Regression Theory when Variances are Non-stationary

Hansen, Bruce E.

Working Paper No. 226
April 1990

University of
Rochester

Regression Theory When Variances are Non-Stationary

Bruce E. Hansen

Rochester Center for Economic Research
Working Paper No. 226

REGRESSION THEORY WHEN VARIANCES ARE NON-STATIONARY

Bruce E. Hansen

University of Rochester
Department of Economics
Harkness Hall
Rochester, NY 14627
(716) 275-7307

Working Paper No. 226

April 1990

ABSTRACT

This paper develops a theory of estimation and inference in linear regression models with non-stationarity arising in the variance. This contrasts with the main thrust of recent theoretical literature on linear regression with integrated processes. Non-stationarity in the variance is a plausible candidate for describing the behavior of many financial variables. A tractable model of non-stationarity in the variance, called bi-integration, is proposed and evaluated. A test for the existence of bi-integration is derived from the theory of Chow sequences.

The primary purpose of the paper is to develop a theory of ordinary least squares and instrumental variables regression under the assumption that the data are bi-integrated. The conditions for identification of the regression slope parameter are obtained. These identification conditions are unconventional in two senses: (i) Consistent estimation of the reduced form is not necessary for consistent estimation of the structural parameter; and (ii) Identification does not depend upon the stochastic structure of the non-stationary variances. T-statistics for the regression slope parameter are asymptotically normal if a variance estimate robust to heteroskedasticity and serial correlation is used.

1. Introduction

The classical linear regression model assumes that the data (regressors and regression errors) are independent over observations and identically distributed. The theory has been gradually extended to cover non–iid situations. It is well known that independence may be replaced by some form of asymptotic independence (such as mixing) and homoskedasticity may be replaced by some form of bounded heteroskedasticity. White (1984) provides an excellent exposition of this theory.

More recently, econometricians have examined more severe departures from the classical model. Many variables measured in levels or log–levels may be usefully described as integrated processes (difference stationary), which are neither asymptotically independent nor of bounded heteroskedasticity. A theory of regression for integrated processes has emerged in a series of papers, including Engle and Granger (1987), Phillips (1987), Park and Phillips (1988) and Phillips and Hansen (1990).

Even after differencing, however, many economic (especially financial) variables are still non–classical in that there appears to be substantial conditional heteroskedasticity. Following the lead of Engle (1982), many researchers have used likelihood techniques to estimate parameterized models of conditional heteroskedasticity, including ARCH (Engle, 1982), GARCH (Bollerslev, 1986) and E–GARCH (Nelson, 1989b), as well as several other variants. Bollerslev and Engle (1986, 1989) suggest that many financial variables are well described by low–order GARCH processes very close to the region of "integration". Consider a GARCH(1,1) process which may be written

$$x_t \quad = \quad \sigma_t \epsilon_t \, , \qquad \epsilon_t \equiv N(0,1)$$

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + \alpha_1\sigma_{t-1}^2 + \alpha_2 x_{t-1}^2 \\ &= \alpha_0 + (\alpha_1 + \alpha_2)\sigma_{t-1}^2\epsilon_{t-1}^2 - \alpha_1\sigma_{t-1}^2(\epsilon_{t-1}^2 - 1) .\end{aligned}$$

This process is "integrated" (I–GARCH) if $\alpha_1 + \alpha_2 = 1$. In this case the unconditional variance of x_t is infinite.

Since the I–GARCH process seems to be an empirically important member of the ARCH family, it is of interest to develop an asymptotic theory for estimation with I–GARCH processes. Nelson (1989a) has made a major step in this direction. The difficulty of the task stems from the fact that I–GARCH processes are non–linear with an infinite unconditional variance. (Standard theory under infinite variances requires linearity.) This makes application of existing theory quite challenging.

I–GARCH models are estimated not because they are "true" descriptions of reality, but because they capture an important characteristic of the data: persistence in the variance. If the question of interest is the behavior of standard statistical techniques (such as ordinary least squares and instrumental variables) using data displaying persistence in the variance, it seems quite reasonable to begin with alternative models which are tractable in a linear framework.

Consider a stochastic process which can be written as

$$x_t = \sigma_t\epsilon_t , \qquad \{\epsilon_t\} \text{ iid}$$

and σ_t is stochastic yet "slow–moving". Note that the GARCH model generates σ_t^2 by an ARMA process with "innovations" x_t^2 . This makes σ_t a complicated non–linear function of $\{\epsilon_j , j < t\}$. To impose linearity, we instead postulate that σ_t follows an ARMA process with some innovation η_t . For simplicity, take an AR(1):

$$\sigma_t = \alpha\,\sigma_{t-1} + \eta_t .$$

Persistence in variance is obtained by setting $\alpha = 1$. (More generally, requiring a unit root in the AR process for σ_t.) We may allow arbitrary dependence between ϵ_t and η_t . This process, which is a special case of the general bi–integrated process introduced in the next section, has the favorable property that an asymptotic theory of regression can be readily obtained. A disadvantage is that maximum likelihood techniques are not immediately available.

A feel for the behavior of bi–integrated processes can be given by sample plots of their trajectories. Figures 1 and 2 display the time paths of two bi–integrated processes where

$$x_t = \sigma_t \epsilon_t$$

$$\sigma_t = \sigma_{t-1} + .6\eta_t \quad , \quad \sigma_0 = 1$$

$$\begin{bmatrix} \epsilon_t \\ \eta_t \end{bmatrix} \text{ iid } N(0,\Omega), \qquad \Omega = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} .$$

The correlation ρ is set to zero in figure 1 and set to unity in figure 2. The sample paths of x_t display the "volatility" which is commonly observed in financial data.

The plan of the paper is as follows. Section 2 introduces bi–integrated processes and derives some general results on the behavior of sample statistics. Section 3 introduces a test for bi–integration using the theory of Chow sequences and reports a simple application. Section 4 develops a theory of regression analysis for bi–integrated variables. For simplicity, attention is restricted to a single regressor and a single instrumental variable. The conditions for identification of the structural parameters are unconventional, in that the reduced form estimates need not converge to constants. Section 5 discusses the problem of inference. It is found that the non–stationarity requires the use of robust variance estimates which yields asymptotic normality of t–statistics under suitable conditions. Section 6 concludes. Proofs are collected in an appendix.

Throughout the paper, I use "$\Longrightarrow$" to denote weak convergence, "$\equiv$" to denote equality in distribution, and "$[\cdot]$" to denote "integer part". Brownian motions $B(r)$ on $r \in [0,1]$ are often written as B, and integrals such as $\int_0^1 B(r)$ are frequently written as $\int_0^1 B$ and sometimes as $\int B$, to achieve notational economy. All limits are taken as the sample size T tends to infinity.

2. Bi–Integrated Processes

Definition. x_t is a *Bi–Integrated* process of order d, denoted BI(d), if x_t has the representation

$$(1) \qquad x_t = \sigma_t \epsilon_t$$

where $\Delta^d \sigma_t \equiv I(0)$, and $\epsilon_t \equiv I(0)$. It is convenient to normalize ϵ_t so that $E(\epsilon_t^2) = 1$.

