1987

The Econometric Analysis of Models with Risk Terms

A. R. Pagan

Aman Ullah

ISSN: 0831-4454
ISBN: 0-7714-0921-4

THE CENTRE FOR DECISION SCIENCES AND ECONOMETRICS

The Econometric Analysis of Models with Risk Terms

Adrian Pagan
and
Aman Ullah

TECHNICAL REPORT NO. 16
JULY 1987

Centre For Decision Sciences And Econometrics
Social Science Centre
The University of Western Ontario
London, Ontario N6A 5C2

The Econometric Analysis of Models with Risk Terms*

Adrian Pagan
The University of Rochester, Rochester, N.Y., U.S.A. 14627

Aman Ullah
The University of Western Ontario, London, Ontario, Canada

April, 1986, Revised February, 1987.

*We are grateful to Hashem Pesaran for his comments upon this paper.

1. Introduction

That the world inhabited by economic agents is a risky one has long been accepted and emphasized by economic theorists. In quantitative models, however, much less importance has traditionally been accorded to such effects. The rational expectations revolution in econometric analysis was largely concerned with how to model the mean of a random variable conditional upon an information set, and only minimal attention was paid to second and higher moments of the corresponding conditional distribution. Even in those instances where higher order moments entered the analysis, as in the Capital Asset Pricing Model, portfolio models, or the first order conditions from many "Euler equation" models, they were typically assumed constant, and so the effects of risk factors became absorbed into the parameter set. Frankel (19S5) is a good illustration of this point.

Although the above seems a fair description of much research, events in the 1970's conspired to interest the applied economist in the difficulties of accounting for changing risk. The initial impetus seems to have come from a rise in inflation rates, and a group of studies was spawned that sought to examine the economic effects of greater unpredictability in either the levels of inflation or relative prices, for example Klein (1977) and Hercowitz (1981). Furthermore, as the 1970's wore on, freer exchange rates and much more flexible monetary policies in many economies meant that some allowance might need to be made for what was perceived as an increasingly risky environment.

There now exist a fair number of studies attempting to allow for a changing risk term in economic models; a very small sample would have to mention Vanderhoff (1983), Mascaro and Meltzer (1983), Lawrence (1983), Engle (1983) and Gylfasson (1981). It is noticeable, however, that there are few theoretical papers on which econometric methods should be employed in this context, and there has been considerable diversity in the way in which each of the researchers has approached the problem. In an earlier paper, Pagan (1984), one of us briefly looked at the area, concluding that at least one of the popular methods was unlikely to be satisfactory, and suggesting some alternatives. However, the treatment was not very comprehensive, and it therefore seems appropriate to re-examine the issues here in somewhat greater depth than before.

Section 2 of this paper considers the estimation of a linear model containing a term representing the risk originating from a failure to perfectly predict some variable. Theoretical models show that this risk is related to the moments of the probability density function of the variable conditioned upon whatever information agents use in their optimization. For expository purposes we assume that the conditional variance is the appropriate moment, propose that it be replaced by functions of the observed data, and then recommend estimation of unknown parameters by instrumental variables, instruments being constructed from the information set. Conditions are set out under which the proposed estimator is consistent and in which correct inferences can be made. Section 3 of the paper looks at issues arising when the level of risk is to be explained rather than when it is an explanator.

Within the literature a variety of proxies for risk can be found. Section 4 looks at these to see how satisfactory each method is when used as a measure of risk in a modeling environment. In general the proxies suffer from a variant of the "errors in variables" problem, and therefore should only be used in conjunction with an instrumental variables estimator; if substituted directly into a regression an underestimate of the effect of risk on decisions is likely. An exception to this rule occurs if a parametric model for the risk term is adopted, but this methodology has its own problems involving potential mis-specification, which can be partially alleviated by following our instrumental variables approach. Finally, section 5 applies some of the ideas in the paper.

2. Risk as a Regressor

2.1 An Instrumental Variables Estimator

Many macro-economic models are formulated by having agents optimally choose a variable y_t on the basis of some information set $\mathcal{F}_t$. It is assumed here that this optimization yields a linear decision rule of the form

$$E(y_t|\mathcal{F}_t) = \bar{x}_t\gamma + \sigma_t^2\delta \qquad t=1,2,\ldots,T \qquad (1)$$

where $\bar{x}_t \in \mathcal{F}_t$ is a (1xs) vector and σ_t^2 is the variance of some variable ψ_t conditional upon $\mathcal{F}_t$. σ_t^2 represents a risk term arising from the failure of agents to be able to correctly predict the variable ψ_t. A model such as (1) can arise in a number of ways. Standard mean/variance analysis produces it, but there exist more general models in which σ_t^2 can be taken as a linear approximation to the expectation of a non-linear function of a random variable, and this is then interpreted as risk, e.g. Leland (1961) for consumption. Each of $\mathcal{F}_t$, Ψ_t, and $\bar{x}_t$ need to be defined by the theoretical context; in Stockman (1978) for example, Ψ_t is a vector of variables involving the domestic and foreign money supplies and the real rate of return, while $\mathcal{F}_t$ will be whatever information an agent is assumed to possess in the optimizing exercise. It is convenient to treat σ_t^2 as a scalar, as the extension to the vector case is obvious. Moreover, although it will be the second moment σ_t^2 which is the focus of this paper, the methods advanced clearly extend to any moment of the density of ψ_t conditional upon $\mathcal{F}_t$, or even the covariance between two variables (say) w_{1t} amd w_{2t}.

(1) may be converted to the form

$$y_t = \bar{x}_t\gamma + \sigma_t^2\delta + e_t \qquad (2)$$

where e_t is an error term with the property $E(e_t|\mathcal{F}_t) = 0$. The estimation problem is that no series on σ_t^2 exists, and theory does not describe exactly how σ_t^2 varies with $\mathcal{F}_t$. In this paper we will address the various solutions that have been proposed to this dilemma. First, however, let us suppose that it was possible to construct a series ϕ_t such that ϕ_t^2 was a function of some index N (that does not depend on t) and that $\phi_t^2(N) \xrightarrow{a.s.} \sigma_t^2$ $\forall t$ as $N \to \infty$. Such a series will be said to possess a "strong property", since for large enough N it is possible to regard ϕ_t^2 as σ_t^2. By contrast it may be that the sole series ϕ_t which is available possesses only a "weak property", $E(\phi_t^2|\mathcal{F}_t) = \sigma_t^2$.

