct.

[7] It should be noted that this trading strategy requires regular transactions, as evidenced by the "average days between changes" in table 6 being approximately two in most cases. This means that the investor must reverse her position to go long in the other stock on average once every two days. For this reason, transactions costs could potentially be high.

TABLE 6.—TRADING STRATEGY PROFITS

	Coke/ Pepsico	Mobil/ Texaco	J&J/ Merck	Coors/ An-Busch
STOPBREAK:				
Average annual wealth	0.188	0.242	0.471	0.365
	(0.296)	(0.180)	(0.229)	(0.447)
'Sharpe Ratio'	0.634	1.342	2.057	0.818
Avg. days between changes	2.62	1.96	1.96	2.36
Exponential Smoother:				
Average annual wealth	0.162	0.214	0.302	0.301
	(0.296)	(0.181)	(0.230)	(0.447)
'Sharpe Ratio'	0.568	1.185	1.314	0.674
Avg. days between changes	2.16	1.94	1.99	2.05
20-day moving average:				
Average annual wealth	0.307	0.121	0.442	0.406
	(0.296)	(0.181)	(0.229)	(0.446)
'Sharpe Ratio'	1.039	0.671	1.927	0.909
Avg. days between changes	1.42	1.29	1.32	3.91

REFERENCES

Andrews, D. W. K., "Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables," *Econometric Theory* 4 (1988), 458–467.

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

Andrews, D. W. K., I. Lee, and W. Ploberger, "Optimal Changepoint Tests for Normal Linear Regression," *Journal of Econometrics* 70 (1996), 9–38.

Cerchi, M., and A. Havenner, "Cointegration and Stock Prices: The Random Walk on Wall Street Revisited," *Journal of Economic Dynamics and Control* 12 (1988), 333–346.

Chelley-Steeley, P. L., and E. J. Pentecost, "Stock Market Efficiency, the Small Firm Effect and Cointegration," *Applied Financial Economics*, 4, (1994), 405–411.

Chow, G. C. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica* 28 (1994), 591–605.

Christiano, L. J., "Searching for a Break in GDP," *Journal of Business and Economic Statistics* 10 (1992), 237–249.

Cook, R. D., and S. Weisburg, "Confidence Curves in Nonlinear Regression," *Journal of the American Statistical Association* 85 (1990), 544–551.

Davidson, J., *Stochastic Limit Theory: An Introduction for Econometricians* (New York: Oxford University Press, 1994).

Davies, R. B., "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika* 64 (1977), 247–254.

de Jong, R. M., "Laws of Large Numbers for Dependent Heterogeneous Processes," *Econometric Theory* 11 (1995), 347–358.

——— "Central Limit Theorems for Dependent Heterogeneous Random Variables," *Econometric Theory* 13 (1997), 353–367.

Granger, C. W. J., "Developments in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics* 48 (1986), 213–228.
Granger, C. W. J., and A. Andersen, "On the Invertibility of Time Series Models," *Stochastic Processes and their Applications* 8 (1978), 87–92.
Granger, C. W. J., and N. R. Swanson, "An Introduction to Stochastic Unit Root Processes," *Journal of Econometrics* 80 (1997), 35–62.
Hamilton, J. D., "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica* 57 (1989), 357–384.
Hansen, B. E., "Testing for Parameter Instability in Regressions with I(1) Process," *Journal of Business and Economic Statistics* 10 (1992), 321–335.
——— "Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis," *Econometrica* 64 (1996), 413–430.
Hendry, D. F., and M. P. Clements, "Intercept Corrections and Structural Change," *Journal of Applied Econometrics* 11 (1996), 475–494.
Schoenberg, R., "Constrained Maximum Likelihood," *Computational Economics* 10 (1997), 251–266.
Shephard, N., "Partial non-Gaussian State Space," *Biometrika* 81 (1994), 115–131.
Stengos, T., and E. Panas, "Testing the Efficiency of the Athens Stock Exchange: Some Results from the Banking Sector," *Empirical Economics* 17 (1992), 239–252.
Tong, H., "Threshold Models in Non-linear Time Series Analysis," *Lecture Notes in Statistics* 21 (Berlin: Springer, 1983).
White, H., *Asymptotic Theory for Econometricians* (New York: Academic Press, 1984).
——— *Estimation, Inference and Specification Analysis* (New York: Cambridge University Press, 1994).
Wooldridge, J. M., "Estimation and Inference for Dependent Processes," in R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics* (Volume 4, New York: North Holland, 1994).

APPENDIX

Proof of Theorem 1

Let $\{\tilde{y}_t\}$, $t = 0, 1, \ldots, T$, be a sequence of realizations from the process in equation (1) and (2) and write

$$\tilde{\epsilon}_t = (1 - \tilde{q}_{t-1})\tilde{\epsilon}_{t-1} + \Delta\tilde{y}_t \qquad (t = 1, \ldots, T), \tag{A.1}$$

where $\{\tilde{\epsilon}_t\}$ denotes the sequence of realizations from $\{\epsilon_t, \mathcal{F}_t\}$ that generated $\{\tilde{y}_t\}$.

Invertibility requires that the innovations can be computed uniquely from the observed time series. Consider the sequence of real numbers, $\hat{\epsilon}_0$, $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_T$, where

$$\hat{\epsilon}_t = (1 - \hat{q}_{t-1})\hat{\epsilon}_{t-1} + \Delta\tilde{y}_t \qquad (t = 1, 2, \ldots, T), \tag{A.2}$$

and $\hat{q}_{t-1} = q(\hat{\epsilon}_{t-1})$. Suppose that $\hat{\epsilon}_0$ is drawn from an unbounded uniform distribution.

From Granger and Andersen (1978), the process is invertible if the sequence $\{\tilde{\epsilon}_t - \hat{\epsilon}_t\}$ converges to zero. Thus, since the recursions generating $\{\tilde{\epsilon}_t\}$ and $\{\hat{\epsilon}_t\}$ are identical, it is sufficient to show that the starting value in equation (A.2) has no effect in the limit. The result here could be termed "weak invertibility," as we show only that convergence occurs wp 1.

Consider the effect on $\hat{\epsilon}_1$ of a perturbation in $\hat{\epsilon}_0$:

$$\frac{\partial\hat{\epsilon}_1}{\partial\hat{\epsilon}_0} = 1 - \left(\hat{q}_0 + \left.\frac{\partial q}{\partial\hat{\epsilon}}\right|_{\hat{\epsilon}_0}\hat{\epsilon}_0\right).$$

This leads to

$$\frac{\partial\hat{\epsilon}_T}{\partial\hat{\epsilon}_0} = \prod_{t=1}^{T}\frac{\partial\hat{\epsilon}_t}{\partial\hat{\epsilon}_{t-1}} = \prod_{t=0}^{T-1}\left(1 - \left(\hat{q}_t + \left.\frac{\partial q}{\partial\hat{\epsilon}}\right|_{\hat{\epsilon}_t}\hat{\epsilon}_t\right)\right)$$
$$= \prod_{t=0}^{T-1}(1 - \hat{q}_t(1 + \hat{\eta}_{qt})).$$

Now, utilizing the law of iterated expectations, we have

$$\begin{aligned} E\left|\frac{\partial\hat{\epsilon}_T}{\partial\hat{\epsilon}_0}\right| &= E\left(\prod_{t=0}^{T-1}|1 - \hat{q}_t(1 + \hat{\eta}_{qt})|\right) \\ &= E\left(|1 - \hat{q}_0(1 + \hat{\eta}_{q0})|\prod_{t=1}^{T-1}E(|1 - \hat{q}_t(1 + \hat{\eta}_{qt})|\,|F_{t-1})\right) \\ &\le E\left(|1 - \hat{q}_0(1 + \hat{\eta}_{q0})|\prod_{t=1}^{T-1}c_t\right) \\ &= E|1 - \hat{q}_0(1 + \hat{\eta}_{q0})|\prod_{t=1}^{T-1}c_t \\ &\to 0 \text{ as } T \to \infty. \end{aligned}$$

Since $E|\partial\hat{\epsilon}_T/\partial\hat{\epsilon}_0| \to 0$ as $T \to \infty$, we say that $\partial\hat{\epsilon}_T/\partial\hat{\epsilon}_0 \to 0$ in L_1 norm. From Davidson (1994, theorem 18.13), this implies convergence in probability; i.e.,

$$\lim_{T\to\infty}\left(\text{prob}\left(\left|\frac{\partial\hat{\epsilon}_T}{\partial\hat{\epsilon}_0}\right| < \delta\right)\right) = 1$$

for every $\delta > 0$, and the process is invertible wp 1. Q.E.D.

The following lemma will be useful in proving theorems 2 and 3 and corollary 4.

Lemma A1

Let X_t be some strictly stationary, nonnegatively valued random variable defined on a complete probability space. Consider the function

$$f_{Tt} = \frac{X_t}{c/\sqrt{T} + X_t^2}$$

where $0 < c < \infty$. Then, for all $p > 0$, $E(f_{Tt}^p) = O(T^{\xi p})$ for any $\xi > p/4(1 + p)$.

Proof of Lemma A1:

For $T < \infty$, we have

$$f_{Tt} = \frac{X_t}{c/\sqrt{T} + X_t^2} \le \frac{T^{1/4}}{2c^{1/2}}.$$

For some $\delta < 1$, we split $E(f_{Tt})$ into two components: one for f_{Tt} less than $T^{\delta/4}/2c^{1/2}$ and the other for f_{Tt} between $T^{\delta/4}/2c^{1/2}$ and the upper bound $T^{1/4}/2c^{1/2}$. We then choose δ to be the smallest value such that the expectation of the latter component shrinks to zero. It follows that the moments grow at the required rate.

Let $I(Y > Z)$ be a function taking the value one if $Y > Z$ and zero otherwise. Now consider

$$\begin{aligned} E(f_{Tt}^p) &= E(f_{Tt}^p I(f_{Tt} \le T^{\delta/4}/2c^{1/2})) + E(f_{Tt}^p I(f_{Tt} > T^{\delta/4}/2c^{1/2})) \\ &\le \left(\frac{T^{\delta/4}}{2c^{1/2}}\right)^p + (E(f_{Tt}^{pq/(q-1)}))^{1-1/q}(E(I(f_{Tt} > T^{\delta/4}/2c^{1/2})))^{1/q} \end{aligned}$$

where $q > 1$ (from Holder's inequality). Using the boundedness of f_{Tt}^p and the binary nature of the function $I(Y > Z)$, we have

$$\begin{aligned} E(f_{Tt}^p) &\le \left(\frac{T^{\delta/4}}{2c^{1/2}}\right)^p + \left(\frac{T^{1/4}}{2c^{1/2}}\right)^p(\text{Prob}\,(f_{Tt} > T^{\delta/4}/2c^{1/2}))^{1/q} \\ &= \left(\frac{T^{\delta/4}}{2c^{1/2}}\right)^p + \left(\frac{T^{1/4}}{2c^{1/2}}\right)^p(\text{Prob}\,(X_t \in [A_T, B_T]))^{1/q}, \end{aligned}$$

where

$$A_T = (c^{1/2}/T^{\delta/4})(1 - (1 - T^{(\delta-1)/2})^{1/2}) = O(T^{-\delta/4}), \quad \text{and}$$
$$B_T = (c^{1/2}/T^{\delta/4})(1 + (1 - T^{(\delta-1/2)})^{1/2}) = O(T^{-\delta/4}).$$