The terminology "bi–integrated" is motivated by the concept of "bi–linear" processes, introduced by Granger and Andersen (1978). A simple bi–linear process is

$$y_t = \alpha y_{t-1} + \epsilon_t + \beta y_{t-1}\epsilon_{t-1} ,$$

which is a non–linear function of the innovations $\{\epsilon_t\}$. The "bi–linearity" arises from the product term $y_{t-1}\epsilon_{t-1}$. Thus the product involved in (1) motivates the term "bi–integration". In this paper, we restrict attention to the leading case of BI(1) processes, so $\sigma_t \equiv I(1)$.

Define $u_t = (\epsilon_t, \Delta\sigma_t)'$. We need

Assumption 1. $\{u_t\}$ is strictly stationary and satisfies

(i) $Eu_t = 0$,

(ii) $E|u_t'u_t|^{2p} < \infty$, for some $p > 1$,

(iii) $\Sigma_{m=1}^{\infty} \alpha_m^{(p-1)/p} < \infty$, where $\{\alpha_m\}$ are the strong mixing coefficients of $\{u_t\}$, and

(iv) $\lim_{T\to\infty} T^{-1}E(S_T S_T') = \Omega$, $S_T = \Sigma_1^T u_t$.

Under assumption 1 we have the invariance principle

$$T^{-1/2}S_{[Tr]} \Rightarrow B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} \equiv BM(\Omega) .$$

We partition Ω as

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$$

and assume $\omega_{11} > 0$, $\omega_{22} > 0$. Some interesting general distributional results are given in the following theorem.

Theorem 1. *Under assumption 1*

(a) $\bar{X} = T^{-1}\Sigma_1^T x_t \Rightarrow \int_0^1 B_2 dB_1 + \lambda_{21}$, $\lambda_{21} = \Sigma_{j=1}^{\infty} E(\Delta\sigma_1 \epsilon_j)$,

(b) $\hat{\sigma}^2 = T^{-1}\Sigma_1^T(x_t - \bar{X})^2$ *satisfies* $T^{-1}\hat{\sigma}^2 \Rightarrow \int_0^1 B_2^2$,

(c) $(\hat{\sigma}^2 T)^{-1} \sum_{t=k+1}^{T} (x_t - \bar{X})(x_{t-k} - \bar{X}) \to_p E(\epsilon_k \epsilon_0)$,

(d) $\sqrt{T}\,\dfrac{(\bar{X} - \lambda_{21})}{\hat{\sigma}} \to_d Z$, $Z = \left[\sqrt{1-\rho^2}\cdot N + \rho\, D_f\right] \sqrt{\omega_{11}}$,

where $\rho = \dfrac{\omega_{12}}{\sqrt{\omega_{11}\omega_{22}}}$, $N \equiv N(0,1)$, *independent of* $D_f \equiv \dfrac{\int W dW}{\sqrt{\int W^2}}$, $W \equiv BM(1)$.

Theorem 1 states the following. Part (a) says that the sample mean of a bi–integrated process converges to a limit random variable. The mean of this limit distribution is λ_{21}. Part (b) notes that the sample variance diverges at rate T, but when appropriately standardized, converges to a random variable. This underscores the fact that a bi–integrated process has a *non–stationary variance.* Part (c) says that estimated correlation coefficients converge to the correlations of the process $\{\epsilon_t\}$. In many applications, it is quite reasonable to suppose that $\{\epsilon_t\}$ is white noise. In these cases, estimated correlation coefficients will be close to zero (at least in large samples). This result points out that bi–integrated processes, while possessing a non–stationary variance, display stationary behavior in the correlations. Part (d) shows that t–statistics have limiting distributions which are given by a scaled mixture of a standard normal and a Dickey–Fuller t–distribution, the mixture depending on the long–run correlation between the processes $\{\epsilon_t\}$ and $\{\Delta\sigma_t\}$.

3. Testing for Bi–integration

It is desirable to have a test which can detect the existence of bi–integration. One consistent test is to split the sample and compare the estimated variances in the subsamples. This is known as the Chow test for change in the variance. The null hypothesis in this case is that the (unconditional) variance is constant throughout the sample, so we have

$$H_o \;:\; x_t \;=\; \sigma\,\epsilon_t, \qquad \epsilon_t \equiv I(0)\;.$$

If the sample split is in period t, the test statistic is

$$\text{(2)} \qquad \text{Chow}_t \;=\; (\hat{\sigma}_t^2 - \hat{\sigma}_T^2)/\hat{V}_t \;,\; \hat{\sigma}_t \;=\; t^{-1}\Sigma_1^t x_j^2 \;,\; \hat{\sigma}_T \;=\; T^{-1}\Sigma_1^T x_j^2$$

and $\hat{V}_t$ is an estimate of the variance

$$\begin{aligned} V_t &= \mathrm{Var}(\hat{\sigma}_t^2 - \hat{\sigma}_T^2) = \mathrm{Var}(\hat{\sigma}_t^2) - \mathrm{Var}(\hat{\sigma}_T^2) \\ &= E(t^{-1}\Sigma_1^t \eta_j)^2 - E(T^{-1}\Sigma_1^T \eta_j)^2 \\ &= \omega_\eta/t - \omega_\eta/T = \left[\frac{1}{t} - \frac{1}{T}\right]\omega_\eta \end{aligned}$$

where

$$\eta_t = x_t^2 - Ex_t^2 \quad , \quad \omega_\eta = \lim_{T\to\infty} E\, T^{-1}(\Sigma_1^T \eta_t)^2 .$$

A consistent estimate of this variance is given by

$$(3) \qquad \hat{V}_t = \left[\frac{1}{t} - \frac{1}{T}\right]\hat{\omega}_\eta , \quad \hat{\omega}_\eta = T^{-1}\left[\Sigma_1^T \hat{\eta}_t^2 + 2\sum_{m=1}^{M} k_m \sum_{t=m+1}^{T} \hat{\eta}_t \hat{\eta}_{t-m}\right] ,$$

$$\hat{\eta}_t = x_t^2 - \hat{\sigma}_T^2 \quad , \quad k_m = 1 - m/(M+1) .$$

The above choice for the function k_m is the Bartlett kernel, but any other standard kernel can alternatively be used. The lag truncation number M should be selected so that M goes to infinity much slower than T. The optimal selection for M in testing problems is unknown, although Andrews (1989a) gives some results for minimizing the mean squared error of estimated variance parameters. Under standard assumptions (namely, assumption 1 plus consistency of $\hat{\omega}_\eta$), this Chow test statistic is asympotically χ_1^2 under the null hypothesis.