As noted above, if ϕ_t^2 had the strong property it would be sensible to replace σ_t^2 in (2) with ϕ_t^2 and to regress y_t against x_t and ϕ_t^2 to obtain an estimate of δ. It is tempting to do the same thing when it has the weak property, but then the model (2) becomes

$$y_t = \bar{x}_t\gamma + \phi_t^2\delta + \epsilon_t = \bar{x}_t\gamma + \phi_t^2\delta + e_t + \delta(\sigma_t^2 - \phi_t^2). \qquad (3)$$

Although $E(\bar{x}_t e_t) = 0$ and it may not be unreasonable to assert that $E(\phi_t^2 e_t) = 0$, it is clear that $E(\phi_t^2\epsilon_t)$ will not be zero as it involves $\delta(E(\sigma_t^4) - E(\phi_t^4))$, which is only zero in degenerate cases. When ϕ_t is normal and $\mathcal{F}_t$ consists of non-stochastic elements, $\delta(E(\sigma_t^4) - E(\phi_t^4))$ is $-2\sigma_t^4$. Pagan (1984, p234) notes that the size of the inconsistency in the OLS estimator of δ can therefore be quite large. It should be observed that no such problems arise if ϕ_t had the strong property since $\delta(\sigma_t^2 - \phi_t^2) \xrightarrow{a.s.} 0$ as $N \longrightarrow \infty$.

Section 4 of the paper discusses the selections of ϕ_t that have been made in the literature. With one exception the proxies for σ_t^2 that have been suggested possess only a weak property (at best), making their use in a regression context suspect. But there is a way out of the difficulty. As the analysis above shows, the inconsistency in the OLS estimator arises from the fact that the "true" regressor σ_t^2 is observed with error, and so OLS suffers from an errors in variables bias. A standard solution to this is to estimate (3) not by OLS but by instrumental variables (IV). From the orthogonality condition $E(e_t+\delta(\sigma_t^2 - \phi_t^2)|\mathcal{F}_t) = 0$, we assume that there exist $p = s+1$

instruments z_t constructed from $\mathcal{F}_t$.[1] Having done so restrictions need to be placed upon the nature of $x_t=(\bar{x}_t \ \phi_t^2)$, z_t and ϵ_t to get desirable asymptotic properties for the IV estimator. It is convenient here to use the terminology and results from the asymptotic theory of mixing processes and we extensively utilize results from White (1984). Definitions and explanations of the terms can be found there.

ASSUMPTIONS

(i) $\{(z_t, x_t, \epsilon_t)\}$ is either a phi (or alpha) mixing sequence with mixing coefficients of size $r'/(r'-1)$, $r' > 1$ (or of size $r'/(r'-1)$, $r' > 1$), where $r' = r+a$ for some $r \geq 1$ and $0 < a \leq r$;

(ii) (a) $E(z_t\epsilon_t) = 0$

(b) $E|z_{ti}\epsilon_t|^{2r'} < \Delta < \infty$ for $r' > 1$, $i = 1, \ldots, p$ and all t

(c) $V_{bT} = \text{var}(T^{-1/2}\sum_{t=b+1}^{b+T} z_t\epsilon_t)$; $V_T = V_{0T}$ and $\exists$ V finite and positive definite such that $V_{bT} - V \rightarrow 0$ as $T \rightarrow \infty$ uniformly in b;

(iii) (a) $E|z_{ti}x_{tj}|^{r'} < \Delta_1 < \infty$ for $r' > 1$ and all $i=1, \ldots, p$; $j=1, \ldots, p$; $t=1, \ldots, T$;

(b) $Q_T = E(Z'X/T)$ has uniformly full column rank

(c) $\hat{P}_T - P_T \xrightarrow{p} 0$ where $\{P_T\}$ represents a p x p weighting matrix for the p instruments and is $O(1)$ and uniformly positive definite.

[1] If more instruments are available than p, generalized instrumental variable (Sargan (1958)) or GMM estimators (Hansen (1982)) will need to replace the simple IV procedure adopted here.

We have chosen to state the conditions in terms of the joint vector (z_t, x_t, ϵ_t) rather than (z_t, x_t, e_t), as it saves a good deal of notation. However, since $\phi_t^2 - \sigma_t^2$ will also mix as in (i), and ϵ_t is the sum of e_t and $\delta(\phi_t^2 - \sigma_t^2)$, it is clear that no generality is lost in doing this.

We can immediately provide the following proposition.

Proposition 1: If (3) is written as $y_t = x_t\beta + \epsilon_t$ or $y = X\beta + \epsilon$, and $p=s+1$, under assumptions (i) - (iii) the instrumental variable estimator of $\beta, \tilde{\beta} = (Z'X)^{-1}Z'y$, is a consistent estimator of β, while $D_T^{-1/2}T^{1/2}(\tilde{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, I_p)$, where $D_T = Q_T' V_T Q_T$.

Proof: Theorem 5-22 of White (1984, p126) with $P_T = (Q_T')^{-1}$.

□

How restrictive is Proposition 1? Loosely, the assumptions restrict the processes x_t, ϕ_t^2 and e_t to be stationary; some non-stationarity in the second moment is possible but it is severely circumscribed. Now, since $\sigma_t^2 = E((\psi_t - E((\psi_t|\mathcal{F}_t)^2|\mathcal{F}_t)$, it is clear that $\phi_t = \psi_t - E(\psi_t|\mathcal{F}_t)$, that is it is the unanticipated part of ψ_t. Since we would expect that ψ_t and $E(\psi_t|\mathcal{F}_t)$ would be co-integrated, using Engle and Granger's (1987) term, the unanticipated quantity should therefore be a stationary process. It also seems reasonable that the correlation of ϕ_t with values in the past should die out fairly quickly the further back one goes.

However, we cannot be so sanguine about other variables that drive the conditional mean from any optimizing exercise resulting in (1). As observed earlier, Stockman's (1978) model has y_t being the spot rate and $\bar{x}_t$ the forward rate, and both of these processes are commonly regarded as ARIMA rather than ARMA processes. A variety of responses can be made to this challenge. First, the optimizing theory may suggest that the non-stationary variables are co-integrated, i.e. there exists a linear combination of them which is stationary. That is the case in Stockman's formulation; the coefficient γ on the forward rate is unity so that (2) could be written as $y_t - \bar{x}_t = \sigma_t^2\delta + e_t$; a simple re-definition of y_t as $y_t - \bar{x}_t$ then allows the assumptions to be invoked. More generally, y_t can be thought of as a linear combination of all the co-integrated variables in the optimizing exercise, while $\bar{x}_t$ contains only those members of $\mathcal{F}_t$ that are stationary. When the co-integrating vector is not known, it can be estimated by regressing the co-integrated variables upon one another, with the residuals from this regression becoming y_t - this is the "two-step" procedure in Engle and Granger (1987) and the reader is referred there to the proof that replacing the co-integrating parameters by estimates does not affect the limiting distribution of the estimators of parameters associated with stationary variables.