Now $B_T - A_T = O(T^{-\delta/4})$; so, given that Prob $(X_t \in [A_T, B_T]) \leq 1$, we have Prob $(X_t \in [A_T, B_T]) = O(T^{-\delta/4})$. Thus,

$$E(f^p_{Tt}) \leq O(T^{\delta p/4}) + O(T^{p/4-\delta/4q}) = O(T^{\xi p}),$$

where

$$\xi = \min_{0<\delta<1} \max\left(\delta/4, 1/4 - \frac{\delta}{4pq}\right) = \frac{pq}{4(1+pq)} > \frac{p}{4(1+p)}. \qquad \text{Q.E.D.}$$

Proof of Theorem 2:

Define

$$w_{Tt} = \frac{\epsilon_t^2}{(c/\sqrt{T} + \epsilon_t^2)^2} \qquad (t = 0, 1, \ldots, T),$$

where $0 < c < \infty$. We use arguments similar to those in Andrews (1988) and de Jong (1995) to show that

$$\left\| T^{-1}\sum_{t=1}^{T} w_{Tt} - E(w_{Tt})\right\|_2 \to 0,$$

where $\|x\|_p \equiv (E(x^p))^{1/p}$. For all $m \geq 1$, we can write

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T} w_{Tt} - E(w_{Tt}) &= T^{-1}\sum_{t=1}^{T} (w_{Tt} - E(w_{Tt})) - E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt}) \\
&\quad + T^{-1}\sum_{t=1}^{T} E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt}) \\
&\quad - E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt-m}) \\
&\quad + T^{-1}\sum_{t=1}^{T} E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt-m}) \\
&= T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{m-1} E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt-j}) \\
&\quad - E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt-j-1}) \\
&\quad + T^{-1}\sum_{t=1}^{T} E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt-m}) \\
&\equiv W_T + V_T.
\end{aligned}$$

Define $Y_{Ttj} \equiv E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt-j}) - E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt-j-1}) = E(w_{Tt}|\mathcal{F}_{Tt-j}) - E(w_{Tt}|\mathcal{F}_{Tt-j-1})$ and note that $\{Y_{Ttj}, \mathcal{F}_{Tt-j}\}$ is a martingale difference sequence for all j. We have

$$\begin{aligned}
\|W_T\|_2 &= \left[E\left(T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{m-1} Y_{Ttj}\right)^2\right]^{1/2} \\
&\leq \sum_{j=0}^{m-1}\left[E\left(T^{-1}\sum_{t=1}^{T} Y_{tj}\right)^2\right]^{1/2} \\
&= \sum_{j=0}^{m-1}\left[T^{-2}\sum_{t=1}^{T} E(Y^2_{Ttj})\right]^{1/2} \\
&= \sum_{j=0}^{m-1} [T^{-1}E(Y^2_{Ttj})]^{1/2} \\
&\leq T^{-1/2}\sum_{j=0}^{m-1} \|E(w_{Tt}|\mathcal{F}_{Tt-j})\|_2 + \|E(w_{Tt}|\mathcal{F}_{Tt-j-1})\|_2
\end{aligned}$$

from Minkowski's inequality. Now, since w_{Tt} is an $\mathcal{F}_{Tt}$-measurable function of finite length, we have from White (1984, theorem 3.49) that $\{w_{Tt}, \mathcal{F}_{Tt}\}$ is mixing of the same size as $\{\epsilon_t, \mathcal{F}_t\}$. Thus, from Davidson (1994, theorem 14.2),

$$\|E(w_{Tt}|\mathcal{F}_{Tt-j})\|_2 \leq 2(1 + \sqrt{2})\,\|w_{Tt}\|_{2p}\alpha(j)^{(p-1)/2p}$$

for some $p > 1$, where $\alpha(j)$ denotes the strong mixing coefficient for ϵ_t. It follows that

$$\begin{aligned}
\|W_T\|_2 &\leq T^{-1/2}2(1 + \sqrt{2})\,\|w_{Tt}\|_{2p}\sum_{j=0}^{m-1} (\alpha(j)^{(p-1)/2p} + \alpha(j-1)^{(p-1)/2p}) \\
&\leq T^{-1/2}4(1 + \sqrt{2})\,\|w_{Tt}\|_{2p}m.
\end{aligned}$$

Now, from lemma (A1), we have $\|w_{Tt}\|_{2p} = O(T^{\eta})$ where $\eta > 2p/(1 + 4p)$. Thus, we let m increase with T by defining

$$m \equiv m_T = \left[\frac{\delta T^{-1/2-\eta}}{4(1+\sqrt{2})}\right],$$

where $\delta > 0$ and $[x]$ denotes the largest integer less than x. This implies

$$\lim_{T\to\infty} \|W_T\|_2 \leq \delta.$$

Using theorem 14.2 of Davidson (1994) on V_T yields

$$\begin{aligned}
\|V_T\|_2 &\leq T^{-1}\sum_{t=1}^{T} \|E(w_{Tt} - E(w_{Tt})|\mathcal{F}_{Tt-m_T})\|_2 \\
&\leq 2(1 + \sqrt{2})\,\|w_{Tt} - E(w_{Tt})\|_{2p}\alpha(m_T)^{(p-1)/2p} \\
&\leq 4(1 + \sqrt{2})\,\|w_{Tt}\|_{2p}\alpha(m_T)^{(p-1)/2p}
\end{aligned}$$

Now, since $\|w_{Tt}\|_{2p} = O(T^{\eta})$, we can use the definition of m_T to write $\|w_{Tt}\|_{2p} = O(m_T^{\lambda})$ where

$$\lambda = \frac{\eta}{1/2 - \eta} > \frac{2p/(1+4p)}{1/2 - 2p/(1+4p)} = 4p.$$

It follows that, for some finite constant K

$$\begin{aligned}
\|V_T\|_2 &\leq K4(1 + \sqrt{2})m_T^{\lambda}\alpha(m_T)^{(p-1)/2p} \\
&= K4(1 + \sqrt{2})(m_T^{2\lambda p/(p-1)}\alpha(m_t))^{(p-1)/2p}.
\end{aligned}$$

Note that $2\lambda p/(p-1) > 8p^2/(p-1) \geq 32$ is minimized at $p = 2$. Then, since we assumed that $\alpha(m_T)$ is of size 32, we have $\|V_T\|_2 \to 0$. Thus $\|T^{-1}\Sigma_{t=1}^{T}w_{Tt} - E(w_{Tt})\|_2 \to 0$ and, therefore, from White (1984, theorem 2.42), $T^{-1}\Sigma_{t=1}^{T}w_{Tt} - E(w_{Tt}) \xrightarrow{p} 0$. Q.E.D.

Proof of Theorem 3:

We have

$$t_{\bar{\gamma}} = \frac{\dfrac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\displaystyle\sum_{t=1}^{T}\Delta y_t\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y^2_{t-1}}}{\left(\dfrac{1}{\omega^2_{T\sigma\bar{\gamma}}T}\displaystyle\sum_{t=1}^{T}\left(\frac{\hat{a}\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y^2_{t-1}}\right)^2\right)^{1/2}}$$

where

$$\omega_{T\sigma\bar{\gamma}} = \left(E\left(\frac{\epsilon_t^2\epsilon^2_{t-1}}{(\bar{c}/\sqrt{T} + \epsilon^2_{t-1})^2}\right)\right)^{1/2}.$$

Note that

$$\Delta y_{t-1} = \epsilon_{t-1} - \frac{c_0/\sqrt{T}}{c_0/\sqrt{T} + \epsilon_{t-2}^2}\epsilon_{t-2}$$

and consider the numerator of $t_{\bar{\gamma}}$:

$$\begin{aligned}
&\frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\Delta y_t \frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2} \\
&= \frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\frac{\left(\epsilon_t - \frac{c_0/\sqrt{T}}{c_0/\sqrt{T}+\epsilon_{t-1}^2}\epsilon_{t-1}\right)\times\left(\epsilon_{t-1} - \frac{c_0/\sqrt{T}}{c_0/\sqrt{T}+\epsilon_{t-2}^2}\epsilon_{t-2}\right)}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2} \\
&= \frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right) \\
&\quad - \frac{c_0}{\omega_{T\sigma\bar{\gamma}}}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-2}}{(c_0/\sqrt{T}+\epsilon_{t-2}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)}\right) \\
&\quad - \frac{c_0}{\omega_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T}\left(\frac{\epsilon_{t-1}^2}{(c_0/\sqrt{T}+\epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)}\right) \\
&\quad + \frac{c_0^2}{\omega_{T\sigma\bar{\gamma}}T^{3/2}}\sum_{t=1}^{T}\left(\frac{\epsilon_{t-1}\epsilon_{t-2}}{(c_0/\sqrt{T}+\epsilon_{t-1}^2)(c_0/\sqrt{T}+\epsilon_{t-2}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)}\right)
\end{aligned}$$

We analyze each of these four terms separately, labeling them (i), (ii), (iii), and (iv), respectively.

$$\begin{aligned}
\text{(i)}\ &\frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2}\right) \\
&= \frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-1}}{\bar{c}/\sqrt{T} + \epsilon_{t-1}^2}\right) + \frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T} \times\left(\frac{\epsilon_t\epsilon_{t-1}(\epsilon_{t-1}^2 - \Delta y_{t-1}^2)}{(\bar{c}/\sqrt{T}+\epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)}\right) \\
&= \frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-1}}{\bar{c}/\sqrt{T} + \epsilon_{t-1}^2}\right) + \frac{1}{\omega_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T} \times\left(\frac{\epsilon_t\epsilon_{t-1}(2c_0\epsilon_{t-1}\epsilon_{t-2})}{(\bar{c}/\sqrt{T}+\epsilon_{t-2}^2)(\bar{c}/\sqrt{T}+\epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)}\right) \\
&\quad - \frac{1}{\omega_{T\sigma\bar{\gamma}}T^{3/2}}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-1}(c_0^2\epsilon_{t-2}^2)}{(\bar{c}/\sqrt{T}+\epsilon_{t-2}^2)^2(\bar{c}/\sqrt{T}+\epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)}\right) \\
&\equiv \frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\epsilon_t f_{Tt-1} + 2c_0\frac{1}{\omega_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T}\epsilon_t f_{Tt-1}f_{Tt-2}f_{yTt-1} \\
&\quad - c_0^2\frac{1}{\omega_{T\sigma\bar{\gamma}}T^{3/2}}\sum_{t=1}^{T}\epsilon_t f_{Tt-1}f_{Tt-2}^2 g_{yTt-1},
\end{aligned}$$

where

$$f_{Tt} = \frac{\epsilon_t}{\bar{c}/\sqrt{T}+\epsilon_t^2},\qquad f_{yTt} = \frac{\epsilon_t}{\bar{c}/\sqrt{T}+\Delta y_t^2},\quad \text{and}$$

$$g_{yTt} = \frac{1}{\bar{c}/\sqrt{T}+\Delta y_t^2}.$$

All three terms here are martingale difference arrays. For the first term, we use the central limit theorem of Davidson (1994, theorem 24.3). For the second and third terms, we proceed in a similar fashion to the proof of theorem 2. One notable difference relates to terms with Δy_t^2 rather than ϵ_t^2 in the denominator. Since

$$\begin{aligned}
|\Delta y_{t-1}| &= \left|\epsilon_{t-1} - \frac{c_0/\sqrt{T}}{c_0/\sqrt{T}+\epsilon_{t-2}^2}\epsilon_{t-2}\right|, \\
&\geq |\epsilon_{t-1}| - \frac{c_0^{1/2}}{2T^{1/4}},
\end{aligned}$$

we have

$$\left|\frac{\epsilon_{t-1}}{c_0/\sqrt{T}+\Delta y_{t-1}^2}\right| \leq \left|\frac{\epsilon_{t-1}}{c_0/\sqrt{T}+(|\epsilon_{t-1}| - c_0^{1/2}/2T^{1/4})^2}\right| \leq \frac{T^{1/4}\sqrt{5}}{6c_0^{1/2}}.$$

Thus, this term is bounded, and the bound and moments diverge at the same rate as the term in lemma (A1).