This Chow test would be useful in testing the hypothesis of a constant variance against the alternative of a variance which changes at a single (known) time t. If the alternative is bi–integration, this may not be particularly powerful. A more reasonable procedure may be to look at the entire sequence of Chow tests evaluated for each changepoint t. Test statistics can then be found by taking either the maximum or the average value of the Chow sequence over some range, such as $R = [.15 \cdot T, .85 \cdot T]$. This test is examined in some detail in Andrews (1989b), Chu (1989), and Hansen (1990), and is quite similar in this context to the CUSUM test proposed by Pagan and Schwert (1989). The statistics are

$$(4) \qquad \text{MaxChow} = \underset{t \in R}{\text{Max}} \ \text{Chow}_t$$

$$\text{MeanChow} = \frac{1}{.6 \cdot T} \sum_{t \in R} \text{Chow}_t$$

Asymptotic critical values under the null of no change in the variance are given in table 1.

Table 1:
Upper percentage points from
Asymptotic Chow Test Distributions

	.100	.075	.050	.025	.010
MaxChow	6.9	7.6	8.5	10.1	12.0
MeanChow	2.15	2.42	2.86	3.62	4.61

Source: Hansen (1990)

The robust covariance estimate of ω_η in (3) allows for the squared x_t to possess serial correlation under the null hypothesis, as long as the squared process satisfies an asymptotic independence condition. This should allow GARCH processes to be included under the null hypothesis. (This is not a rigorous statement, for the existing theorems assume that the data satisfies a strong mixing condition, yet it is unknown whether GARCH processes are mixing.) The theorems require that η_t^2 has at least $2+\delta$ moments finite, which requires x_t to possess finite unconditional $4+2\delta$ moments. This does not hold for all GARCH processes, especially for those close to the region of I–GARCH. When $E(\eta_t^2) = \infty$, the asymptotic theory is invalid. Although an asymptotic theory has not been developed, it is a reasonable conjecture that if one could be found, it would involve convergence not to Brownian motions, but to Levy processes (continuous time martingales whose finite dimensional distributions are stable). The test statistics would not diverge under the null, but converge to an alternative distribution, and therefore tests using the critical values from table 1 would have misleading size. It is thus unclear how well the Chow–based test can discriminate between bi–integrated and GARCH processes with infinite unconditional fourth moments.

I applied these tests to the U.S.–Swiss exchange rate series discussed in Engle and Bollerslev (1986). The data are weekly observations on the differenced logarithm of the exchange rate between the U.S. and Switzerland from July 1973 through August 1985. Figure 3 displays the Chow test sequence for the hypothesis of constant variance. The sequence was calculated using a long–run variance estimate using a Bartlett kernel with four lags (M = 4 in (3)). Table 2 reports the test statistics, both of which are significant, suggesting that the variance is non–constant in the sample. Note that Engle–Bollerslev found that a fitted GARCH model was in the neighborhood of "integration" in the I–GARCH sense.

Table 2 : Tests for Constant Variance

MaxChow	11.43	(significant at 5% level)
MeanChow	2.32	(significant at 10% level)

4. Bi–integrated regression: Identification

This section will examine conditions for consistent estimation of a simple model of bi–integrated regression. The regression model is

(5) $$y_t = \mu + \alpha x_{2t} + x_{1t} , \quad t = 1, \dots , T .$$

The scalar variables y_t and x_{2t} are observed, and x_{1t} is an unobserved disturbance term. The variable x_{2t} may be endogenous, so some observable x_{3t} will be used as an instrument for x_{2t}. Least squares obtains as the special case $x_{3t} = x_{2t}$. The maintained assumption in this section is that regressor, instrumental variable and regression error are all bi–integrated processes. Since bi–integrated processes have non–stationary variances (unconditionally infinite), it is an interesting question how standard regression techniques will behave.

The variables (x_{1t}, x_{2t}, x_{3t}) are generated by

(6) $$x_{1t} = \sigma_{1t}u_{1t}, \qquad x_{2t} = \sigma_{2t}u_{2t} , \qquad x_{3t} = \sigma_{3t}u_{3t} .$$
$$\Delta\sigma_{1t} = v_{1t} , \qquad \Delta\sigma_{2t} = v_{2t} , \qquad \Delta\sigma_{3t} = v_{3t} .$$

Define the innovation vectors $v_t = (v_{1t}, v_{2t}, v_{3t})'$, $u_t = (u_{1t}, u_{2t}, u_{3t})'$, and $\eta_t = x_{3t}x_{1t} - E(x_{3t}x_{1t})$.

Assumption 2: $\{v_t, u_t\}$ is strictly stationary and satisfies

(i) $Ev_t = 0 \qquad Eu_t = 0$

(ii) $\{v_t, u_t\}$ has strong mixing coefficients α_m of size $-p/(p-2)$ for some $p > 2$,

(iii) $E|v_t'v_t|^p < \infty , \quad E|u_t'u_t|^{2p} < \infty ,$

(iv) $\lim_{T\to\infty} T^{-1}ES_TS_T' = \Omega , \quad S_t = \Sigma_{j=1}^{t}(v_j', u_j', \eta_j)' ,$

Assumption 2 is sufficient for the invariance principle

$$(7) \qquad T^{-1/2}S_{[Tr]} \Rightarrow B(r) \equiv BM(\Omega) .$$

Partition B and Ω as follows

$$B = (V', U', \eta)' , \qquad V = (V_1, V_2, V_3) , \qquad U = (U_1, U_2, U_3) .$$

$$\Omega = \begin{pmatrix} \Omega_{vv} & \Omega_{vu} & \Omega_{v\eta} \\ \Omega_{uv} & \Omega_{uu} & \Omega_{u\eta} \\ \Omega_{\eta v} & \Omega_{\eta u} & \Omega_{\eta\eta} \end{pmatrix} , \quad \Omega_v = [\omega_{vij}]_{ij} \quad \Omega_u = [\omega_{uij}]_{ij} .$$

To exclude degenerate cases we need $\omega_{vii} > 0$, $\omega_{uii} > 0$, $i = 1,2,3$. Define the covariance function for u:

$$\gamma_{ij} = E(u_{it}u_{jt}).$$

Equation (5) is estimated by instrumental variables, using x_{3t} as an instrument for x_{2t} (least squares is the special case $x_{3t} = x_{2t}$). We have the following condition for identification of bi–integrated regression coefficients.

Theorem 2. *The necessary and sufficient conditions for consistent estimation of* α *in (5) under assumption 2 are*

(i) $\gamma_{32} \neq 0$, (ii) $\gamma_{31} = 0$.

Condition (i) is analogous to the classic relevance condition, and is satisfied by the OLS estimator. Condition (ii) is analogous to the classic assumption of orthogonality between the instruments and the regression errors. Note, however, that the conditions for consistent estimation are completely independent of the properties of $\{\sigma_{1t}, \sigma_{2t}, \sigma_{3t}\}$, the "variance parts" of the regressors, errors, and instruments. Identification is obtained solely by the behavior of the process $\{u_t\}$.