However, it is not in fact necessary to regress out all the non-stationary variables. Suppose that (3) is re-parameterized so that $\bar{x}_t$ is separated into $\bar{x}_{1t}$ (non-stationary) and $\bar{x}_{2t}$ (stationary) variables. The γ vector is partitioned conformably into $\gamma' = (\gamma_1' \; \gamma_2')$, while $\beta_2' = (\gamma_2' \; \delta)$ and $\beta_1 = \gamma_1$. Let $\bar{x}_{1t}$ be its own instrument and denote the instruments for $\bar{x}_{2t}$ and ϕ_t^2 as z_{1t}. Then, if $\bar{x}_{1t}$ is ARIMA (0,g,0) while $(\bar{x}_{2t}, z_{1t}, \phi_t^2)$ are strictly stationary processes, Lemma 2 of Sims, Stock and Watson (1986) shows that cov $(T^g(\tilde{\beta}_1 - \beta_1), T^{1/2}(\tilde{\beta}_2 - \beta_2))$ is zero, so that the limiting normal distribution in Proposition 1 applies to the estimators of the parameters of the stationary variables. For $\tilde{\beta}_1$, the limiting distribution may or may not be normal depending upon whether there is drift in the ARIMA process or not. Because of this separability result, estimation of δ can proceed as if y_t and $\bar{x}_t$ were stationary, and hence Proposition 1 and its extensions given below apply directly.

Proposition 1 shows that inference about β cannot be based upon the standard covariance matrix for $T^{1/2}(\tilde{\beta}-\beta)$ provided by most instrumental variables computer programs. That estimator is $\sigma^2(TQ_T)^{-1}T^{-1}Z'Z(TQ_T)^{-1}$, where σ^2 is an estimate of the variance of ϵ_t, and, even if ϵ_t were independently distributed, does not equal V_T asymptotically owing to heteroskedasticity in the errors ϵ_t. It is necessary therefore to provide an estimator of the covariance matrix of $T^{1/2}(\tilde{\beta}-\beta)$ which will be robust to the general form of evolutionary behaviour assumed for ϕ_t. What is required is that V_T be consistently estimated, and for this we have the following proposition.

<u>Proposition 2:</u> Under the assumptions earlier but with mixing coefficients of size 2 (or of size $2(r+a-1)^{-1}(r+a)$, $r > 1$), and $r' = 2(r+b)$ for some $b > 0$ in (ii)(b) and (iii)(a), the estimator

$$\hat{V}_T = T^{-1}\sum_{t=1}^{T} z_t'z_t\tilde{\epsilon}_t^2 + T^{-1}\sum_{\tau=1}^{\ell}\sum_{t=\tau+1}^{T}(z_t'\tilde{\epsilon}_t\tilde{\epsilon}_{t-\tau}z_{t-\tau} + z_{t-\tau}'\tilde{\epsilon}_{t-\tau}\tilde{\epsilon}_t z_t)$$

is a consistent estimator of V_T when $\tilde{\epsilon}_t = y_t - x_t\tilde{\beta}$ and ℓ is $o(T^{1/4})$[2].

---

[2] Newey and West (1986) give a weighted version which is a consistent estimator but ensures that $\hat{V}_T$ would be positive definite.

<u>Proof:</u> White (1984, Theorem 6.20), except the restriction upon ℓ is the correction found in Phillips (1985) and Newey and West (1986). □

Consistent estimation of δ when a ϕ_t^2 is available such that $E(\phi_t^2|\mathcal{F}_t) = \sigma_t^2$ is therefore easy, and Propositions 1 and 2 give the limiting theory. Unfortunately, since $\phi_t = \psi_t - E(\psi_t|\mathcal{F}_t)$, and it is going to be rare that $E(\psi_t|\mathcal{F}_t)$ is available (or known), ϕ_t will need to be estimated from available data. In most econometric work $E(\psi_t|\mathcal{F}_t)$ is made a linear function of q elements (w_t) taken from $\mathcal{F}_t$, that is $\psi_t = E(\psi_t|\mathcal{F}_t) + \phi_t$ is assumed to be $\psi_t = w_t\theta_0 + \phi_t$, where θ_0 is the true value of some parameters. ϕ_t would then be estimated by replacing θ_0 by some estimate $\hat{\theta}$.

Proposition 3 is concerned with the properties of the instrumental variables estimates when ϕ_t^2 is replaced by $\hat{\phi}_t^2$ and $\hat{\phi}_t = \psi_t - w_t\hat{\theta}$. It may be possible to state the conditions needed for Proposition 3 to hold in terms of the more primitive assumptions made earlier for Proposition 1, but it is easier to impose conditions directly upon w_t. Then Proposition 3 states that the limiting distribution of the "feasible" IV estimator $T^{1/2}(\hat{\beta}-\beta)$, $\hat{\beta} = (Z'\hat{X})^{-1}Z'y$, where $\hat{X}$ has t^{th} row $(\bar{x}_t\ \hat{\phi}_t^2)$, is the same as that of $T^{1/2}(\tilde{\beta}-\beta)$.

<u>Proposition 3:</u> If (i) $\hat{\theta} - \theta_0$ is $O_p(T^{-1/2})$, (ii) the set of conditions given earlier for Proposition 1 hold but with (z_t', x_t', ϵ_t) augmented with w_t, (iii) $E|(z_{ti}w_{tj}\phi_t)|^{r'+a} < \Delta_1 < \infty$ for some a > 0 $\forall t$, (iv) $E|z_{ti}w_{tj}w_{tk}|^{r'+a} < \Delta_2 < \infty$

for some $a > 0\ \forall t$, $i=1,\ldots,p, j=1,\ldots,q$

$$T^{1/2}(\tilde{\beta}-\beta) - T^{1/2}(\hat{\beta}-\beta) \text{ is } o_p(1).$$

<u>Proof:</u> As $\tilde{\beta} = (Z'X)^{-1}Z'y$ and $\hat{\beta} = (Z'\hat{X})^{-1}Z'y$ we first need to show that $T^{-1}Z'X - T^{-1}Z'\hat{X}$ is $o_p(1)$.