Define

$$Z_{Tt} = \frac{T^{-1/2}\epsilon_t\epsilon_{t-1}}{\omega_{T\sigma\bar{\gamma}}(\bar{c}/\sqrt{T}+\epsilon_{t-1}^2)}.$$

Since $\{\epsilon_t, \mathcal{F}_t\}$ is a stationary martingale difference sequence, $\{Z_{Tt}, \mathcal{F}_{Tt}\}$ is also a martingale difference sequence for all $\bar{c} > 0$. To use Davidson's CLT, we need to show that $\max_{1\leq t\leq T}|Z_{Tt}| \xrightarrow{p} 0$ and $\sum_{t=1}^{T} Z_{Tt}^2 \xrightarrow{p} 1$. We have

$$|Z_{Tt}| \leq \frac{T^{-1/2}\epsilon_t}{\omega_{T\sigma\bar{\gamma}}}\frac{T^{1/4}}{2\bar{c}^{1/2}} \xrightarrow{p} 0,$$

implying that the first condition is satisfied. For the second condition, we follow the proof of theorem 2 and write

$$\begin{aligned}
\sum_{t=1}^{T} Z_{Tt}^2 - 1 &= \omega_{T\sigma\bar{\gamma}}^{-2}T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{m-1}(E(\epsilon_t^2 f_{Tt-1}^2|F_{Tt-j}) \\
&\quad - E(\epsilon_t^2 f_{Tt-1}^2|F_{Tt-j-1})) + T^{-1}\sum_{t=1}^{T}E(\omega_{T\sigma\bar{\gamma}}^{-2}\epsilon_t^2 f_{Tt-1}^2 - 1|F_{Tt-m}) \\
&\equiv W_T + V_T.
\end{aligned}$$

Since the summands of W_T are martingale differences, we have

$$\|W_T\|_2 \leq \omega_{T\sigma\bar{\gamma}}^2 T^{-1/2}\sum_{j=0}^{m-1}\|E(\epsilon_t^2 f_{Tt-1}^2|F_{Tt-j})\|_2 + \|E(\epsilon_t^2 f_{Tt-1}^2|F_{Tt-j-1})\|_2.$$

Then, since $\epsilon_t^2 f_{Tt-1}^2$ is mixing of the same size as ϵ_t we can use theorem 14.2 of Davidson (1994) and then Holder's inequality to obtain

$$\begin{aligned}
\|W_T\|_2 &\leq \omega_{T\sigma\bar{\gamma}}^{-2}T^{-1/2}2(1+\sqrt{2})\|\epsilon_t^2 f_{Tt-1}^2\|_{2p}\sum_{j=0}^{m-1}(\alpha(j)^{(p-1)/2p} + \alpha(j-1)^{(p-1)/2p}) \\
&\leq \omega_{T\sigma\bar{\gamma}}^{-2}T^{-1/2}4(1+\sqrt{2})\|\epsilon_t^2\|_{2pq}\|f_{Tt-1}^2\|_{(2pq/q-1)}m
\end{aligned}$$

for some $p, q > 1$, where $\alpha(j)$ denotes the strong mixing coefficient for ϵ_t. From lemma (A1), we have $\|f_{Tt-1}^2\|_{(2pq/q-1)} = O(T^{\eta})$, where $\eta > 2pq/(q-1+4pq)$. Now, we let m increase with T by defining

$$m \equiv m_T = \left[\frac{\delta T^{1/2-\eta}}{2(1+\sqrt{2})\|\epsilon_t^2\|_{2pq}}\right],$$

where $\delta > 0$ and $[x]$ denotes the largest integer less than x. Then, since ϵ_t has $4pq$ finite moments, we have

$$\lim_{T\to\infty}\|W_T\|_2 \leq \delta\omega_{T\sigma\bar{\gamma}}^{-2} \to 0.$$

(Note at this point that the mixing conditions on ϵ_t could be relaxed if we had a nontrivial lower bound on the rate of divergence of $E(\epsilon_t^2 f_{Tt}^2) = \omega_{T\sigma\bar{\gamma}}^2$. In that case, we could allow m_T to diverge at a faster rate, which in turn would put less of a constraint on the rate that the mixing coefficients converge to zero.)

Using theorem 14.2 of Davidson (1994) on V_T yields

$$\|V_T\|_2 \leq 2(1 + \sqrt{2})\omega_{T\sigma\bar{\gamma}}^{2} \|\epsilon_t^2\|_{2pq} \|f_{Tt-1}^2\|_{(2pq/q-1)} \alpha(m_T)^{(p-1)/2p}$$

and, since $\|f_{Tt-1}^2\|_{(2pq/q-1)} = O(T^{\eta})$, we can use the definition of m_T to write $\|f_{Tt-1}^2\|_{(2pq/q-1)} = O(m_T^{\lambda})$ where

$$\lambda = \frac{\eta}{1/2 - \eta} > \frac{2pq/(q-1+4pq)}{1/2 - 2pq/(q-1+4pq)} = \frac{4pq}{q-1}.$$

It follows that, for some finite constant K,

$$\|V_T\|_2 \leq K m_T^{\lambda} \alpha(m_T)^{(p-1)/2p}$$
$$= K(m_T^{2\lambda p/(p-1)} \alpha(m_T))^{(p-1)/2p}.$$

Now $2\lambda p/(p-1) > 8p^2 q/(p-1)(q-1)$, and so, from our assumption on the size of $\alpha(m_T)$, we have $\|V_T\|_2 \to 0$. Thus, $\|\Sigma_{t=1}^T Z_{Tt}^2 - 1\|_2 \to 0$, which implies $\Sigma_{t=1}^T Z_{Tt}^2 \xrightarrow{p} 1$ from White (1984, theorem 2.42). We can now invoke the CLT in Davidson (1994, theorem 24.3) to obtain $\Sigma_{t=1}^T Z_{Tt} \xrightarrow{d} N(0, 1)$.

Since $\{\epsilon_t f_{Tt-1} f_{Tt-2} f_{yTt-1}, \mathcal{F}_{Tt}\}$ is a martingale difference array, we can proceed as for W_T above to obtain

$$\left\| \frac{2c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \epsilon_t f_{Tt-1} f_{Tt-2} f_{yTt-1} \right\|_2 \leq 2c_0 \omega_{T\sigma\bar{\gamma}}^{-1} T^{-1/2} \|\epsilon_t f_{Tt-1} f_{Tt-2} f_{yTt-1}\|_2$$

From the boundedness of f_{Tt-2} and f_{yTt-1}, $|\epsilon_t f_{Tt-1} f_{Tt-2} f_{yTt-1}| \leq KT^{\delta/2} \times |\epsilon_t f_{Tt-1}|$ with probability approaching one (wpa 1), for some $\delta < 1$, $K < \infty$. It follows that

$$\lim_{T\to\infty} 2c_0 \omega_{T\sigma\bar{\gamma}}^{-1} T^{-1/2} \|\epsilon_t f_{Tt-1} f_{Tt-2} f_{yTt-1}\|_2 \leq \lim_{T\to\infty} 2c_0 \omega_{T\sigma\bar{\gamma}}^{-1} T^{-1/2} \|\epsilon_t f_{Tt-1} K T^{\delta/2}\|_2$$
$$= \lim_{T\to\infty} 2c_0 \|\epsilon_t f_{Tt-1}\|_2^{-1} T^{-1/2} \|\epsilon_t f_{Tt-1}\|_2 K T^{\delta/2}$$
$$= \lim_{T\to\infty} 2c_0 T^{(\delta-1)/2} = 0.$$

Thus, we have

$$\left\| \frac{2c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \epsilon_t f_{Tt-1} f_{Tt-2} f_{yTt-1} \right\|_2 \to 0,$$

and, therefore,

$$\frac{2c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \epsilon_t f_{Tt-1} f_{Tt-2} f_{yTt-1} \xrightarrow{p} 0.$$

Similarly, we have

$$\lim_{T\to\infty} \left\| \frac{c_0^2}{\omega_{T\sigma\bar{\gamma}} T^{3/2}} \sum_{t=1}^{T} \epsilon_t f_{Tt-1} f_{Tt-2}^2 g_{yTt-1} \right\|_2$$
$$\leq \lim_{T\to\infty} c_0^2 \omega_{T\sigma\bar{\gamma}}^{-1} T^{-1} \|\epsilon_t f_{Tt-1} f_{Tt-2}^2 g_{yTt-1}\|_2$$
$$\leq \lim_{T\to\infty} c_0^2 \|\epsilon_t f_{Tt-1}\|_2^{-1} T^{-1} \|\epsilon_t f_{Tt-1}\|_2 K T^{\delta}$$
$$= 0,$$

and, therefore,

$$c_0^2 \frac{1}{\omega_{T\sigma\bar{\gamma}} T^{3/2}} \sum_{t=1}^{T} \epsilon_t f_{Tt-1} f_{Tt-2}^2 g_{yTt-1} \xrightarrow{p} 0.$$

Thus, for term (i),

$$\frac{1}{\omega_{T\sigma\bar{\gamma}} \sqrt{T}} \sum_{t=1}^{T} \left(\frac{\epsilon_t \epsilon_{t-1}}{\bar{c}/\sqrt{T} + \Delta y_{t-1}^2} \right) \xrightarrow{d} N(0, 1).$$

(ii) Consider

$$\frac{c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \left(\frac{\epsilon_t \epsilon_{t-2}}{(c_0/\sqrt{T} + \epsilon_{t-2}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)} \right) = \frac{c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \epsilon_t f_{Tt-2} g_{yTt-1}.$$

Since $\{\epsilon_t f_{Tt-2} g_{yTt-1}, \mathcal{F}_{Tt}\}$ is a martingale difference sequence, we proceed as above to get

$$\lim_{T\to\infty} \left\| \frac{c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \epsilon_t f_{Tt-2} g_{yTt-1} \right\|_2 \leq \lim_{T\to\infty} c_0 \omega_{T\sigma\bar{\gamma}}^{-1} T^{-1/2} \|\epsilon_t f_{Tt-2} g_{yTt-1}\|_2$$
$$\leq \lim_{T\to\infty} c_0 \|\epsilon_t f_{Tt-1}\|_2^{-1} T^{-1/2} \|\epsilon_t f_{Tt-2} K T^{\delta/2}\|_2$$
$$= 0,$$

which implies that

$$\frac{c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \epsilon_t f_{Tt-2} g_{yTt-1} \xrightarrow{p} 0.$$

$$\text{(iii)} \quad \frac{c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \left(\frac{\epsilon_{t-1}^2}{(c_0/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)} \right)$$
$$= \frac{c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \left(\frac{\epsilon_{t-1}^2}{(c_0/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \epsilon_{t-1}^2)} \right) + \frac{c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \left(\frac{\epsilon_{t-1}^2(\epsilon_{t-1}^2 - \Delta y_{t-1}^2)}{(c_0/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)} \right)$$
$$= \frac{c_0}{\omega_{T\sigma\bar{\gamma}} T} \sum_{t=1}^{T} \left(\frac{\epsilon_{t-1}^2}{(c_0/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \epsilon_{t-1}^2)} \right) + \frac{2c_0^2}{\omega_{T\sigma\bar{\gamma}} T^{3/2}} \sum_{t=1}^{T} \left(\frac{\epsilon_{t-1}^3 \epsilon_{t-2}}{(c_0/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)(c_0/\sqrt{T} + \epsilon_{t-2}^2)} \right)$$
$$- \frac{c_0^3}{\omega_{T\sigma\bar{\gamma}} T^2} \sum_{t=1}^{T} \left(\frac{\epsilon_{t-1}^2 \epsilon_{t-2}^2}{(c_0/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \epsilon_{t-1}^2)(\bar{c}/\sqrt{T} + \Delta y_{t-1}^2)(c_0/\sqrt{T} + \epsilon_{t-2}^2)^2} \right)$$
$$= \frac{c_0 \omega_{T\gamma\bar{\gamma}}}{\omega_{T\sigma\bar{\gamma}}} + \frac{2c_0^2}{\omega_{T\sigma\bar{\gamma}} T^{3/2}} \sum_{t=1}^{T} f_{Tt-1}^2 f_{yTt-1} f_{Tt-2} - \frac{c_0^3}{\omega_{T\sigma\bar{\gamma}} T^2} \sum_{t=1}^{T} f_{Tt-1}^2 f_{Tt-2}^2 g_{yTt-1} + o_p(1)$$