A standard way of thinking about identification in instrumental regression is to view the IV estimator as indirect least squares (ILS) solved from OLS estimation of the reduced form. In this case the estimated reduced form equations are

$$\begin{aligned} x_{2t} &= \hat{\mu}_2 + \hat{\beta}_2 x_{3t} + \hat{w}_{2t} \\ y_t &= \hat{\mu}_y + \hat{\beta}_y x_{3t} + \hat{w}_{yt} \end{aligned}$$

so that

$$\hat{\alpha} = \hat{\beta}_y / \hat{\beta}_2 \ .$$

The relevance condition is normally thought of as holding when the first stage regression of the regressor x_2 on the instrument x_3 asymptotically yields a significant coefficient; i.e. if $\hat{\beta}_2 \rightarrow_p \beta_2 \neq 0$. This does not necessarily hold in the context of identified bi–integrated regression.

Theorem 3. *If assumption 2 and the identification conditions are satisfied,*

(a) $$\hat{\beta}_2 \Rightarrow \left[\int_0^1 V_3^2 \gamma_{33}\right]^{-1} \int_0^1 V_3 V_2 \gamma_{32}$$

(b) $$\hat{\beta}_y \Rightarrow \left[\int_0^1 V_3^2 \gamma_{33}\right]^{-1} \int_0^1 V_3 V_2 \gamma_{32} \alpha$$

(c) $$\hat{\alpha} \rightarrow_p \alpha \ .$$

The reduced form coefficients converge in probability to constants only if σ_{2t} *and* σ_{3t} *are cointegrated, i.e.*

(CI) $$V_2 = cV_3 \ ,$$

in which case

(d) $$\hat{\beta}_2 \rightarrow_p c\gamma_{32}/\gamma_{33} \ , \qquad \hat{\beta}_y \rightarrow_p \alpha\, c\gamma_{32}/\gamma_{33} \ .$$

Theorem 3 states that the reduced form coefficients converge in general to non–degenerate random variables, although the structural coefficient is consistently estimated. The only case in which the reduced form coefficients are consistent for the underlying parameters is when the variance processes of the regressors and the instruments are cointegrated, which obtains for OLS. Similar results were found by Phillips and Hansen (1990) in the context of IV estimation of cointegrating regressions among I(1) variables. In a study of identification of simultaneous equation models with identities, Brown (1985) found that structural identification does not necessarily require identification of the reduced form.

The next result gives the asymptotic distribution of the IV estimators.

<u>Theorem 4</u>. *If assumption 2 and the identification conditions are satisfied,*

(a) $$(\hat{\mu} - \mu) \Rightarrow \int V_1 dU_1 ,$$

(b) $$\sqrt{T}(\hat{\alpha} - \alpha) \Rightarrow \left[\int V_3 V_2 \gamma_{32}\right]^{-1}\left[\int V_3 V_1 d\eta + \int (V_1, V_3)\lambda_{v\eta}\right]$$

where $$\lambda_{v\eta} = \Sigma_{j=1}^{\infty} E\left(\begin{bmatrix} v_{11} \\ v_{31} \end{bmatrix} \eta_j\right) .$$

Theorem 4 points out that the constant term μ is not estimated consistently by linear techniques. This is equivalent to the failure of the sample mean to consistently estimate the population mean of a bi–integrated process in theorem 1a.

5. Bi–integrated Regression: Inference

Theorem 4 in section 4 gave the asymptotic distribution of the instrumental variables estimator of the regression coefficients in a bi–integrated regression. The asymptotic representation in that theorem is not particularly useful for inferential purposes. To keep the attention concentrated, we will restrict attention in this section to identified least squares estimation. Theorem 4 simplifies in the OLS case to

$$\sqrt{T}(\hat{\alpha} - \alpha) \quad \Rightarrow \quad \left[\int V_2^2 \gamma_{22}\right]^{-1} \left[\int V_2 V_1 d\eta \ + \ \int (V_1, V_2) \lambda_{v\eta}\right] \quad = \ f_\alpha \text{ , say.}$$

Here, $\lambda_{v\eta} \ = \ \Sigma_{j=1}^{\infty} E\left[\begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} \eta_j\right]$, and $\eta_j = u_{2j} u_{1j}$.

This distribution is not generically a mixture of normals. To facilitate inference we add an assumption.

Assumption 3. For all t, j , $E[\ u_{1t} u_{2t} \mid v_{t+j}\ ] \ = 0$.

Theorem 5. *Under assumption* 3, $\lambda_{v\eta} = 0$, $\Omega_{v\eta} = 0$, *and*

$$f_\alpha \ \equiv \ S^{1/2} N ,$$

where $N \ \equiv \ N(0,\ 1)$ *is independent of* $S \ \equiv \ \left[\int V_2^2 \gamma_{22}\right]^{-1} \left[\int V_2^2 V_1^2 \omega_{\eta\eta}\right] \left[\int V_2^2 \gamma_{22}\right]^{-1}$.

Theorem 5 shows that assumption 3 allows $\hat{\alpha}$ to have an asymptotic mixture of normals distribution. The limiting variance, however, is of the form which requires a robust estimator. Consider

$$\hat{S} \ = \ \hat{S}_{11}^{-1} \hat{S}_{12} \hat{S}_{11}^{-1} , \quad \hat{S}_{11} \ = \ T^{-2} \Sigma_1^T (x_{2t} - \bar{X}_2)^2 ,$$

$$\hat{S}_{12} \ = \ T^{-3} \sum_{m=-M}^{M} k_m \sum_{t=m+1}^{M} x_{2t} x_{2t-m} \hat{x}_{1t} \hat{x}_{1t-m}$$

where k_m is a kernel function as discussed earlier, and $\hat{x}_{1t}$ are the OLS residuals.

Theorem 6 *Under assumptions* 2 *and* 3 , and $M = o(T^{1/4})$,

(a) $\hat{S}_{12} \Rightarrow \int V_2^2 V_1^1 \omega_{\eta\eta}$,

(b) $\sqrt{T/\hat{S}}\ (\hat{\alpha} - \alpha) \rightarrow_d N(0, 1)$.

Theorem 6 suggests that standard inferential methods can proceed conventionally even in bi–integrated regressions if robust variance estimates are used.

6. Conclusion

This paper studied the properties of ordinary least squares and instrumental variables regression in models with non–stationary variances. Part of the motivation for a study of this nature is the empirical observation that economic and financial data seem to display non–constancies in the variance. Although the model used in this paper – bi–integration – is distinct from the popular GARCH model, the insights found here can be interpreted as suggestive for a broader class of models with time–varying variances. Identification using instrumental variables is possible, but in the unconventional sense that reduced form parameters need not be consistently estimated in order to achieve consistent estimation of the structural parameters. Inference requires variance estimates which are robust to heteroskedasticity and serial correlation, as is frequently done in practice. Kim and Schmidt (1989) found in a Monte Carlo study of unit root tests in the presence of GARCH errors that the White heteroskedasticity–robust variance estimate improved the behavior of their test statistics. This study suggests that such results may be generic.

Appendix

The proofs will make repeated use of three lemmas proved in Hansen (1989). For reference, they are reported below as lemma A.1, A.2, and A.3.