$$T^{-1}Z'X - T^{-1}Z'\hat{X} = T^{-1}Z'(X-\hat{X}) = \begin{bmatrix} 0 \\ T^{-1}\Sigma z_t'(\phi_t^2 - \hat{\phi}_t^2) \end{bmatrix} \quad (4)$$

The i^{th} element in $T^{-1}\Sigma z_t'(\phi_t^2 - \hat{\phi}_t^2)$ is

$$T^{-1}\Sigma z_{it}(\phi_t^2 - \hat{\phi}_t^2) = 2(T^{-1}\Sigma z_{it}w_t\phi_t)(\hat{\theta}-\theta_0) - (\hat{\theta}-\theta_0)'(T^{-1}\Sigma z_{it}w_t'w_t)(\hat{\theta}-\theta_0). \quad (5)$$

By Proposition 3.50 of White (1984), $(z_{it}w_t\phi_t)$ is a mixing sequence of order $(r' - 1)^{-1}r'$, while by Theorem 3.49 of White $z_{it}w_t'w_t$ also mixes. From the strong law of large numbers for mixing processes - White, Corollary 3.48 - and conditions (ii) and (iii) of the proposition, $T^{-1}\Sigma z_{it}w_t\phi_t$ and $T^{-1}\Sigma z_{it}w_t'w_t$ both converge to their expectations. Hence (5) is $o_p(1)$ and

$$\tilde{\beta} - \hat{\beta} = (T^{-1}Z'X)^{-1}T^{-1/2}\Sigma z_t'(\phi_t^2 - \hat{\phi}_t^2)\delta + o_p(1). \quad (6)$$

Inspection of (5) multiplied by $T^{1/2}$ shows that with condition (i) and $T^{-1}\Sigma z_{it}w_t\phi_t \xrightarrow{p} 0$ (because $E(\phi_t|z_{it}, w_t) = 0$), $T^{-1/2}\Sigma z_t'(\phi_t^2 - \hat{\phi}_t^2)$ is $o_p(1)$, verifying the proposition. □

2.2 Problems in Constructing Proxies and Instruments

An emphasis on the IV method for handling the errors in variables problem arising when ϕ_t^2 replaces σ_t^2 fits well with the recent tendency in macroeconomic research to generate estimators from orthogonality conditions. The popularity of this strategy is partly explained by the robustness of the IV estimator to any failure of the econometrician to have the complete information set available to the agent. Thus if (1) was $E(y_t|\mathcal{F}_t) = E(\psi_t|\mathcal{F}_t)\gamma$, and the only information available to an econometrician was $\mathcal{G}_t \subset \mathcal{F}_t$, an estimable equation is

$$y_t = E(\psi_t|\mathcal{G}_t)\gamma + e_t + (E(\psi_t|\mathcal{F}_t) - E(\psi_t|\mathcal{G}_t))\gamma. \tag{7}$$

If IV is applied to (7) with instruments chosen from $\mathcal{G}_t$, the estimator of γ is consistent as $E(E(\psi_t|\mathcal{F}_t) - E(\psi_t|\mathcal{G}_t))|\mathcal{G}_t) = 0$ from the law of iterated expectations (Nelson (1975), Wickens (1982)). Unhappily, such a felicitous outcome does not carry over to the estimation of the effects of risk. Proposition 4 deals with this point.

Proposition 4: Unless $\mathcal{G}_t = \mathcal{F}_t$, variables in $\mathcal{G}_t$ will generally be correlated with the error term in any model in which $\sigma_t^2 = E(\psi_t^2|\mathcal{F}_t)$ is replaced by $\bar{\sigma}_t^2 = E(\psi_t^2|\mathcal{G}_t)$.

Proof: Let $\bar{\phi}_t = \psi_t - E(\psi_t|\mathcal{G}_t)$ so that $\sigma_t^2 = E(\phi_t^2|\mathcal{F}_t), \bar{\sigma}_t^2 = E(\bar{\phi}_t^2|\mathcal{G}_t)$ and the model estimated by the investigator is based upon

$$y_t = \bar{x}_t\gamma + \bar{\sigma}_t^2\delta + e_t + (\sigma_t^2 - \bar{\sigma}_t^2)\delta. \tag{8}$$

Evaluating the expectation of the last term in (8) conditional upon $\mathcal{G}_t$ gives $\delta E((\sigma_t^2 - \bar{\sigma}_t^2)|\mathcal{G}_t) = \delta E[(E(\phi_t^2|\mathcal{F}_t) - E(\bar{\phi}_t^2|\mathcal{G}_t))|\mathcal{G}_t]$. Substituting $\phi_t = \bar{\phi}_t + E(\psi_t|\mathcal{G}_t) - E(\psi_t|\mathcal{F}_t)$, and simplifying using the law of iterated expectations shows that $\delta E((\sigma_t^2-\bar{\sigma}_t^2)|\mathcal{G}_t) = \delta\{[E(\psi_t|\mathcal{G}_t)]^2 - E[(E(\psi_t|\mathcal{F}_t))^2|\mathcal{G}_t]\}$. This expression is generally non-zero as Jensen's inequality for conditional expectations has $E[(E(\psi_t|\mathcal{F}_t))^2|\mathcal{G}_t] \geq [E(\psi_t|\mathcal{G}_t)]^2$.

□

Because of the correlation of members of the set $\mathcal{G}_t$ with $(\sigma_t^2 - \bar{\sigma}_t^2)$, consistent estimation of δ will almost always require that $\mathcal{F}_t$ be known. *Only if* $\delta = 0$, in which case $\delta(\sigma_t^2 - \bar{\sigma}_t^2)$ does not appear in the error term of (8), will consistent estimation be possible with a truncated information set. This makes it crucial that $\mathcal{G}_t$ be set wide enough to encompass $\mathcal{F}_t$, when generating the conditional expectation. Such a lack of robustness makes the estimation of models with risk terms a very difficult task. Throughout the remainder of this paper we maintain the assumption that $\mathcal{G}_t = \mathcal{F}_t$.