from theorem 2. From the assumption that $\epsilon_t|F_{t-1}$ is distributed symmetrically about zero, $\{f^2_{Tt-1}f_{yTt-1}f_{Tt-2}, F_{Tt-1}\}$ is a martingale difference sequence. Thus, proceeding as for W_T above and using Holder's inequality yields

$$\left\| T^{-3/2}\sum_{t=1}^{T} f^2_{Tt-1}f_{yTt-1}f_{Tt-2}\right\|_2 \le T^{-1}\|f^2_{Tt-1}\|_{2p}\|f_{yTt-1}\|_{4p/(p-1)}\|f_{Tt-2}\|_{4p/(p-1)} = O(T^{\eta})$$

for any $p > 1$ and any

$$\eta > \frac{2p}{1+4p} + \frac{2p}{5p-1} - 1.$$

Thus, we can choose p close enough to one such that $\eta < 0$ implying, from White (1984, theorem 2.42),

$$\frac{2c_0^2}{\omega_{T\sigma\bar{\gamma}}T^{3/2}}\sum_{t=1}^{T} f^2_{Tt-1}f_{yTt-1}f_{Tt-2} \xrightarrow{p} 0.$$

Similarly, using Holders inequality and the boundedness of g_{yTt}, we have

$$\left\| T^{-2}\sum_{t=1}^{T} f^2_{Tt-1}f^2_{Tt-2}g_{yTt-1}\right\|_p \le T^{-1}\|f^2_{Tt-1}\|_{2p}\|f^2_{Tt-2}\|_{2p}T^{1/2} = O(T^{\eta})$$

for any $p > 0$ and any

$$\eta > \frac{4p}{1+4p} - \frac{1}{2}.$$

We can choose p such that $\eta < 0$, which implies

$$\frac{c_0^3}{\omega_{T\sigma\bar{\gamma}}T^2}\sum_{t=1}^{T} f^2_{Tt-1}f^2_{Tt-2}g_{yTt-1} \xrightarrow{p} 0$$

from White (1984, theorem 2.42).

Thus,

$$\frac{c_0}{\omega_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T}\left(\frac{\epsilon^2_{t-1}}{(c_0/\sqrt{T}+\epsilon^2_{t-1})(\bar{c}/\sqrt{T}+\Delta y^2_{t-1})}\right) = \frac{c_0\omega_{T\bar{\gamma}\bar{\gamma}}}{\omega_{T\sigma\bar{\gamma}}} + o_p(1).$$

(iv) From the assumption that $\epsilon_t|\mathcal{F}_{t-1}$ is distributed symmetrically about zero, the summands of (iv) are a martingale difference sequence with respect to $\mathcal{F}_{Tt-1}$. Then, as above, we have

$$\left\| T^{-3/2}\sum_{t=1}^{T} f_{Tt-1}f_{Tt-2}g_{yTt-1}\right\|_2 \le T^{-1}\|f_{Tt-1}\|_4\|f_{Tt-2}\|_4T^{1/2} = O(T^{\eta})$$

for any $\eta > 2/5 - 1/2$. Thus, we can choose $\eta < 0$, and we have, from White (1984, theorem 2.42),

$$\frac{c_0^2}{T^{3/2}}\sum_{t=1}^{T}\left(\frac{\epsilon_{t-1}\epsilon_{t-2}}{(c_0/\sqrt{T}+\epsilon^2_{t-1})(c_0/\sqrt{T}+\epsilon^2_{t-2})(\bar{c}/\sqrt{T}+\Delta y^2_{t-1})}\right) \xrightarrow{p} 0.$$

Thus, for the numerator of $t_{\bar{\gamma}}$, we have

$$\frac{1}{\omega_{T\sigma\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\Delta y_t\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}} - \left(z - \frac{c_0\omega_{T\bar{\gamma}\bar{\gamma}}}{\omega_{T\sigma\bar{\gamma}}}\right) = o_p(1)$$

where $z \sim N(0, 1)$.

Consider the denominator of $t_{\bar{\gamma}}$. Now,

$$\hat{u}_t = \Delta y_t - \hat{\phi}\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}},$$

where

$$\hat{\phi} = \frac{\frac{1}{T}\sum_{t=1}^{T}\Delta y_t\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}}}{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}}\right)^2}.$$

Since we have shown that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_t\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}} - (\omega_{T\sigma\bar{\gamma}}z - c_0\omega_{T\bar{\gamma}\bar{\gamma}}) = o_p(1),$$

using lemma (A1), we have that the numerator of $\hat{\phi}$ converges to zero. Then, replacing ϵ_t with Δy_t in the proof of theorem 2, we obtain that the denominator is asymptotically equivalent to a nonzero constant. Thus, $\hat{\phi} \xrightarrow{p} 0$.

We have, for the denominator of $t_{\bar{\gamma}}$,

$$\begin{aligned}
&\frac{1}{\omega^2_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T}\left(\frac{\hat{u}_t\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}}\right)^2\\
&= \frac{1}{\omega^2_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T}\left(\frac{\epsilon_t\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}}\right)^2 + o_p(1)\\
&= \frac{1}{\omega^2_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}} - \frac{c_0\epsilon_t\epsilon_{t-2}}{\sqrt{T}(\bar{c}/\sqrt{T}+\Delta y^2_{t-1})(c_0/\sqrt{T}+\epsilon^2_{t-2})}\right)^2 + o_p(1)\\
&= \frac{1}{\omega^2_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}}\right)^2 + o_p(1)\\
&= \frac{1}{\omega^2_{T\sigma\bar{\gamma}}T}\sum_{t=1}^{T}\left(\frac{\epsilon_t\epsilon_{t-1}}{\bar{c}/\sqrt{T}+\epsilon^2_{t-1}}\right)^2 + o_p(1)\\
&= 1 + o_p(1)
\end{aligned}$$

from (i) above. This leaves

$$t_{\bar{\gamma}} - \left(z - \frac{c_0\omega_{T\bar{\gamma}\bar{\gamma}}}{\omega_{T\sigma\bar{\gamma}}}\right) = o_p(1),$$

and the result follows. Q.E.D.

Proof of Corollary 4:

We have

$$t_{\bar{\gamma}} = \frac{\frac{1}{\omega_{T\bar{\alpha}\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\Delta y_t\sum_{i=1}^{t}\bar{\alpha}^{i-1}\frac{\Delta y_{t-i}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-i}}}{\left(\frac{1}{\omega^2_{T\bar{\alpha}\bar{\gamma}}T}\sum_{t=1}^{T}\left(\hat{u}_t\sum_{i=1}^{t}\bar{\alpha}^{i-1}\frac{\Delta y_{t-i}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-i}}\right)^2\right)^{1/2}}
= \frac{\frac{1}{\omega_{T\bar{\alpha}\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\Delta y_t\bar{C}(L)\left(\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}}\right)}{\left(\frac{1}{\omega^2_{T\bar{\alpha}\bar{\gamma}}T}\sum_{t=1}^{T}\left(\hat{u}_t\bar{C}(L)\left(\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y^2_{t-1}}\right)\right)^2\right)^{1/2}}$$

where $\bar{C}(L) = 1 + \bar{\alpha}L + \bar{\alpha}^2L^2 + \cdots$, and L denotes the lag operator. Define

$$\bar{x}_{Tt-1} = \bar{C}(L)\left(\frac{\epsilon_{t-1}}{\bar{c}/\sqrt{T} + \epsilon_{t-1}^2}\right), \qquad x_{Tt-1} = C_0(L)\left(\frac{\epsilon_{t-1}}{c_0/\sqrt{T} + \epsilon_{t-1}^2}\right).$$

where $C_0(L) = 1 + \alpha_0 L + \alpha_0^2 L^2 + \cdots$, and note that the process in equation (7) can be written as

$$\Delta y_t = \epsilon_t + \frac{c_0}{\sqrt{T}}(\alpha_0 - 1)x_{Tt-1}.$$

The only change from the case in theorem 3 is that, since x_{Tt} depends on the infinite past of ϵ_t, we cannot directly map the mixing properties of ϵ_t to x_{Tt}. However, we can utilize the concept of near-epoch dependence (NED). From Davidson (1994, definition 17.1), x_{Tt} is L_p-NED on $\{\epsilon_t\}$ if

$$(E|x_{Tt} - E(x_{Tt}|F_{t-m}^{t+m})|^p)^{1/p} \leq d_{Tt}\zeta_m$$

where $p > 0$, $\zeta_m \to 0$ and $\{d_{Tt}\}$ is a positive sequence of constants. We have

$$\begin{aligned} &E|x_{Tt} - E(x_{Tt}|F_{t-m}^{t+m})| \\ &= E\left|\sum_{i=0}^{\infty}\alpha^i\frac{\epsilon_{t-i}}{c/\sqrt{T}+\epsilon_{t-i}^2} - E\left(\sum_{i=0}^{\infty}\alpha^i\frac{\epsilon_{t-i}}{c/\sqrt{T}+\epsilon_{t-i}^2}\middle|\mathscr{F}_{t-m}^{t+m}\right)\right| \\ &= E\left|\sum_{i=m+1}^{\infty}\alpha^i\left(\frac{\epsilon_{t-i}}{c/\sqrt{T}+\epsilon_{t-i}^2} - E\left(\frac{\epsilon_{t-i}}{c/\sqrt{T}+\epsilon_{t-i}^2}\middle|\mathscr{F}_{t-m}^{t+m}\right)\right)\right| \\ &\leq \sum_{i=m+1}^{\infty}\alpha^i E\left|\frac{\epsilon_{t-i}}{c/\sqrt{T}+\epsilon_{t-i}^2} - E\left(\frac{\epsilon_{t-i}}{c/\sqrt{T}+\epsilon_{t-i}^2}\middle|\mathscr{F}_{t-m}^{t+m}\right)\right| \\ &\leq 2E\left|\frac{\epsilon_t}{c/\sqrt{T}+\epsilon_t^2}\right|\sum_{i=m+1}^{\infty}\alpha^i. \end{aligned}$$

Let $\zeta_m = \Sigma_{i=m+1}^{\infty}\alpha^i \to 0$ and

$$d_{Tt} = 2E\left|\frac{\epsilon_t}{c/\sqrt{T}+\epsilon_t^2}\right| = O(T^{\xi})$$

where $\xi > 1/8$. We can then proceed with the type of "telescoping sum" proofs given for theorem (2) and (3) (i.e., decomposing the statistic into sums of martingale differences and utilizing the mixing properties of ϵ_t and lemma (A1). The results go through in identical fashion to those in the proof of theorem (3). We present an outline below.