Lemma A.1. *Under assumption 2, for any strictly stationary, ergodic and square integrable* $\{e_t\}$ *, and for* $i = 1,2,3,\ j = 1,2,3$:

$$T^{-2}\Sigma_1^T \sigma_{it}\sigma_{jt}e_t \ \Rightarrow\ \int_0^1 V_i V_j E(e_t)\ .$$

Lemma A.2. *Under assumption 2, and for* $i = 1,2,3,\ j = 1,2,3$:

$$T^{-3/2}\Sigma_1^T \sigma_{it}\sigma_{jt}\eta_t \ \Rightarrow\ \int_0^1 V_i V_j d\eta \ +\ \int_0^1 V_i \Sigma_{k=1}^{\infty} E(v_{j1}\eta_k) \ +\ \int_0^1 V_j \Sigma_{k=1}^{\infty} E(v_{i1}\eta_k).$$

Lemma A.3 *Under assumption 2,* $M = o(T^{1/2})$ *, and if*

$$T^{-1}\sum_{m=-M}^{M} k_m \sum_{t=m+1}^{M} \eta_t\eta_{t-m} \ \rightarrow_p\ \omega_{\eta\eta}$$

then

$$T^{-3}\sum_{m=-M}^{M} k_m \sum_{t=m+1}^{M} \sigma_{2t}^2\sigma_{1t}^2\eta_t\eta_{t-m} \ \Rightarrow\ \int_0^1 V_2^2 V_1^2\ \omega_{\eta\eta}\ .$$

Proof of Theorem 1. Part (a) is given in Phillips (1988). For part (b) note that

$$T^{-1}\hat{\sigma}^2 \ =\ T^{-2}\Sigma_1^T x_t^2 \ -\ T^{-1}\bar{X}^2 \ =\ T^{-2}\Sigma_1^T \sigma_t^2\epsilon_t^2 \ +\ O_p(T^{-1})$$

$$\Rightarrow\ \int_0^1 B_2^2\ E(\epsilon_t^2) \ =\ \int_0^1 B_2^2$$

using lemma A.1. For part (c) note that

$$T^{-2}\Sigma_t(x_t-\bar{X})(x_{t-k}-\bar{X}) \ =\ T^{-2}\Sigma_t\sigma_t\sigma_{t-k}\epsilon_t\epsilon_{t-k} \ +\ O_p(T^{-1})$$

$$=\ T^{-2}\Sigma_t\sigma_t\sigma_t\epsilon_t\epsilon_{t-k} \ +\ o_p(1) \ \Rightarrow\ \int_0^1 B_2^2 E(\epsilon_t\epsilon_{t-k})$$

again by lemma A.1. Combining with result (b) we find

$$(\hat{\sigma}^2 T)^{-1} \sum_{t=k+1}^{T} x_t x_{t-k} \Rightarrow \left[\int_0^1 B_2^2\right]^{-1} \int_0^1 B_2^2 E(\epsilon_k \epsilon_0) = E(\epsilon_k \epsilon_0) \quad \text{a.s.}$$

since $\int_0^1 B_2^2 > 0$ a.s. Weak convergence to a constant is convergence in probability. To show part (d), we combine parts (a) and (b) to find

$$\sqrt{T}\,\frac{(\bar{X} - \lambda_{21})}{\hat{\sigma}} = \left[T^{-1}\hat{\sigma}^2\right]^{-1/2}\left[\bar{X} - \lambda_{21}\right] \Rightarrow \left[\int_0^1 B_2^2\right]^{-1/2} \int_0^1 B_2 dB_1 .$$

Decompose B_1 as $B_1 = \sqrt{\omega_{11}}\,\left[(1-\rho^2)^{1/2} W_1 + \rho W_2\right]$, where $W_2 = \omega_{22}^{-1/2} B_2$ is independent of $W_1 \equiv BM(1)$. This random variable equals

$$\left[\left[\int_0^1 B_2^2\right]^{-1/2} \int_0^1 B_2 dW_1 (1-\rho^2)^{1/2} + \left[\int_0^1 B_2^2\right]^{-1/2} \int_0^1 B_2 dW_2 \rho\right] \sqrt{\omega_{11}}$$

$$= \left[\, N(0,1)(1-\rho^2)^{1/2} + \left[\int_0^1 W_2^2\right]^{-1/2} \int_0^1 W_2 dW_2 \rho \,\right] \sqrt{\omega_{11}} . \qquad \square$$

<u>Proof of theorem 2</u>. We first prove sufficiency. Under assumption 2,

$$\text{(A1)} \qquad \begin{bmatrix} \hat{\mu} - \mu \\ \hat{\alpha} - \alpha \end{bmatrix} = \begin{bmatrix} T & \Sigma_1^T x_{2t} \\ \Sigma_1^T x_{3t} & \Sigma_1^T x_{3t} x_{2t} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_1^T x_{1t} \\ \Sigma_1^T x_{3t} x_{1t} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & T^{-1}\Sigma_t x_{2t} \\ T^{-2}\Sigma_t x_{3t} & T^{-2}\Sigma_t x_{3t} x_{2t} \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\Sigma_1^T x_{1t} \\ T^{-2}\Sigma_1^T x_{3t} x_{1t} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & \int V_2 dU_2 \\ 0 & \int V_3 V_2 \gamma_{32} \end{bmatrix}^{-1} \begin{bmatrix} \int V_1 dU_1 \\ \int V_3 V_1 \gamma_{31} \end{bmatrix} = \begin{bmatrix} 1 & \int V_2 dU_2 \\ 0 & \int V_3 V_2 \gamma_{32} \end{bmatrix}^{-1} \begin{bmatrix} \int V_1 dU_1 \\ 0 \end{bmatrix} .$$

Since $\int V_3 V_2 \neq 0$ a.s and $\gamma_{32} \neq 0$, $(\hat{\alpha} - \alpha) \rightarrow_p 0$.

We now show necessity. If $\gamma_{32} \neq 0$ and $\gamma_{31} \neq 0$, then (A1) shows that $(\hat{\alpha} - \alpha)$ converges to a non-degenerate random variable, so $\hat{\alpha}$ is inconsistent. If

$\gamma_{32} = 0$ and $\gamma_{31} = 0$, then

(A2) $$\begin{bmatrix} \hat{\mu} - \mu \\ \hat{\alpha} - \alpha \end{bmatrix} = \begin{bmatrix} 1 & T^{-1}\Sigma_t x_{2t} \\ T^{-3/2}\Sigma_t x_{3t} & T^{-3/2}\Sigma_t x_{3t}x_{2t} \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\Sigma_1^T x_{1t} \\ T^{-3/2}\Sigma_1^T x_{3t}x_{1t} \end{bmatrix}.$$

Now $T^{-3/2}\Sigma_t x_{3t}x_{2t} = T^{-3/2}\Sigma_t \sigma_{3t}\sigma_{2t}\eta_{32t}$, where $\eta_{32t} = u_{3t}u_{2t}$ is a zero–mean square integrable stationary random variable by assumptions 2 and 3 and the fact that $\gamma_{32} = 0$. By lemma A.2, this sum converges to a limit random variable which is non–zero almost surely. Similarly, $T^{-3/2}\Sigma_t x_{3t}x_{2t}$ converges to an almost surely non–zero random variable. Thus (A2) converges weakly to an almost surely non–zero random variable, and $\hat{\alpha}$ is inconsistent. □