Even if the set of variables in $\mathcal{F}_t$ can be identified however, the assumption that $E(\psi_t|\mathcal{F}_t)$ is *linear* in w_t could well be incorrect. Moreover, no guidance has been given concerning the selection of instruments z_t. It is likely that z_t will have to be a non-linear function of $\mathcal{F}_t$. Suppose that $E(\psi_t|\mathcal{F}_t) = \theta\psi_{t-1}$ and that ϕ_t was normal and identically distributed with zero mean and conditional variance $\sigma_t^2=\alpha_0+\alpha_1\phi_{t-1}^2$. Because $\psi_{t-1} \in \mathcal{F}_t$, it might be taken as an instrument for ϕ_t^2, but clearly $E(\psi_{t-1}\phi_t^2) = 0$ when $|\theta| < 1$, violating the requirement for a good instrument in assumption (iii)(b). If, however, ψ_{t-1}^2

is taken as the instrument, $E(\psi_{t-1}^2\phi_t^2) \neq 0$ and it satisfies the conditions. Of course an ideal instrument for ϕ_t^2 would be σ_t^2, but economic theory is not of much help in indicating how σ_t^2 can be expected to vary with $\mathcal{F}_t$ - all models that allow for risk merely note the dependence of decisions upon the conditional variance, assuming that agents will learn about how the available information maps into this moment.

How then could estimates of the conditional mean and variance of ψ_t be found that do not rely on linearity? Our objective is to estimate $E(g(\psi_t)|w_t)$, where w_t represents the basic elements in $\mathcal{F}_t$ and $g(\psi_t)$ is either ψ_t or ψ_t^2. Non-parametric estimation procedures have been advanced in recent years to estimate such conditional moments, with the main econometric references being Bierens (1985), Robinson (1983), Robinson (1986), and Ullah (1986). Readers are referred to these papers for a more detailed discussion of the methods employed.

Robinson (1983) is perhaps the most general treatment of those mentioned above. He considered how to estimate quantities of the form $E(g(\psi_t)|w_t)$ where g is a Borel function on R^1 such that $E|g(\psi_t)| < \infty$ and w_t contains a finite number of lags of ψ_t. Robinson, and the others referred to above, employ the kernal method of moment estimation pioneered by Nadaraya (1964) and Watson (1964) to produce an estimator of the conditional mean of ψ_t at $w_t = \bar{w}_t$ of the form

$$\hat{E}(g(\psi_t)|\bar{w}_t) = \frac{[(\gamma_T^q T)^{-1}\sum_{j=1}^{T} g(\bar{\psi}_j)K((\bar{w}_t-\bar{w}_j)/\gamma_T)]}{[(\gamma_T^q T)^{-1}\sum_{j=1}^{T} K((\bar{w}_t-\bar{w}_j)/\gamma_T)]}, \qquad (9)$$

where $\bar{w}_j$, $\bar{\psi}_j$ are observed data, γ_T is a "bandwidth" parameter that is typically proportional to $T^{-1/(4+q)}$, and K( ) is a kernal function that aims to smooth the density. Many types of kernals might be employed. A popular one used later in Section 5 is the normal kernal (see Singhand Ullah (1985, p.31) for this).

Under various restrictions upon γ_T, differentiability of the density function of ψ_t, w_t, boundedness of $g(\psi_t)$, and assuming (ψ_t, w_t) are strictly stationary stochastic processes that are α-mixing with mixing coefficients α_j such that $T \sum_{j=T}^{\infty} \alpha_j$ is o(1), Robinson shows that $\hat{E}(g(\psi_t)|\bar{w}_t) \xrightarrow{p} E(g(\psi_t)|\bar{w}_t)$ and $(a_T T)^{1/2}(\hat{E}(g(\psi_t)|\bar{w}_t) - E(g(\psi_t)|\bar{w}_t))$ has a limiting normal distribution, where $a_T = \gamma_T^q$. If $g(\psi_t)$ is not bounded but has bounded moments of order h, Robinson also showed that the normality result goes through provided $T \sum_{j=T}^{\infty} \alpha_j^{1-(2/h)}$ is O(1), (h>2). Hence it is possible to estimate the conditional mean and variance of ψ_t without specifying the exact way in which the conditional moments depend on w_t. As well, the central limit theorem enables confidence bands to be placed around estimates.

A number of quantities can be constructed from this result. Suppose non-parametric estimates of $E(\psi_t|\bar{w}_t)$ and $E(\psi_t^2|\bar{w}_t)$ are available. Then we can define $\hat{\sigma}_t^2 = \hat{E}(\psi_t^2|\bar{w}_t) - [\hat{E}(\psi_t|\bar{w}_t)]^2$ and thereby derive a consistent estimate of σ_t^2 by non-parametric methods. The non-parametric mean estimator $\hat{m}_t = \hat{E}(\psi_t|\bar{w}_t)$ yields non-parametric residuals $\hat{\phi}_t = \psi_t - \hat{m}_t$. Previously such estimates were utilized to set up an IV estimator, and it is natural to do that again. Thus the equation to be estimated is:

$$y_t = \bar{x}_t\gamma + \hat{\phi}_t^2\delta + e_t + \delta(\sigma_t^2 - \phi_t^2) + \delta(\phi_t^2 - \hat{\phi}_t^2) \quad (10)$$

with instruments z_t. Proposition 5 gives a statement of conditions under which the IV estimator of $\beta' = (\gamma'\delta)$ is consistent when $\hat{m}_t$ is found from a non-parametric estimation method.

Proposition 5: Under (i) the conditions set out in Proposition 1,and, (ii) that $\hat{m}_t - m_t$ is uniformly consistent, i.e. $\sup_{w \in R^q} (\hat{m}_t - m_t)$ is $o_p(1)$, $\tilde{\beta} = (Z'X)^{-1}Z'y$ is a consistent estimator of β.

Proof: It is only necessary to show that $T^{-1}\Sigma z_t(\phi_t^2 - \hat{\phi}_t^2) \overset{p}{\rightarrow} 0$ as $T^{-1}\Sigma z_t e_t$ and $T^{-1}\Sigma z_t(\sigma_t^2 - \phi_t^2)$ were shown to converge to zero in the proof of Proposition 1. A sufficient condition for this is that $T^{-1}\Sigma z_t(\phi_t^2 - \hat{\phi}_t^2) \overset{p}{\rightarrow} 0$ or $-2T^{-1}\Sigma z_{it}\phi_t(\hat{\phi}_t - \phi_t) + T^{-1}\Sigma z_{it}(\hat{\phi}_t - \phi_t)^2 \overset{p}{\rightarrow} 0 \ \forall i = 1, \ldots, p$. As the first element in this expression is of higher order than the second we concentrate upon it, re-writing it as $T^{-1}\Sigma z_{it}\phi_t(m_t - \hat{m}_t)$. A sufficient condition for $T^{-1}\Sigma z_{it}\phi_t(m_t - \hat{m}_t) \overset{p}{\rightarrow} 0$ is that $T^{-1}\Sigma|z_{it}\phi_t| \, |m_t - \hat{m}_t| \overset{p}{\rightarrow} 0$, and the latter is bounded by $[T^{-1}\Sigma|z_{it}\phi_t|] \sup_{w \in R^q} |\hat{m}_t - m_t|$. As $\sup |(m_t - \hat{m}_t)|$ is $o_p(1)$, $T^{-1}\Sigma z_{it}\phi_t(m_t - \hat{m}_t)$ is $o_p(1)$ if $T^{-1}\Sigma|z_{it}\phi_t|$ is $O_p(1)$. Because z_{it} and ϕ_t are stationary and ergodic processes, from the ergodic theorem (White (1984, Theorem 3.34, p.42)), $T^{-1}\Sigma|z_{it}\phi_t| \overset{p}{\rightarrow} E(|z_{it}\phi_t|)$ if $E|z_{it}\phi_t| < \infty$. But from the Cauchy-Schwartz inequality $E|z_{it}\phi_t| \leq [E(z_{it}^2)E(\phi_t^2)]^{1/2} < \infty$ under the assumptions of Proposition 1.