For the numerator of $t_{\bar{\gamma}}$:

$$\begin{aligned} &\frac{1}{\omega_{T\bar{\alpha}\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\Delta y_t\bar{C}(L)\left(\frac{\Delta y_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right) \\ &= \frac{1}{\omega_{T\bar{\alpha}\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\left(\epsilon_t + \frac{c_0}{\sqrt{T}}(\alpha_0-1)x_{Tt-1}\right)\bar{C}(L)\left(\frac{\epsilon_{t-1}+\frac{c_0}{\sqrt{T}}(\alpha_0-1)x_{Tt-2}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right) \\ &= \frac{1}{\omega_{T\bar{\alpha}\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{C}(L)\left(\frac{\epsilon_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right) \\ &+ \frac{c_0}{\omega_{T\bar{\alpha}\bar{\gamma}}T}(\alpha_0-1)\sum_{t=1}^{T}x_{Tt-1}\bar{C}(L)\left(\frac{\epsilon_{t-1}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right) \\ &+ \frac{c_0}{\omega_{T\bar{\alpha}\bar{\gamma}}T}(\alpha_0-1)\sum_{t=1}^{T}\epsilon_t\bar{C}(L)\left(\frac{x_{Tt-2}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right) \\ &+ \frac{c_0^2}{\omega_{T\bar{\alpha}\bar{\gamma}}T^{3/2}}(\alpha_0-1)^2\sum_{t=1}^{T}x_{Tt-1}\bar{C}(L)\left(\frac{x_{Tt-2}}{\bar{c}/\sqrt{T}+\Delta y_{t-1}^2}\right) \\ &= \frac{1}{\omega_{T\bar{\alpha}\bar{\gamma}}\sqrt{T}}\sum_{t=1}^{T}\epsilon_t\bar{x}_{Tt-1} + \frac{c_0}{\omega_{T\bar{\alpha}\bar{\gamma}}T}(\alpha_0-1)\sum_{t=1}^{T}x_{Tt-1}\bar{x}_{Tt-1} + o_p(1) \\ &= z + (\alpha_0-1)\frac{c_0\omega_{T\bar{\alpha}\gamma\bar{\gamma}}}{\omega_{T\bar{\alpha}\bar{\gamma}}} + o_p(1) \end{aligned}$$

where $z \sim N(0, 1)$.

For the denominator, we have

$$\begin{aligned} &\frac{1}{\omega_{T\bar{\alpha}\bar{\gamma}}^2 T}\sum_{t=1}^{T}\left(\hat{u}_t\sum_{i=1}^{t}\bar{\alpha}^{i-1}\frac{\Delta y_{t-i}}{\bar{c}/\sqrt{T}+\Delta y_{t-i}^2}\right)^2 \\ &= \frac{1}{\omega_{T\bar{\alpha}\bar{\gamma}}^2 T}\sum_{t=1}^{T}\left(\epsilon_t\sum_{i=1}^{t}\bar{\alpha}^{i-1}\frac{\epsilon_{t-i}}{\bar{c}/\sqrt{T}+\epsilon_{t-i}^2}\right)^2 + o_p(1) \\ &= 1 + o_p(1) \end{aligned}$$

Thus,

$$t_{\bar{\gamma}} - c_0(\alpha_0-1)\frac{\omega_{T\bar{\alpha}\gamma\bar{\gamma}}}{\omega_{T\bar{\alpha}\bar{\gamma}}} \xrightarrow{d} N(0, 1),$$

and the result follows. Q.E.D.

Proof of Theorem 5:

Gaussianity of $L(y^T, \varphi)$ and compactness of Ψ provide sufficient regularity for the existence of the QMLE (White, 1994), theorem (2.12).

We have

$$\begin{aligned} T^{-1}L(y^T, \varphi) &= -\frac{1}{2T\sigma^2}\sum_{t=1}^{T}(\Delta y_t - \alpha\Delta y_{t-1} + \theta_{t-1}\epsilon_{t-1})^2 - \frac{1}{2}\log(2\pi\sigma^2) \\ &= T^{-1}\sum_{t=1}^{T}\log f_t(y^t, \varphi). \end{aligned}$$

Repeated substitution for ϵ_{t-1} reveals that $f_t(y^t, \varphi)$ is a function of the increasing sequence $\{y_i: i = 1, 2 \ldots, t\}$. Thus, to show consistency, we utilize the uniform law of large numbers (ULLN) for heterogeneous processes in theorem (4.2) of Wooldridge (1994). To invoke this ULLN, we need to show that $\{\log f_t(y^t, \varphi): t = 1, 2 \ldots\}$ satisfies a weak law of large numbers (WLLN) for all $\varphi \in \Psi$ and that there exists a function $c_t(y^t) \geq 0$ that satisfies a WLLN and has the property that $|\log f_t(y^t, \varphi_1) - \log f_t(y^t, \varphi_2)| \leq c_t(y^t)\langle\varphi_1 - \varphi_2\rangle$ for all $\varphi_1, \varphi_2 \in \Psi$, where $\langle x\rangle$ denotes the Euclidean norm of the vector x. Once the ULLN has been shown to hold, we can claim consistency directly from theorem (4.3) of Wooldridge (1994).

For some $\bar{\varphi} \in \Psi$, we have

$$\log f_t(y^t, \bar{\varphi}) = -\frac{1}{2\bar{\sigma}^2}\bar{\epsilon}_t^2 - \frac{1}{2}\log(2\pi\bar{\sigma}^2),$$

where $\bar{\epsilon}_t = \Delta y_{t-1} - \bar{\alpha}\Delta y_{t-1} + \bar{\theta}_{t-1}\bar{\epsilon}_{t-1}$. We will show that $\bar{\epsilon}_t$ is L_2-near epoch department (NED) on the mixing sequence $\{\epsilon_t\}$, and then use the WLLN in theorem (1) of Andrews (1988).

Following example (17.4) in Davidson (1994), we define a function $\tilde{g}_t(\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_0)$ such that $\tilde{\epsilon}_t \equiv \tilde{g}_t$ and consider a F^t_{t-m}-measurable function $\tilde{g}^m_t(\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-m}) = \tilde{g}_t(\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-m}, 0, 0, \dots)$. From a first-order Taylor expansion, we obtain

$$\tilde{g}_t - \tilde{g}^m_t = \sum_{j=m+1}^{t} \left(\frac{\partial \tilde{g}_t}{\partial \epsilon_{t-j}}\right)^* \epsilon_{t-j},$$

where * indicates that the derivatives are evaluated at points between 0 and ϵ_{t-j}.

Now

$$\|\tilde{\epsilon}_t - E(\tilde{\epsilon}_t | \mathcal{F}^{t+m}_{t-m})\|_2 \le \|\tilde{\epsilon}_t - \tilde{g}^m_t\|_2$$

$$= \left\| \sum_{j=m+1}^{t} \left(\frac{\partial \tilde{g}_t}{\partial \epsilon_{t-j}}\right)^* \epsilon_{t-j} \right\|_2$$

$$\le \sum_{j=m+1}^{t} \left\| \left(\frac{\partial \tilde{g}_t}{\partial \epsilon_{t-j}}\right)^* \epsilon^{t-j} \right\|_2$$

where $\|x\|_p \equiv (E|x|^p)^{1/p}$. Differentiating yields

$$\frac{\partial \tilde{g}_t}{\partial \epsilon_{t-j}} = \frac{\partial \Delta y_t}{\partial \epsilon_{t-j}} - \tilde{\alpha} \frac{\partial \Delta y_{t-1}}{\partial \epsilon_{t-j}} + \left(\tilde{\theta}_{t-1} + \frac{\partial \tilde{\theta}_{t-1}}{\partial \tilde{\epsilon}_{t-1}} \tilde{\epsilon}_{t-1}\right) \frac{\partial \tilde{g}_{t-1}}{\partial \epsilon_{t-j}}$$

$$= \left(\frac{\partial \Delta y_t}{\partial \epsilon_{t-j}} - \tilde{\alpha} \frac{\partial \Delta y_{t-1}}{\partial \epsilon_{t-j}}\right) + \tilde{k}_{t-1} \left(\frac{\partial \Delta y_{t-1}}{\partial \epsilon_{t-j}} - \tilde{\alpha} \frac{\partial \Delta y_{t-2}}{\partial \epsilon_{t-j}}\right)$$

$$+ \tilde{k}_{t-1}\tilde{k}_{t-2} \left(\frac{\partial \Delta y_{t-2}}{\partial \epsilon_{t-j}} - \frac{\partial \Delta y_{t-3}}{\partial \epsilon_{t-j}}\right) + \cdots + \tilde{k}_{t-1}\tilde{k}_{t-2} \cdots$$

$$\tilde{k}_{t-j+1} \left(\frac{\partial \Delta y_{t-j+1}}{\partial \epsilon_{t-j}} - \frac{\partial \Delta y_{t-j}}{\partial \epsilon_{t-j}}\right) + \tilde{k}_{t-1}\tilde{k}_{t-2} \dots \tilde{k}_{t-j} \frac{\partial \Delta y_{t-j}}{\partial \epsilon_{t-j}},$$

where $\tilde{k}_{t-1} \equiv \tilde{\theta}_{t-1} + (\partial \tilde{\theta}_{t-1}/\partial \tilde{\epsilon}_{t-1})\tilde{\epsilon}_{t-1} = 1 - (1 - \tilde{\alpha})(1 + \tilde{\eta}_{t-1})\tilde{q}_{t-1}$ and $\tilde{\eta}_t = (\partial \tilde{q}_t/\partial \tilde{\epsilon}_t)(\tilde{\epsilon}_t/\tilde{q}_t)$.

Using $\Delta y_t = \epsilon_t + (\alpha_0 - 1)\Sigma_{i=1}^{t-1}\alpha_0^{i-1}(1 - q_{t-i})\epsilon_{t-i}$, we have

$$\frac{\partial \tilde{g}_t}{\partial \epsilon_{t-j}} = (\alpha_0 - 1)(\alpha_0^{j-1} - \tilde{\alpha}\alpha_0^{j-2}) \left(1 - q_{t-j} - \frac{\partial q_{t-j}}{\partial \epsilon_{t-j}} \epsilon_{t-j}\right)$$

$$+ \tilde{k}_{t-1}(\alpha_0 - 1)(\alpha_0^{j-2} - \alpha_0^{j-3}) \left(1 - q_{t-j} - \frac{\partial q_{t-j}}{\partial \epsilon_{t-j}} \epsilon_{t-j}\right)$$

$$+ \cdots + \tilde{k}_{t-1}\tilde{k}_{t-2} \dots \tilde{k}_{t-j+1}(\alpha_0 - 1) \left(1 - q_{t-j} - \frac{\partial q_{t-j}}{\partial \epsilon_{t-j}} \epsilon_{t-j}\right)$$

$$- \tilde{k}_{t-1}\tilde{k}_{t-2} \dots \tilde{k}_{t-j+1}\tilde{\alpha} + \tilde{k}_{t-1}\tilde{k}_{t-2} \dots \tilde{k}_{t-j}$$

which implies that

$$\left|\frac{\partial \tilde{g}_t}{\partial \epsilon_{t-j}}\right| \le |\alpha_0 - 1|\,|\alpha_0 - \tilde{\alpha}|\,|\alpha_0^{j-2} + \tilde{k}_{t-1}\alpha_0^{j-3} + \tilde{k}_{t-1}\tilde{k}_{t-2}\alpha_0^{j-4}$$

$$+ \cdots + \tilde{k}_{t-1}\tilde{k}_{t-2} \dots \tilde{k}_{t-j+2}|\,|(1 - (1 + \eta_{t-j})q_{t-j})|$$

$$+ |\alpha_0 - 1|\,|\tilde{k}_{t-1}\tilde{k}_{t-2} \dots \tilde{k}_{t-j+1}|\,|(1 - (1 + \eta_{t-j})q_{t-j})|$$

$$+ |\tilde{k}_{t-1}\tilde{k}_{t-2} \dots \tilde{k}_{t-j+1}|\,|\tilde{k}_{t-j} - \tilde{\alpha}|.$$