<u>Proof of theorem 3</u>. Conventional formulae yield

$$\hat{\beta}_2 = \left[T^{-2}\Sigma_1^T(x_{3t}-\bar{x}_3)^2\right]^{-1}\left[T^{-2}\Sigma_1^T(x_{3t}-\bar{x}_3)x_{2t}\right]$$

and

$$\hat{\beta}_y = \left[\Sigma_1^T(x_{3t}-\bar{x}_3)^2\right]^{-1}\left[\Sigma_1^T(x_{3t}-\bar{x}_3)y_t\right]$$

$$= \left[T^{-2}\Sigma_1^T(x_{3t}-\bar{x}_3)^2\right]^{-1}\left[T^{-2}\Sigma_1^T(x_{3t}-\bar{x}_3)x_{2t}\alpha + T^{-2}\Sigma_1^T(x_{3t}-\bar{x}_3)x_{1t}\right],$$

where $\bar{x}_3 = T^{-1}\Sigma_1^T x_{3t} = O_p(1)$. Note that

(A3) $$T^{-2}\Sigma_1^T(x_{3t}-\bar{x}_3)^2 = T^{-2}\Sigma_1^T x_{3t}^2 - T^{-1}\left[T^{-1}\Sigma_1^T x_{3t}\right]^2$$

$$= T^{-2}\Sigma_1^T \sigma_{3t}^2 u_{3t}^2 + O_p(T^{-1}) \Rightarrow \int V_3^2\gamma_{33} > 0 \text{ a.s.}$$

using lemma A.1. Similarly, we find that

(A4) $$T^{-2}\Sigma_1^T(x_{3t}-\bar{x}_3)x_{2t} \Rightarrow \int V_3V_2\gamma_{32}$$

(A5) $$T^{-2}\Sigma_1^T(x_{3t}-\bar{x}_3)x_{1t} \Rightarrow \int V_3V_1\gamma_{31} = 0$$

under the identification condition. The stated results follow by the application of the continuous mapping theorem to (A3), (A4) and (A5). □

Proof of theorem 4. Under assumption 2 and the identification condition

$$\begin{bmatrix} (\hat{\mu} - \mu) \\ \sqrt{T}(\alpha - \alpha) \end{bmatrix} = \begin{bmatrix} 1 & T^{-3/2}\Sigma_t x_{2t} \\ T^{-3/2}\Sigma_t x_{3t} & T^{-2}\Sigma_t x_{3t}x_{2t} \end{bmatrix}^{-1} \begin{bmatrix} T^{-1}\Sigma_t x_{1t} \\ T^{-3/2}\Sigma_t x_{3t}x_{1t} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & 0 \\ 0 & \int V_3V_2\gamma_{32} \end{bmatrix}^{-1} \begin{bmatrix} \int V_1 dU_1 \\ \int V_3V_1 d\eta + \int (V_3, V_1)\lambda_{v11} \end{bmatrix}$$

where the final convergence applies lemma A.2. □

Proof of theorem 5. By assumption 3, for all j, $E(v_1\eta_j) = E[v_1 E(u_{1t}u_{2t}|v_1)] = 0$, thus

$$\lambda_{v\eta} = \Sigma_{j=1}^{\infty} E\left[\begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}\eta_j\right] = 0 ,$$

and $f_\alpha = \left[\int V_2^2\gamma_{22}\right]^{-1}\int V_2V_1 d\eta$. Similarly

$$\Omega_{v\eta} = \Sigma_{j=-\infty}^{\infty} E\left[\begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}\eta_j\right] = 0 ,$$

which implies that η is independent of $V = (V_1\ V_2)'$. Define the σ–field $\mathscr{F}_v = \sigma(V(r)\ ;\ 0 \le r \le 1)$. Then conditional on $\mathscr{F}_v$, f_α is distributed

$$N(0, S) = S^{1/2}N \qquad S \equiv \left[\int V_2^2\gamma_{22}\right]^{-1}\left[\int V_2^2V_1^2\omega_{\eta\eta}\right]\left[\int V_2^2\gamma_{22}\right]^{-1}$$

where $N \equiv N(0,1)$ does not depend upon S, and is therefore independent of S. □

Proof of theorem 6. We have

$$\hat{S}_{12} = T^{-3}\sum_{m=-M}^{M} k_m \sum_{t=m+1}^{M} x_{2t}x_{2t-m}\hat{x}_{1t}\hat{x}_{1t-m}$$

$$= T^{-3}\sum_{m=-M}^{M} k_m \Sigma_t x_{2t}x_{2t-m}\hat{x}_{1t}x_{1t-m}$$

$$- \; T^{-3} \sum_{m=-M}^{M} k_m \sum_t x_{2t} x_{2t-m} \hat{x}_{1t} [x_{2t}(\hat{\alpha} - \alpha) + (\hat{\mu} - \mu)]$$

$$= T^{-3} \sum_{m=-M}^{M} k_m \sum_t x_{2t} x_{2t-m} \hat{x}_{1t} x_{1t-m} + o_p(1)$$

$$= T^{-3} \sum_{m=-M}^{M} k_m \sum_t x_{2t} x_{2t-m} x_{1t} x_{1t-m} + o_p(1)$$

$$= T^{-3} \sum_{m=-M}^{M} k_m \sum_t \sigma_{2t} \sigma_{2t} \sigma_{1t} \sigma_{1t} \eta_t \eta_{t-m} + o_p(1)$$

$$\Longrightarrow \int_0^1 V_2^2 V_1^2 \, \omega_\eta$$

by lemma A.3, and the fact that our assumptions give

$$T^{-1} \sum_{m=-M}^{M} k_m \sum_t \eta_t \eta_{t-m} \longrightarrow_p \omega_\eta$$

(see Newey and West (1987)). This gives (a).

By the continuous mapping theorem,

$$\hat{S} \Longrightarrow S \equiv \left[\int V_2^2 \gamma_{22}\right]^{-1} \left[\int V_2^2 V_1^2 \omega_{\eta\eta}\right] \left[\int V_2^2 \gamma_{22}\right]^{-1} .$$

Assumptions 2 and 3 guarantee by theorems 4 and 5 that

$$\sqrt{T}(\hat{\alpha} - \alpha) \Longrightarrow S^{1/2} N$$

thus

$$\sqrt{T/\hat{S}} \; (\hat{\alpha} - \alpha) \Longrightarrow S^{-1/2} S^{1/2} N = N \equiv N(0, 1).$$

Weak convergence in $\mathbb{R}$ is convergence in distribution, completing the proof of part (b). $\square$

References

Andrews, D. W. K. (1989a): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Cowles Foundation Discussion Paper No. 877–R.

________________ (1989b): "Tests of parameter instability and structural change with unknown change point," manuscript, Yale University.

Bollerslev, T. and R.F. Engle (1989): "Common persistence in conditional variances," unpublished manuscript, Northwestern University.