□

Of course, the pertinent question concerns the likelihood of uniform convergence of $\hat{m}_t$. This has been shown under a number of different assumptions about the way in which ψ_t and w_t are generated. Bierens (1983) allows the processes to be ϕ-mixing or to have a ϕ-mixing base and shows uniform consistency under the condition that $T\gamma_T^{2q} / \sum_{j=0}^{T} \phi_j^{1/2} \to \infty$ as $T \to \infty$, where ϕ_j are here the mixing coefficients. Sufficient conditions for this are $T\gamma_T^{2q} \to \infty$ and $\sum_0^{\infty} \phi_j^{1/2} < \infty$. When processes are α-mixing, using the proof in Singh and Ullah (1987), $\hat{m}_t - m_t$ can be shown to be uniformly consistent provided $T\gamma_T^{2q} \to \infty$ and $\sum_{j=0}^{\infty} \alpha_j^{1-(2/h)} < \infty$, where the h'th moment of $\Psi'_T (h>2)$ is bounded.[3] Thus uniform convergence of $\hat{m}_t$ seems a reasonable assumption for most time series, although there is no completely general treatment yet available in the literature.

It is instructive to compare this outcome with that when the residuals $\hat{\phi}_t$ derive from a parametric regression. In that case the convergence to zero was at the rate $T^{-1/2}$, whereas $\sup|\hat{m}_t - m_t|$ normally converges to zero at a rate $T^{-1/2}a_T^{-1}$, and so the use of non-parametric procedures slows the rate at which consistency is achieved by an amount directly dependent upon the dimension of w_t, i.e. q. This conclusion supports the intuition that a shift to non-parametric methods requires larger numbers of observations if they are to be successfully employed.

---

[3]This correction to Singh and Ullah's (1985) result on uniform consistency was suggested by P. M. Robinson in a personal communication to the second author, who is grateful for this information.

Having established consistency under fairly general conditions, questions of inference come to the fore, and these prove to be far more complex. One important case can be dispatched immediately: when it is desired to test if $\delta=0$. A good deal of econometric work involves testing such an hypothesis, i.e. checking for the *existence* of a risk premium, giving Proposition 6 below a substantial range of application.

Proposition 6: Under the same conditions as in Propositions 2 and 5, *and* $\delta=0$, the covariance matrix of the IV estimator, formed by generating $\hat{\phi}_t$ non-parametrically, is the same as that of Proposition 1, and may be consistently estimated as in Proposition 2.

Proof: When $\delta=0$ the error in (10) is just e_t, and it is only necessary to show that $T^{-1}Z'\hat{X}$ is $O_p(1)$. The only new term is $T^{-1}\Sigma z_{it}\hat{\phi}_t^2$ and this converges to $\text{plim}_{T\to\infty} T^{-1}\Sigma z_{it}\phi_t^2$ if $T^{-1}\Sigma z_{it}\phi_t(\phi_t - \hat{\phi}_t) \xrightarrow{p} 0$, an outcome demonstrated in the proof of Proposition 5. □

Unfortunately, once the realm of testing the sharp hypothesis $\delta=0$ is left, it becomes difficult to provide a precise answer to how the covariance matrix of the IV estimator of Proposition 4 should be computed.[4] Ideally, one would like to extend Proposition 3, but this does not seem to be possible. The situation is akin to that in Carroll (1982) in that it requires $T^{-1/2}\Sigma z_t e_t(m_t - \hat{m}_t) \xrightarrow{p} 0$. Carroll demonstrated this when all data was

[4]Note that if all one was interested in was testing if $\delta=0$ *any* proxy for σ_t^2 could be entered into a regression of y_t against $\bar{x}_t$ and the proxy and the OLS results used to test if $\delta=0$. Presumably the power of the test depends upon the correlation of the proxy and σ_t^2, and so demands a careful choice.

independently and identically distributed, σ_t^2 was a function of only a single variable w_t, and w_t had bounded, compact support. But it seems very hard to generalize this result. Fundamentally, the difficulty in doing so arises from the fact that $\hat{m}_t$ is a ratio of random variables and, with standard kernels such as the normal, there is no guarantee that its finite-sample variance exists. Hence, it becomes hard to apply standard limit theorems. Robinson (1986b) has managed to extend Carroll's result by replacing the kernel estimator m_t by a nearest-neighbour estimator, and it may be necessary to follow that route here as well. Clearly, this is an area that will repay future research.

3. Risk as a Regressand

It is not always the case that the model to be estimated has risk as a regressor. Sometimes information is sought upon the importance of different influences upon the level of risk, in which case it appears as the dependent variable in a regression. Two examples might be cited. The first arises in the debate over whether the variability in inflation was an increasing function of its level or not. Applications exist in which measures of "inflation risk" are related to the level of inflation, for example Hercowitz (1981), Pagan et al (1983) and Engle (1983). A second type comes from the volatility debate initiated by Shiller (1981), which argued that stock market prices fluctuated much more than dividends. Malkiel (1979) and Pindyck (1984) attribute the decline in real stock market values between the mid-1960's and the early 1980's as due to a rise in risk. This could be captured by assuming that $\sigma_t^2 = \sigma^2 + \alpha D_t$, where D_t is a dummy variable that is unity after the time in which risk is assumed to rise. Alternative hypotheses in the same line are the investigation by Poterba and Summers (1986), which has σ_t^2 following an autoregression $\sigma_t^2 = \alpha_o + \alpha_1\sigma_{t-1}^2$, or French et al (1986) who model it as an ARIMA process.