Now, the function $(1 + \tilde{\eta}_{t-1})\tilde{q}_{t-1}$ is bounded above by 9/8 and is positive wp 1. Thus, there exists a constant $\bar{k}$ such that $|\tilde{k}_t| \le \bar{k} < 1$ wp 1 for all t. Also, since $\alpha_0, \tilde{\alpha} \in [0, 1)$ we have $\alpha_0, \tilde{\alpha} \le \bar{k}$. It follows that

$$\left|\frac{\partial \tilde{g}_t}{\partial \epsilon_{t-j}}\right| \le |(j-1)\,\bar{k}^{j-1} + \bar{k}^{j-1}|\,|(1 - (1 + \eta_{t-j})q_{t-j})| + 2|\bar{k}^j|$$

$$\le (j + 2)\bar{k}^j$$

$$\to 0 \quad \text{as} \quad j \to \infty.$$

Now, from above

$$\|\tilde{\epsilon}_t - E(\tilde{\epsilon}_t | \mathcal{F}^{t+m}_{t-m})\|_2 \le \sum_{j=m+1}^{t} \left\| \left(\frac{\partial \tilde{g}_t}{\partial \epsilon_{t-j}}\right)^* \epsilon_{t-j} \right\|_2$$

$$\le \sum_{j=m+1}^{t} (j + 2)\bar{k}^j \|\epsilon_{t-j}\|_2$$

$$= d_t \zeta_m,$$

where $d_t = \|\epsilon_t\|_2$ and $\zeta_m = \Sigma_{j=m+1}^{\infty}(j + 2)\bar{k}^j$. Thus, $\tilde{\epsilon}_t$ is L_2-NED on $\{\epsilon_t\}$, which implies that $\tilde{\epsilon}_t^2$ is L_1-NED from Davidson (1994, theorem (17.9)). Then, from theorem (17.5) of Davidson (1994), $\{\tilde{\epsilon}_t^2 - E(\tilde{\epsilon}_t^2), \mathcal{F}_t\}$ is an L_1-mixingale with constants $d_t^* \equiv d^* = 2\|\tilde{\epsilon}_t\|_{2p}$ for some $p > 1$. To invoke the WLLN in Andrews (1988, theorem (1), we need the sequence $\{\tilde{\epsilon}_t^2\}$ to the uniformly integrable and d^* to be finite. These conditions will be satisfied if $E(\tilde{\epsilon}_t^{2p}) < \infty$ (Davidson, 1994, theorem (12.10)). Now, from Minkowski's inequality and repeated substitution,

$$E(\tilde{\epsilon}_t^{2p}) \le (\|\Delta y_t\|_{2p} + \|\tilde{\alpha}\Delta y_t\|_{2p} + \|\tilde{\theta}_{t-1}\tilde{\epsilon}_{t-1}\|_{2p})^{2p}$$

$$\le (2\|\Delta y_t\|_{2p} + \bar{k}\|\tilde{\epsilon}_{t-1}\|_{2p})^{2p}$$

$$\le \left(\frac{2}{1 - \bar{k}} \|\Delta y_t\|_{2p}\right)^{2p}$$

$$< \infty$$

since $\|\epsilon_t\|_{2p} < \infty$ implies $\|\Delta y_t\|_{2p} < \infty$. Thus, from Andrews WLLN, we have

$$T^{-1} \sum_{t=1}^{T} \log f_t(y^t, \tilde{\varphi}) \xrightarrow{p} E\left(T^{-1} \sum_{t=1}^{T} \log f_t(y^t, \tilde{\varphi})\right)$$

for each $\tilde{\varphi} \in \Psi$.

Consider a mean value expansion of $\log f_t(y^t, \varphi)$ around $\varphi = \varphi_2$

$$\log f_t(y^t, \varphi) - \log f_t(y^t, \varphi_2) = (\nabla_\varphi \log f_t(y^t, \varphi))^{*\prime}(\varphi - \varphi_2)$$

where * indicates evaluation of the derivative at a point between φ and φ_2. Thus, we have, for $\varphi = \varphi_1$,

$$|\log f_t(y^t, \varphi_1) - \log f_t(y^t, \varphi_2)| \le \langle \nabla_\varphi \log f_t(y^t, \varphi)^* \rangle \langle \varphi_1 - \varphi_2 \rangle$$

where $\langle x \rangle$ denotes the Euclidean norm of the vector x. To use the ULLN of Wooldridge (1994), we define $c_t(y^t) = \sup_{\varphi \in \Psi} \langle \nabla_\varphi \log f_t(y^t, \varphi) \rangle$ and show that it satisfies a WLLN. We have

$$\langle \nabla_\varphi \log f_t(y^t, \tilde{\varphi}) \rangle = (\tilde{\sigma}^{-4}\tilde{w}_t^2\tilde{\epsilon}_t^2 + \tilde{\sigma}^{-4}\tilde{v}_t^2\tilde{\epsilon}_t^2 + \tilde{\sigma}^{-6}(\tilde{\epsilon}_t^2 - \tilde{\sigma}^2)^2)^{1/2}$$

$$\le \tilde{\sigma}^{-2}|\tilde{w}_t\tilde{\epsilon}_t| + \tilde{\sigma}^{-2}|\tilde{v}_t\tilde{\epsilon}_t| + \tilde{\sigma}^{-3}|\tilde{\epsilon}_t^2 - \tilde{\sigma}^2|.$$

Consider

$$|\tilde{w}_t| \le |\tilde{b}_{t-1}\tilde{w}_{t-1}| + (1 - \tilde{\alpha}) \left|\frac{\tilde{\epsilon}_{t-1}^3}{(\tilde{\gamma} + \tilde{\epsilon}_{t-1}^2)^2}\right|.$$

Now

$$\left|\frac{\tilde{\epsilon}_{t-1}^3}{(\tilde{\gamma} + \tilde{\epsilon}_{t-1}^2)^2}\right| \leq \frac{3\sqrt{3}}{16\sqrt{\tilde{\gamma}}} < \frac{1}{\sqrt{\tilde{\gamma}}}$$

and $|\tilde{b}_{t-1}| < 1$ wp 1, which implies $|\tilde{w}_t| < \bar{k}|\tilde{w}_{t-1}| + (1 - \tilde{\alpha})\tilde{\gamma}^{-1/2}$ wp 1. Then, since $w_0 = 0$ wp 1, we have

$$|\tilde{w}_t| < \frac{(1 - \tilde{\alpha})}{(1 - \bar{k})\sqrt{\tilde{\gamma}}} < \infty \text{ wp 1 } \forall \tilde{\varphi} \in \Psi.$$

Similarly for $\tilde{v}_t$ we have

$$|\tilde{v}_t| \leq \sum_{i=1}^{t} \bar{k}^{i-1} |\tilde{q}_{t-i}\tilde{\epsilon}_{t-i} - \Delta y_{t-i}|$$

$$\leq \sum_{i=1}^{t} \bar{k}^{i-1}(|\tilde{\epsilon}_{t-i}| + |\Delta y_{t-i}|).$$

Now consider

$$|\tilde{\epsilon}_t| \leq |\Delta y_t| + |\tilde{\alpha}\Delta y_{t-1}| + |\tilde{\theta}_{t-1}\tilde{\epsilon}_{t-1}|$$

$$\leq |\Delta y_t| + \bar{k}|\Delta y_{t-1}| + \bar{k}|\tilde{\epsilon}_{t-1}|$$

$$\leq 2\sum_{i=0}^{t} \bar{k}^i |\Delta y_{t-i}|$$

$$= 2\sum_{i=0}^{t} \bar{k}^i \left|\epsilon_{t-i} + (\alpha_0 - 1)\sum_{j=1}^{t-i-1} \alpha_0^{j-1}(1 - q_{t-i-j})\epsilon_{t-i-j}\right|$$

$$\leq 2\sum_{i=0}^{t} \bar{k}^i \left(|\epsilon_{t-i}| + |\alpha_0 - 1| \sum_{j=1}^{t-i-1} \alpha_0^{j-1}\frac{\sqrt{\gamma_0}}{2}\right)$$

$$\leq 2\sum_{i=0}^{t} \bar{k}^i |\epsilon_{t-i}| + K_1$$

where K_1 is a finite constant.

Thus, we have

$$\langle\nabla_\varphi \log f_t(y^t, \tilde{\varphi})\rangle \leq \frac{1 - \alpha}{\tilde{\sigma}^2(1 - \bar{k})\sqrt{\tilde{\gamma}}}|\tilde{\epsilon}_t| + \frac{1}{\tilde{\sigma}^2}|\tilde{v}_t||\tilde{\epsilon}_t| + \frac{1}{\tilde{\sigma}^3}|\tilde{\epsilon}_t|^2$$

$$\leq \frac{1 - \tilde{\alpha}}{\tilde{\sigma}^2(1 - \bar{k})\sqrt{\tilde{\gamma}}}\left(2\sum_{i=1}^{t} \bar{k}^i |\epsilon_{t-i}| + K_1\right)$$

$$+ \frac{1}{\tilde{\sigma}^2}\left(2\sum_{i=1}^{t} \bar{k}^i \sum_{j=0}^{t-i} \bar{k}^j |\epsilon_{t-i-j}| + \sum_{i=1}^{t} \bar{k}^i |\epsilon_{t-i}| + K_2\right)$$

$$\times \left(2\sum_{i=1}^{t} \bar{k}^i |\epsilon_{t-i}| + K_1\right)$$

$$+ \frac{1}{\tilde{\sigma}^3}\left(2\sum_{i=1}^{t} \bar{k}^i |\epsilon_{t-i}| + K_1\right)^2 \leq c_t(y^t),$$

where $c_t(y^t) \equiv K_3((c_{1t} + K_1) + (c_{1t} + \Sigma_{i=1}^t \bar{k}^i c_{1t-i} + K_2)(c_{1t} + K_1) + (c_{1t} + K_1)^2)$ with $c_{1t} \equiv 2\Sigma_{i=0}^t \bar{k}^i |\epsilon_{t-i}|$ and K_2 and K_3 finite constants. Now consider

$$\|c_{1t} - E(c_{1t}|\mathcal{F}_{t-m}^{t+m})\|_2 = 2\left\|\sum_{i=m+1}^{t} \bar{k}^i |\epsilon_{t-i}|\right\|_2$$

$$\leq 2\sum_{i=m+1}^{t} \bar{k}^i \|\epsilon_t\|_2.$$

Thus c_{1t} is L_2-NED with $d_t = \|\epsilon_t\|_2$ and $\zeta_m = \Sigma_{i=m+1}^{\infty} \bar{k}^i$. It follows that c_{1t}^2 is L_1-NED and, since $\|\epsilon_t\|_{2p} < \infty$ for some $p > 1$, c_{1t}^2 is also uniformly integrable. Thus $c_t(y^t)$, which is a function of c_{1t}^2, obeys the WLLN in Andrews (1988, theorem (1)).

We can now assert that $\{\log f_t(y^t, \varphi)\}$ satisfies the ULLN in Wooldridge (1994, theorem (4.2)) and thus, from theorem (4.3) of Wooldridge (1994), we have $\hat{\varphi} \xrightarrow{p} \varphi_0$. Q.E.D.