Brown, B.W. (1985): "The identification problem in simultaneous equation models with identities," *International Economic Review*, 26, 45–66.

Chu, C–S. J. (1989): "New tests for parameter constancy in stationary and nonstationary regression models," manuscript, UCSD.

Engle, R. F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation," *Econometrica*, 50, 987–1008.

Engle, R. F. and T. Bollerslev (1986): "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

Engle, R. F. and C.W.J. Granger (1987): "Co–integration and error–correction: representation, estimation and testing," *Econometrica*, 55, 251–276.

Granger, C.W.J., and A.P. Andersen (1978): *An Introduction to Bilinear Time Series Models*, Gottingen: Vandenhock and Ruprecht.

Hansen, B.E. (1989): "Bi–integration and heteroskedastic cointegration", manuscript, University of Rochester.

________________ (1990): "Testing for structural change of unknown form in models with non–stationary regressors," manuscript, University of Rochester.

Kim, K. and P. Schmidt (1989): "Unit root tests with conditional heteroskedasticity," manuscript, Michigan State University.

Nelson, Daniel B. (1989a): "Stationarity and persistence in the GARCH(1,1) model," manuscript, Univ. of Chicago Business School.

________________, (1989b): "Conditional heteroskedasticity in asset returns: a new approach," manuscript, Univ. of Chicago Business School.

Newey, W.H. and K.D. West (1987): "A simple, positive semi–definite, heteroskedastiticy and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Pagan, A.R. and G.W. Schwert (1989): "Alternative models for conditional stock volatility," manuscript, University of Rochester.

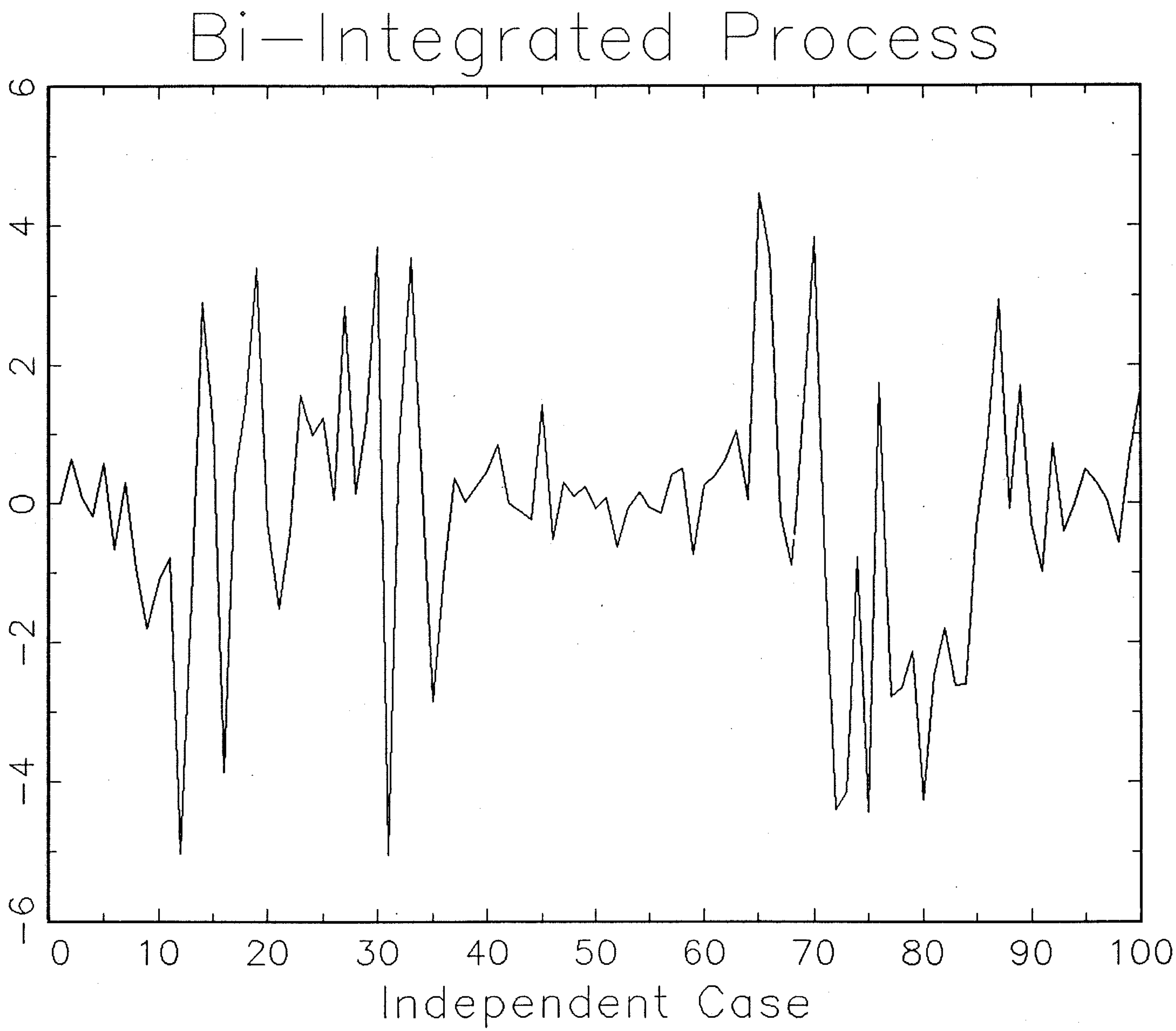
Park, J.Y. and P.C.B. Phillips (1988): "Statistical inference in regressions with integrated processes," *Econometric Theory*, 4, 468–498.