All of these possibilities can be captured in (11)[5]

$$\sigma_t^2 = \sum_{j=1}^{k} \alpha_j\sigma_{t-j}^2 + \bar{x}_t\gamma \tag{11}$$

which can be converted to a suitable estimating equation

$$\phi_t^2 = \sum_{j-1}^{k} \alpha_j\phi_{t-j}^2 + \bar{x}_t\gamma + \sum_{j=0}^{k} \alpha_j(\sigma_{t-j}^2 - \phi_{t-j}^2) \ , \ (\alpha_0 = -1) \tag{12}$$

provided that ϕ_t exhibits the weak property. It is the divergence between ϕ_{t-j}^2 and σ_{t-j}^2 that makes (12) a stochastic relation.

If the α_j in (11) were all zero, (12) would be a standard linear model to which least squares could be applied provided $\bar{x}_t$ appeared in the information set: if it did not an instrumental variables estimator would need to be invoked. Thus a consistent estimator of γ would be readily available, and the only complication is the need to allow for the heterogeneity in the error term when performing inferences.

---

[5]Poterba and Summers add a disturbance term to (11). Although this just augments the error term in (12) and does not change our argument in any way, it is not at all clear why they should add an error term to (11). By definition σ_t^2 is a conditional expectation and therefore a function of variables entering the conditioning set. It would therefore be necessary to argue that the appended error appears in the conditioning set used by agents but is absent from that of the econometrician's.

When lagged terms of ϕ_t^2 appear in (12), OLS would generally be inconsistent. Supose that $\bar{x}_t$ is absent from (12) and that k = 1. Then the OLS estimator of α_1 is $\hat{\alpha}_1 = \alpha_1 + (\sum_{t=1}^{T} \phi_{t-1}^4)^{-1} (\sum_{t=1}^{T} \phi_{t-1}^2(\sigma_{t-1}^2 - \phi_{t-1}^2))\alpha_1$, and the inconsistency is $\alpha_1\mu_4^{-1}(v - \mu_4)$, where $v = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(\sigma_{t-1}^4)$ and $\mu_4 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(\phi_{t-1}^4)$. If ϕ_t was conditionally normally distributed then

$E(\phi_t^4) = E[E(\phi_t^4|\mathcal{F}_t)] = E[3\sigma_t^4]$ and $\mu_4 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} 3E(\sigma_{t-1}^4)$, making the inconsistency equal to $-(2/3)\alpha_1$. Hence, even if the true value of α_1 is close to unity, the OLS estimate will tend to be very low by comparison.[6]

Estimation of (12) therefore needs to be done by IV. However, the presence of an MA(k) in the disturbance term of (12) introduces a new complication, since although members of $\mathcal{F}_t$ are uncorrelated with $(\phi_t^2 - \sigma_t^2)$, they need not be with $\{\phi_{t-j}^2 - \sigma_{t-j}^2\}_{j=1}^k$. What is certain, however, is that elements of $\mathcal{F}_{t-k-1}$ will be, and this means that instruments need to be constructed by lagging the information set at least (k+1) times. Of course, this is an old strategy for getting consistent estimators of the AR part of ARMA models.

[6]Poterba and Summers (1986) conclude (p.1147) that the inconsistency is small, despite the fact that they assume normality for the ϕ_t (footnote 10, p.1147).

4. The Properties of Alternative Risk Measures

As mentioned in the introduction, there are quite a number of papers that have been concerned to calibrate the effects of risk and have entered some measure of it into regressions. Because these measures are widespread it is important to understand why they can be defective, and we turn to an investigation of that here. Some of these procedures are specific to inflation risk, although they might be adapted to other contexts as well. Four methods are singled out for analysis.

(a) Moving Average Measures

This is by far the most popular technique; references would be legion but a small sample is Gylfason (1981), Klein (1977) Ibrahim and Williams (1978) and Pindyck (1984). The strategy here is replace σ_t^2 by ϕ_t^2 constructed as

$$\phi_t^2 = m^{-1} \sum_{k=0}^{m-1} (\psi_{t-k} - \mu_{t-k})^2 \text{ where } \mu_t = m^{-1} \sum_{\ell=0}^{m-1} \psi_{t-\ell}.$$

It is not entirely clear what argument for this definition is. Moreover, it is not hard to construct cases where it can be very misleading. For example, suppose that $\psi_t = bt + dt^2$, and is therefore perfectly predictable. In such circumstances σ_t^2 must be zero, and any risk measure should reflect this. But ϕ_t^2 as defined above varies systematically with t whenever $m > 1$.

The origin of these difficulties is the failure to fully specify the information set underlying the construction of σ_t^2. One might attempt to surmount them by defining $\mathcal{F}_t$ as the elements making up μ_t, although that would force the moving average to be based only upon past values of ψ_t. With this proviso, it is apparent that the proposed definition is motivated by the formula for the sample variance of a stationary series (see Klein (p.700)); for non-stationary series it is a misleading indicator of σ_t^2, producing at best an average of the σ_t^2's as $m \to \infty$. Consequently, variance changes tend to get blurred.

As a regressor, the use of $\phi_t^2 = m^{-1}\sum_{j=0}^{m-1} (\psi_{t-j} - \mu_{t-j})^2$ in place of σ_t^2 must lead to inconsistent estimators if the current value of ψ_t appears in it. Few authors appreciate this point; Huizinga and Mishkin (1985) being an exception. One of the few arguments that might be advanced in support of ϕ_t^2 as a regressor is that (1) should have $m^{-1}\sum_{j=0}^{m-1} \delta_{t-j}^2$ in place of δ_t^2. Then ϕ_t^2 has the strong property that $\phi_t^2 \xrightarrow{p} m^{-1}\sum_{j=0}^{m-1} \sigma_{t-j}^2$ as $m \to \infty$, and, with $m/T \to 0$ as $T \to \infty$, δ would be consistently estimated by regression. However this would not seem to be what investigators using this procedure actually had in mind, as they almost invariably write the risk terms with no lags.

(b) Relative Price Measures

One of the most common measures of σ_t^2 when it relates to prices or inflation is that derived from a series showing the relationship of individual prices to the aggregate level price. With P_i the log of the prices of the i^{th} commodity and P the log of the general price level, σ_t^2 is replaced by $\phi_t^2 = m^{-1}\Sigma(P_{it} - P_t)^2$, where the summation is over m commodities. To analyze the utility of this formula, some model of $P_{it} - P_t$ is necessary; the one adopted here being that of Lucas (1973) as formulated by Cukierman and Wachtel (1979) and generalized in Pagan et al. (1983). In the latter paper it was shown that, as $m \to \infty$, $P_{it} - P_t = D_t a_i + \eta_t b_i + v_{it} + o_p(1)$, where η_t is a macroeconomic shock common to all markets, v_{it} is a market-specific shock and D_t represents the mean value of non-price variables shifting demand and supply curves. Cukierman and Wachtel's formulation has $P_{it} - P_t = v_{it} + o_p(1)$, as they assume $D_t = 0$ and that all markets identical making $b_i = 0$.