Proof of Theorem 6:

Consider a first-order mean value expansion of the first-order condition around φ_0. Define $\nabla_\varphi^2 L(y^T, \varphi)$ as the hessian. The mean value expansion yields

$$\nabla_\varphi L(y^T, \hat{\varphi}) = 0 = \nabla_\varphi L(y^T, \varphi_0) + \nabla_\varphi^2 \ddot{L}(y^T, \varphi)(\hat{\varphi} - \varphi_0)$$

where $\nabla_\varphi^2 \ddot{L}(y^T, \varphi)$ has each row evaluated at a mean value (possibly different for each row) lying between $\hat{\varphi}$ and φ_0. We can rearrange to get

$$\sqrt{T}(\hat{\varphi} - \varphi_0) = H_0^{-1}T^{-1/2}\nabla_\varphi L(y^T, \varphi_0) + ((-T^{-1}\nabla_\varphi^2 \ddot{L}(y^T, \varphi))^{-1} - H_0^{-1})T^{-1/2}\nabla_\varphi L(y^T, \varphi_0)$$
$$= H_0^{-1}T^{-1/2}\nabla_\varphi L(y^T, \varphi_0) + o_p(1).$$

To verify this, we need to show that the hessian obeys a uniform law of large numbers (ULLN).

We partition the scores vector $T^{-1/2}\nabla_\varphi L(y^T, \varphi)$ as

$$T^{-1/2}\nabla_\varphi L(y^T, \varphi) = \begin{bmatrix} -\sigma^{-2}T^{-1/2}\sum_{t=1}^{T} s_t\epsilon_t \\ \sigma^{-3}T^{-1/2}\sum_{t=1}^{T}(\epsilon_t^2 - \sigma^2)\end{bmatrix},$$

where $s_t = (w_t, v_t)'$ as defined in equations (15) through (17). Similarly, we partition the hessian into the matrix

$$-T^{-1}\nabla_\varphi^2 L(y^T, \varphi) = H = \begin{bmatrix} H_1 & H_3 \\ H_3' & H_2 \end{bmatrix}$$

and consider each of the components H_1, H_2, and H_3 separately. As in the proof of theorem (5), showing that H satisfies a ULLN requires verification of two conditions (Wooldridge, 1994, theorem (4.2)). First, we show that the elements of H have summands that are L_1-NED on $\{\epsilon_t\}$ and satisfy the WLLN in Andrews (1988, theorem (1)) for each $\varphi \in \Psi$. Second, we require existence of some matrix function $c_t(y^t)$ that satisfies a WLLN and has the property that $|h(y^t, \varphi_1) - h(y^t, \varphi_2)| \leq c_t(y^t)|\varphi_1 - \varphi_2|$ for all φ_1, $\varphi_2 \in \Psi$, where $h(y^t, \varphi_1)$ denotes the summands of H.

For some $\tilde{\varphi} \in \Psi$, consider

$$\tilde{H}_1 = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}\left(\tilde{s}_t\tilde{s}_t' + \tilde{\epsilon}_t\frac{\partial\tilde{s}_t}{\partial\tilde{\varphi}_{12}}\right)$$

where $\varphi_{12} = (\gamma\alpha)$. We showed in the proof of theorem (5) that $\{\tilde{\epsilon}_t\}$ is L_2-NED on $\{\epsilon_t\}$, and we show below that $\{\tilde{s}_t\}$ and $\{\partial\tilde{s}_t/\partial\tilde{\varphi}_{12}\}$ are L_2-NED on $\{\epsilon_t\}$. It then follows from Davidson (1994, theorem (17.9)) that the summands of $\tilde{H}_1$ are L_1-NED on $\{\epsilon_t\}$.

Recall that $s_t = (w_t, v_t)'$. As in the proof of theorem (5), consider a function $\tilde{w}_t^m$, which is an $\mathcal{F}_{t-m}^t$-measurable approximation $\tilde{w}_t$ defined such that a first-order Taylor expansion produces

$$\tilde{w}_t - \tilde{w}_t^m = \sum_{j=m+1}^{t}\left(\frac{\partial\tilde{w}_t}{\partial\epsilon_{t-j}}\right)^* \epsilon_{t-j},$$

where * indicates that the derivatives are evaluated at points between 0 and ϵ_{t-j}. Now,

$$\frac{\partial \tilde w_t}{\partial \epsilon_{t-j}} = \tilde b_{t-1}\frac{\partial \tilde w_{t-1}}{\partial \epsilon_{t-j}} + \tilde w_{t-1}\frac{\partial \tilde b_{t-1}}{\partial \tilde\epsilon_{t-1}}\frac{\partial \tilde g_{t-1}}{\partial \epsilon_{t-j}} + \frac{3\tilde\gamma\tilde\epsilon_{t-1}^2 - \tilde\epsilon_{t-1}^4}{(\tilde\gamma+\tilde\epsilon_{t-1}^2)^3}\frac{\partial \tilde g_{t-1}}{\partial \epsilon_{t-j}}$$

where $\tilde g_t$ is as defined in the proof of theorem (5). Now, since there exists $\bar k<1$ such that $\tilde b_{t-1}\le \bar k$ wp 1 for all t, we have

$$\left|\frac{\partial \tilde w_t}{\partial\epsilon_{t-j}}\right| \le \left|\frac{1}{1-\bar k}\left(\tilde w_{t-1}\frac{\partial \tilde b_{t-1}}{\partial\tilde\epsilon_{t-1}} + \frac{3\tilde\gamma\tilde\epsilon^2_{t-1}-\tilde\epsilon^4_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^3}\right)\frac{\partial\tilde g_{t-1}}{\partial\epsilon_{t-j}}\right| \text{ wp 1.}$$

Note that $|\tilde w_t|$ is bounded wp 1 (see the proof of theorem (5)) and that

$$\left|\frac{\partial\tilde b_{t-1}}{\partial\tilde\epsilon_{t-1}}\right| = (1-\tilde\alpha)\left|\frac{6\tilde\gamma^2\tilde\epsilon_{t-1} - 2\tilde\gamma\tilde\epsilon^3_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^3}\right| < \infty \text{ wp 1.}$$

This leads to

$$\left|\frac{\partial\tilde w_t}{\partial\epsilon_{t-j}}\right| \le K_w\left|\frac{\partial\tilde g_{t-1}}{\partial\epsilon_{t-j}}\right| \text{ wp 1}$$

for some $K_w<\infty$. Thus,

$$\|\tilde w_t - E(\tilde w_t|\mathcal{F}_{t-m}^{t+m})\|_2 \le \sum_{j=m+1}^t \left\|\left(\frac{\partial\tilde w_t}{\partial\epsilon_{t-j}}\right)^*\epsilon_{t-j}\right\|_2$$

$$\le K_w\sum_{j=m+1}^t\left\|\left(\frac{\partial\tilde g_t}{\partial\epsilon_{t-j}}\right)^*\epsilon_{t-j}\right\|_2 .$$

$$\le K_w\sum_{j=m+1}^t (j+2)\bar k^j\|\epsilon_{t-j}\|_2$$

from the proof of theorem (5). It follows that, since $\|\epsilon_t\|_2<\infty$ and $\lim_{m\to\infty}\sum_{j=m+1}^t(j+2)\bar k^j = 0$, $\{\tilde w_t\}$ is L_2-NED on $\{\epsilon_t\}$ (Davidson, 1994, definition (17.1)).

Similarly, we have

$$\left|\frac{\partial\tilde v_t}{\partial\epsilon_{t-j}}\right| = \left|\tilde b_{t-1}\frac{\partial\tilde v_{t-1}}{\partial\epsilon_{t-j}} + \tilde v_{t-1}\frac{\partial\tilde b_{t-1}}{\partial\tilde\epsilon_{t-1}}\frac{\partial\tilde g_{t-1}}{\partial\epsilon_{t-j}} + \frac{\tilde\epsilon^2_{t-1}(\tilde\epsilon^2_{t-1}+3\tilde\gamma)}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^2}\frac{\partial\tilde g_{t-1}}{\partial\epsilon_{t-j}} + \frac{\partial\Delta y_{t-1}}{\partial\epsilon_{t-j}}\right|$$

$$\le \sum_{i=1}^j \bar k^i\left|{}_{t-i}\frac{\partial\tilde b_{t-i}}{\partial\tilde\epsilon_{t-i}}\frac{\partial\tilde g_{t-i}}{\partial\epsilon_{t-j}} + \frac{\tilde\epsilon^2_{t-i}(\tilde\epsilon^2_{t-i}+3\tilde\gamma)}{(\tilde\gamma+\tilde\epsilon^2_{t-i})^2}\frac{\partial\tilde g_{t-i}}{\partial\epsilon_{t-j}} + \frac{\partial\Delta y_{t-i}}{\partial\epsilon_{t-j}}\right|$$

$$\le K_{v1}\sum_{i=1}^j\bar k^i(|\tilde v_{t-i}|+K_{v2})\left|\frac{\partial_{t-i}}{\partial\epsilon_{t-j}}\right| + \sum_{i=1}^j\bar k^i\left|\frac{\partial\Delta y_{t-i}}{\partial\epsilon_{t-j}}\right|$$

$$\le \bar k^j(j+2)K_{v1}\sum_{i=1}^j\bar k^i(|\tilde v_{t-i}|+K_{v2}) + j\bar k^j$$

for some finite constants K_{v1} and K_{v2}. As for $\tilde w_t$, we can write

$$\|\tilde v_t - E(\tilde v_t|\mathcal F^{t+m}_{t-m})\|_2 \le \sum_{j=m+1}^t\left\|\left(\frac{\partial\tilde v_t}{\partial\epsilon_{t-j}}\right)^*\epsilon_{t-j}\right\|_2$$

Then, expressing $\tilde v_t$ as a function of ϵ_t as in the proof of theorem (5), and using the fact that $\|\epsilon_t\|_{4p}<\infty$ for some $p>1$ leads to the result that $\{\tilde v_t\}$ is L_2-NED on $\{\epsilon_t\}$.

Now $\partial\tilde s_t/\partial\tilde\varphi_{12}$ is a symmetric 2×2 matrix with the following components:

$$\frac{\partial\tilde w_t}{\partial\tilde\gamma} = \tilde b_{t-1}\frac{\partial\tilde w_{t-i}}{\partial\tilde\gamma} + \frac{\partial\tilde b_{t-1}}{\partial\tilde\gamma}\tilde w_{t-1} - (1-\tilde\alpha)\frac{2\tilde\epsilon^3_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^3}$$
$$- (1-\tilde\alpha)\frac{3\tilde\gamma\tilde\epsilon^2_{t-1}-\tilde\epsilon^4_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^3}\tilde w_{t-1},$$

$$\frac{\partial\tilde w_t}{\partial\tilde\alpha} = \frac{\partial\tilde v_t}{\partial\tilde\gamma} = \tilde b_{t-1}\frac{\partial\tilde v_{t-1}}{\partial\tilde\gamma} + \frac{\partial\tilde b_{t-1}}{\partial\tilde\gamma}\tilde v_{t-1} - \frac{\tilde\epsilon^3_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^2}$$
$$+ \frac{3\tilde\gamma\tilde\epsilon^2_{t-1}+\tilde\epsilon^4_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^2}\tilde w_{t-1},$$

$$\frac{\partial\tilde v_t}{\partial\tilde\alpha} = \tilde b_{t-1}\frac{\partial\tilde v_{t-1}}{\partial\tilde\alpha}+\frac{\partial\tilde b_{t-1}}{\partial\tilde\alpha}\tilde v_{t-1},$$

where

$$\frac{\partial\tilde b_{t-1}}{\partial\tilde\gamma} = (1-\tilde\alpha)\frac{(3\tilde\gamma-\tilde\epsilon^2_{t-1})\tilde\epsilon^2_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^3} - (1-\tilde\alpha)\frac{6\tilde\gamma^2\tilde\epsilon_{t-1}-2\tilde\gamma\tilde\epsilon^3_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^3}\tilde w_{t-1},$$

$$\frac{\partial\tilde b_{t-1}}{\partial\tilde\alpha} = \frac{(3\tilde\gamma+\tilde\epsilon^2_{t-1})\tilde\epsilon^2_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^2} - (1-\tilde\alpha)\frac{6\tilde\gamma^2\tilde\epsilon_{t-1}-2\tilde\gamma\tilde\epsilon^3_{t-1}}{(\tilde\gamma+\tilde\epsilon^2_{t-1})^3}\tilde v_{t-1}.$$

By processing in a manner analogous to that for w_t and v_t above, we can show that the elements of $\{\partial\tilde s_t/\partial\tilde\varphi_{12}\}$ are L_2-NED on $\{\epsilon_t\}$. Thus, from Davidson (1994, theorem (17.9)), the summands of $\tilde H_1$ are L_1-NED on $\{\epsilon_t\}$. Further, since $\tilde s_t$, $\partial\tilde s_t/\partial\tilde\varphi_{12}$ and $\tilde\epsilon_t$ have more than two finite moments, we can invoke the WLLN in Andrews (1988, theorem (1)). Thus, we have $\tilde H_1 - E(\tilde H_1)\xrightarrow{p}0$ for each $\tilde\varphi\in\Psi$.