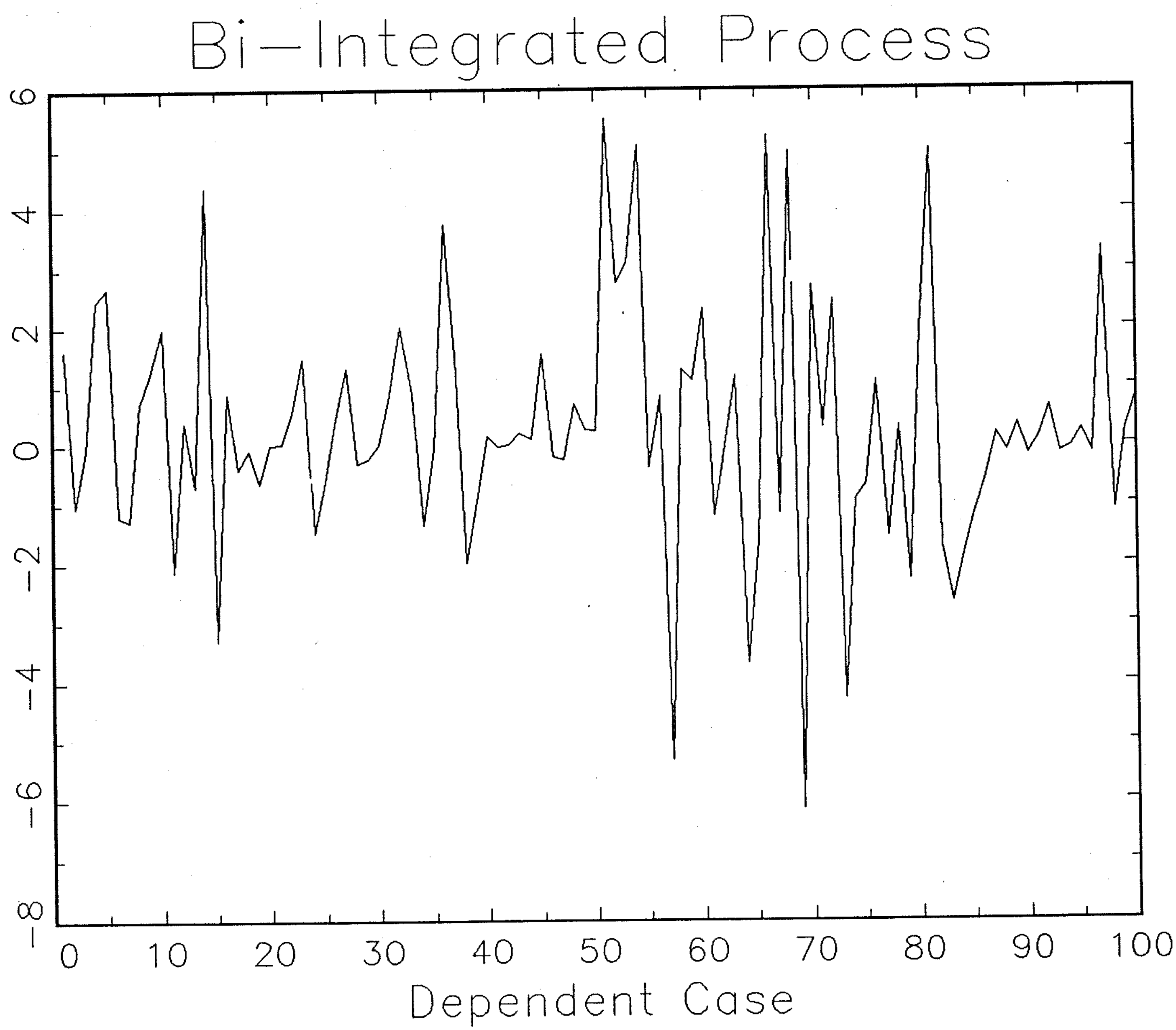
Phillips, P.C.B. (1987): "Time series regression with a unit root," *Econometrica*, 55, 277–301.

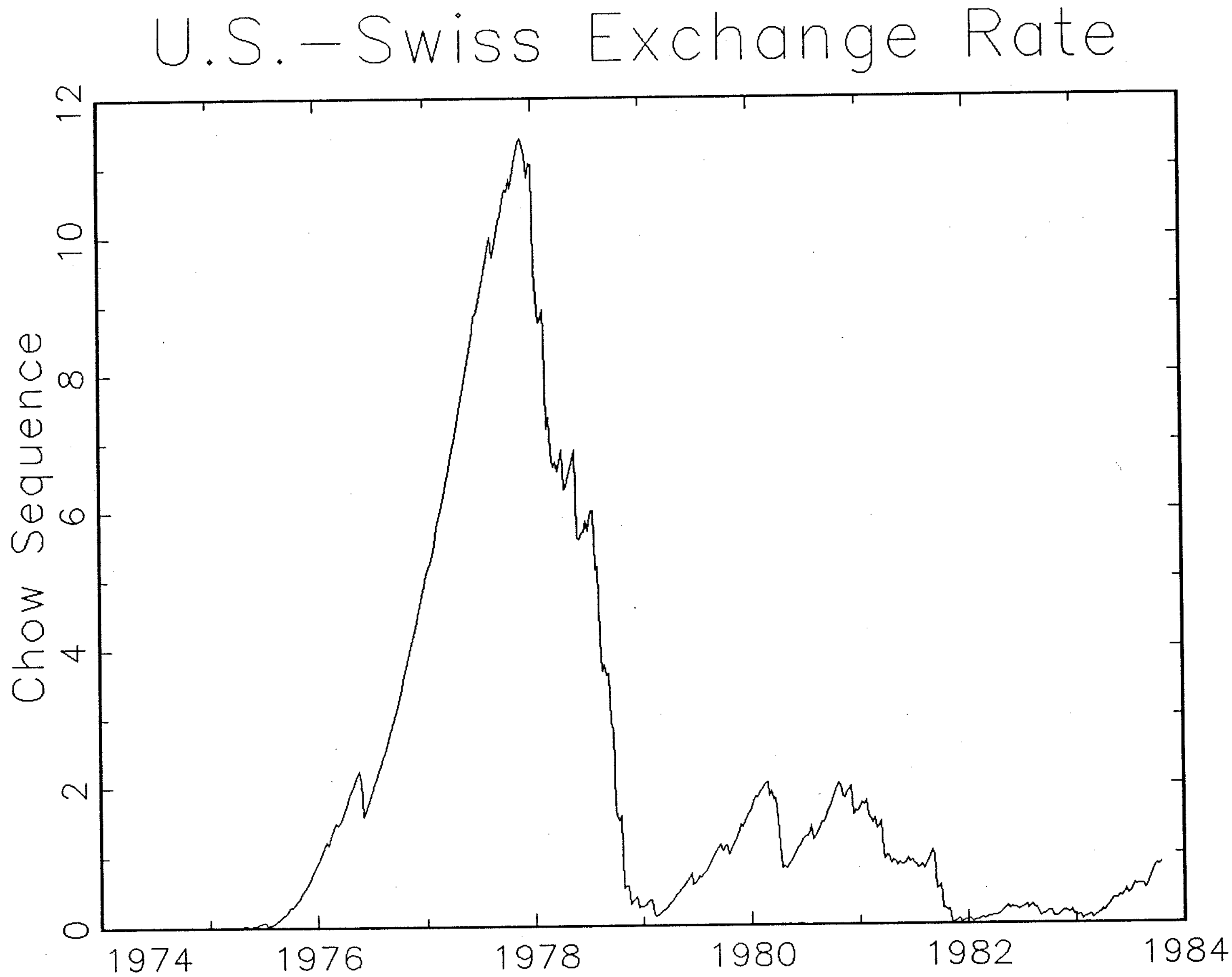
________________, (1988): "Weak convergence of to the matrix stochastic integral $\int_0^1 BdB'$," *Journal of Multivariate Analysis*, 24, 252–264.

Phillips, P.C.B. and B.E. Hansen (1990): "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, 57, 473–496.

White, H. (1984): *Asymptotic theory for econometricians*, Academic Press.


Bi-Integrated Process
6
4
2
0
-2
-4
-6
0
10
20
30
40
50
60
70
80
90
100
Independent Case


Bi-Integrated Process
6
4
2
0
-2
-4
-6
-8
0
10
20
30
40
50
60
70
80
90
100
Dependent Case


U.S.-Swiss Exchange Rate
Chow Sequence
0
2
4
6
8
10
12
1974
1976
1978
1980
1982
1984