Now the (1, 1) element of the matrix H_1 is

$$H_{11} = \frac{1}{T\sigma^2}\sum_{t=1}^t\left(w_t^2+\epsilon_t\frac{\partial w_t}{\partial\gamma}\right) \equiv \frac1T\sum_{t=1}^T h_{t11}.$$

Consider a mean value expansion of h_{t11} around $\varphi=\check\varphi$

$$h_{t11}-\check h_{t11} = (\nabla_\varphi h_{t11})^{*\prime}(\varphi-\check\varphi)$$

where * indicates evaluation of the scores function at a point between φ and $\check\varphi$. We have, for $\varphi=\check\varphi$,

$$|\hat h_{t11}-\check h_{t11}|\le\langle\nabla_\varphi h^*_{t11}\rangle\langle\hat\varphi-\check\varphi\rangle,$$

where $\langle x\rangle$ denotes the Euclidean norm of the vector x. Note we must show that there exists a function $c_{t11}(y^t)$ independent of φ that satisfies a WLLN and has the property that $\langle\nabla_\varphi h^*_{t11}\rangle\le c_{t11}(y^t)$. The elements of $\nabla_\varphi\tilde h_{t11}$ for some $\varphi=\tilde\varphi$ are as follows:

$$\left|\frac{\partial\tilde h_{t11}}{\partial\tilde\gamma}\right| = \left|\frac{1}{\tilde\sigma^2}\left(2\tilde w_t\frac{\partial\tilde w_t}{\partial\tilde\gamma}+\tilde\epsilon_t\frac{\partial^2\tilde w_t}{\partial\tilde\gamma^2}+\tilde w_t\frac{\partial\tilde w_t}{\partial\tilde\gamma}\right)\right| \le |K_{1\gamma}+K_{2\gamma}\tilde\epsilon_t|$$

$$\left|\frac{\partial\tilde h_{t11}}{\partial\tilde\alpha}\right| = \left|\frac{1}{\tilde\sigma^2}\left(2\tilde w_t\frac{\partial\tilde w_t}{\partial\tilde\alpha}+\tilde\epsilon_t\frac{\partial^2\tilde w_t}{\partial\tilde\alpha\partial\tilde\gamma}+\tilde w_t\frac{\partial\tilde w_t}{\partial\tilde\alpha}\right)\right| \le |K_{1\alpha}+K_{2\alpha}\tilde\epsilon_t|$$

$$\left|\frac{\partial\tilde h_{t11}}{\partial\tilde\sigma}\right| = \left|\frac{2}{\tilde\sigma^3}\left(\tilde w_t^2+\tilde\epsilon_t\frac{\partial\tilde w_t}{\partial\tilde\gamma}\right)\right| \le |K_{1\sigma}+K_{2\sigma}\tilde\epsilon_t|$$

where $K_{ij} < \infty$ are constants independent of φ. The inequalities follow from wp 1 boundedness of the terms in parentheses and imply that there exist finite constants K_0, K_1 and K_2 such that

$$|\nabla_\varphi h_{t11}| \leq |K_0 + K_1\tilde{\epsilon}_t + K_2\tilde{\epsilon}_t^2|$$

for all $\tilde{\varphi} \in \Psi$. Now since $|K_0 + K_1\tilde{\epsilon}_t + K_2\tilde{\epsilon}_t^2|$ obeys a WLLN (from the proof of theorem (5)), the ULLN in Wooldridge (1994, theorem (4.2)) holds for the (1, 1) element of H_1. Although terms involving $\tilde{v}_t$ are not bounded wp 1, we can use arguments similar to those above and in the proof of theorem (5) to show that a function c_t exists for the other elements of H_1. It follows that $H_1 - E(H_1) \xrightarrow{p} 0$ uniformly on Ψ.

Consider

$$\tilde{H}_2 = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}(3\tilde{\epsilon}_t^2/\tilde{\sigma}^2 - 1).$$

Since $\tilde{\sigma}^2 > 0$ and since we showed in theorem (5) that $\{\tilde{\epsilon}_t^2\}$ obeys a ULLN, we can assert that $H_2 - E(H_2) \xrightarrow{p} 0$ uniformly on Ψ.

Consider

$$\tilde{H}_3 = -\frac{2}{T\tilde{\sigma}^3}\sum_{t=1}^{T}\tilde{s}_t\tilde{\epsilon}_t.$$

Since both $\{\tilde{s}_t\}$ and $\{\tilde{\epsilon}_t\}$ are L_2-NED on $\{e_t\}$, we have from Davidson (1994, theorem (17.9)) that $\{\tilde{s}_t\tilde{\epsilon}_t\}$ is L_1-NED on $\{\epsilon_t\}$ and satisfies Andrew's WLLN. Further, using arguments similar to those above and the fact that $\{\tilde{\epsilon}_t^2\}$ satisfies a WLLN, we can show that there exists a function $c_{t3}(y^t)$ independent of φ that satisfies a WLLN and has the property that $|h_3(y^t, \varphi_1) - h_3(y^t, \varphi_2)| \leq c_{t3}(y^t)(\varphi_1 - \varphi_2)$ for all $\varphi_1, \varphi_2 \in \Psi$, where $h_3(y^t, \varphi_1)$ denotes the summands of H_3. Thus, $H_3 - E(H_3) \xrightarrow{p} 0$ uniformly on Ψ.

From the above arguments, we have that $T^{-1}\nabla_\varphi^2 L(y^T, \varphi) - E(T^{-1}\nabla_\varphi^2 L \times (y^T, \varphi)) \xrightarrow{p} 0$ uniformly on Ψ, and we must now verify that $(-T^{-1}\nabla_\varphi^2 \ddot{L}(y^T, \varphi)) - H_0 \xrightarrow{p} 0$. Consider

$$(-T^{-1}\nabla_\varphi^2 L(y^T, \hat{\varphi})) - H_0 = (-T^{-1}\nabla_\varphi^2 L(y^T, \hat{\varphi}) - E(-T^{-1}\nabla_\varphi^2 L(y^T, \hat{\varphi}))) + (E(-T^{-1}\nabla_\varphi^2 L(y^T, \hat{\varphi})) - H_0).$$

Now, the first term goes to zero from the above ULLN. Since $T^{-1}\nabla_\varphi^2 \times L(y^T, \varphi)$ is a continuous function of φ, consistency of $\hat{\varphi}$ for φ_0 ensures that the second term goes to zero in probability (White, 1984, proposition (2.11)). Thus, $(-T^{-1}\nabla_{\hat{\varphi}}^2 L(y^T, \hat{\varphi})) - H_0 \xrightarrow{p} 0$, which implies that $(-T^{-1}\nabla_\varphi^2 \times \ddot{L}(y^T, \varphi)) - H_0 \xrightarrow{p} 0$ since the arguments of $\ddot{L}(y^T, \varphi)$ always lie between $\hat{\varphi}$ and φ_0.

Note at this point that

$$H_{01} = E\left(\frac{1}{T\sigma_0^2}\sum_{t=1}^{T}\left(s_t s_t' + \epsilon_t\frac{\partial s_t}{\partial \varphi_{12}}\right)\right) = \frac{1}{T\sigma_0^2}\sum_{t=1}^{T}E(s_t s_t'),$$

$$H_{02} = E\left(\frac{1}{T\sigma_0^2}\sum_{t=1}^{T}(3\epsilon_t^2/\sigma_0^2 - 1)\right) = \frac{2}{\sigma_0^2} \quad \text{and}$$

$$H_{03} = E\left(-\frac{2}{T\sigma_0^3}\sum_{t=1}^{T}s_t\epsilon_t\right) = 0$$

from the martingale difference property of $\{\epsilon_t, \mathcal{F}_t\}$.

Now consider

$$\lambda' V_0^{-1/2}T^{-1/2}\nabla_\varphi L(y^T, \varphi_0) = \lambda' V_0^{-1/2}\begin{bmatrix} -\sigma_0^{-2}T^{-1/2}\sum_{t=1}^{T} s_t\epsilon_t \\ \sigma_0^{-3}T^{-1/2}\sum_{t=1}^{T}(\epsilon_t^2 - \sigma_0^2)\end{bmatrix} \equiv \lambda' V_0^{-1/2}T^{-1/2}\sum_{t=1}^{T}Z_t$$

where $\lambda'\lambda = 1$. We use the central-limit theorem for mixingales in theorem (1) of de Jong (1997) to show that $\lambda' V_0^{-1/2}T^{-1/2}\Sigma_{t=1}^T Z_t \xrightarrow{p} N(0, 1)$. It then follows from the Cramér-Wold device (White, 1984, proposition (5.1)) that $V_0^{-1/2}T^{-1/2}\Sigma_{t=1}^T Z_t \xrightarrow{p} N(0, I)$. For all $m > 0$, we can utilize the martingale difference property of $\{s_t\epsilon_t, \mathcal{F}_t\}$ to obtain

$$\|E(\lambda' V_0^{-1/2}T^{-1/2}Z_t|\mathcal{F}_{Tt-m})\|_2 = \|E(\lambda_3^* T^{-1/2}\sigma_0^{-3}(\epsilon_t^2 - \sigma_0^2)|\mathcal{F}_{Tt-m})\|_2 \leq 2(1 + \sqrt{2})\sigma_0^{-3}\lambda_3^* T^{-1/2}\alpha(m)^{(p-1)/2p}\|\epsilon_t^2\|_{2p}$$

from Davidson (1994, theorem (14.2)), where λ_3^* denotes the third element of the vector $\lambda' V_0^{-1/2}$ and $p > 1$. Thus, since $\|\epsilon_t^2\|_{2p} < \infty$ and $\alpha(m)$ is of size $p/(p-1)$, the sequence $\{\lambda' V_0^{-1/2}T^{-1/2}Z_t, \mathcal{F}_{Tt}\}$ is an L_2-mixingale of size $-1/2$ with mixiangle indices $a_{Tt} = 2(1 + \sqrt{2})\sigma_0^3\lambda_3^* T^{-1/2}\|\epsilon_t^2\|_{2p}$. The assumption that $\|\epsilon_t^2\|_{2p} < \infty$ is then sufficient for the remaining conditions of de Jong's Theorem. Thus, we have

$$V_0^{-1/2}T^{-1/2}\nabla_\varphi L(y^T, \varphi_0) \xrightarrow{p} N(0, 1),$$

and the result follows. Q.E.D.