With the identical market assumption and *no systematic factors*, $m^{-1}\Sigma(P_{it}-P_t)^2 = m^{-1}\Sigma v_{it}^2 = \phi_t^2$ and it is clear that $\phi_t^2 \to \sigma_t^2$ as $m \to \infty$ if $\sigma_t^2 = E(v_{it}^2)$. Therefore, the desirable strong property of a good variance measure alluded to earlier holds, and OLS delivers consistent estimators.[7] Undoubtedly, it is this type of argument which underlines the popularity of such a measure. It should be clear however, that the strong property fails to hold when markets are not identical. Then the macroeconomic shock common to all has differential effects upon the prices in each market, the diversity being dependent upon the demand and supply responses. Only the weak property for ϕ_t^2 can be invoked in the general case, and thus the estimation difficulties outlined in Sections 2 and 3 recur.

(c) Survey Measures

Sometimes measures of uncertainty due to inflation or interest rate fluctuations have been derived from the dispersion of responses given by

---

[7]Note that many studies actually use the inflation rate rather than the price level so that $\phi_t^2 \to 2E(v_{it}^2)$. This may raise questions concerning the definition of σ_t^2.

individuals to questions about their anticipations, e.g. Levi and Makin (1979). Analyzed in the context of the same model as that used to establish the characteristics of relative price measures, the survey based approaches are found to have similar deficiencies. This is most clearly seen by examining the polar case where each individual transacts in only one market,[8] so that his anticipated price level combines information from the i^{th} market with a "macro" projection of the aggregate price level. Designating this by $\hat{P}_t(i)$, it was show in in Pagan et al (1983) that $\hat{P}_t(i) - P_t = \theta_i(\eta_t\beta_i + \epsilon_{it}) - \eta_t\Sigma w_k\beta_k - \Sigma w_k\epsilon_{kt}$, where θ_i is the weight given to the i^{th} price in forming the aggregate anticipation, w_k are the weights used in constructing the aggregate price level, and β_i depends on the demand and supply curve parameters in the i^{th} market. It is apparent from this formula that the macro shock appears in the survey measure just as it does in indices constructed from relative price dispersion, and therefore the two procedures share the same set of difficulties. An alternative definition is to relate $\hat{P}_t(i)$ to $\Sigma w_k \hat{P}_t(i)$, i.e., to the average of respondents' anticipations, in which case $\hat{P}_t(i) - \Sigma w_k\hat{P}_t(i) = \theta_i(\eta_t\beta_i + \epsilon_i) - \Sigma\theta_k(\eta_t\beta_k + \exists_k)$, and the macro shock disappears only if all markets are identical. When markets differ, the macro shock has different effects upon the anticipations of actors, depending upon which market they transact in.

(d) Measures Based on Particular Parameterisations

All of the above methods could be construed as attempts to avoid an explicit parameterisation for σ_t^2. Some authors have gone much further

[8] Nothing in the argument below depends on this assumption, but the formula given for the price differential will have different weights attached to the shocks η_t and ϵ_{it}.

however, and explicitly parameterised σ_t^2, e.g. Hansen and Hodrick (1983), Domowitz and Hakkio (1984) and Engle et al (1984). What differentiates these studies are the variables σ_t^2 is assumed to depend upon and the estimation methods employed to obtain estimates of the parameters of interest, γ and δ. We therefore turn to an examination of some of the main features of this literature.

Existing studies invariably assume that σ_t^2 is linearly related to some variable z_t in the form $\sigma_t^2 = \sigma^2 + z_t\alpha$. If observations on ϕ_t^2 are available one could re-write this as

$$\phi_t^2 = \sigma^2 + z_t\alpha + (\phi_t^2 - \sigma_t^2) = \sigma^2 + z_t\alpha + v_t \tag{13}$$

and joint estimation of (3) with (13) could be performed. One estimator of (3) and (13) would be the instrumental variables estimator described earlier, but alternative procedures would be to apply FIML or to engage in two-stage procedures (one variant of the latter would be to regress ϕ_t^2 against unity and z_t, producing $\tilde{\sigma}^2$ and $\tilde{\alpha}$, and then use $\tilde{\sigma}_t^2 = \tilde{\sigma}^2 + z_t\tilde{\alpha}$ as a regressor in (3)). Two-stage procedures will only provide valid inferences provided the adjustments to the covariance matrix described in Pagan (1984, p. 240) are done, whereas FIML does not share this problem. However, FIML is far more sensitive to specification errors than the instrumental variables estimator. If the set of variables driving (13) is larger than z_t, FIML will generally be inconsistent, whereas (IV) will not be. Thus, even if σ_t^2 is parameterised, there are sound reasons for employing the instrumental variable estimators discussed in Section 2.

Donowitz and Hakkio (1985) examine risk premia in the exchange rate. In their theoretical model ψ_t represents foreign and domestic monies, which they take as following a VAR, so that $\hat{\phi}_t$ would be the residuals from the fitted

VAR, and estimation could be done as described above. However, in practice they depart from their theoretical model, making σ_t^2 a function of the past history of e_t, as defined in (3); in particular they force σ_t^2 to be Engle's (1982) ARCH process, say $\sigma_t^2 = \sigma^2 + e_{t-1}^2\alpha$, and then apply MLE to (3) to obtain estimates of γ, β and α.

A potential problem with modelling σ_t^2 as an ARCH process is that the MLE of δ is almost certain to be inconsistent if σ_t^2 is not of the ARCH form, since the conditional mean in (1) is then mis-specified. Unfortunately, it is not possible to derive an estimate of ϕ_t (in this case e_t) that could be used to define the IV estimator of section 2 if σ_t^2 really is an ARCH process; the reason being that $\mathcal{F}_t$ would be a function of the *entire* past history of $\{y_t, \bar{x}_{t+1}\}$, and so non-parametric methods which require $\mathcal{F}_t$ to be finite dimensional cannot be applied. If only a finite number of members of $\mathcal{F}_t$ are used, $\mathcal{G}_t$, Proposition 4 shows that the IV estimator using $\hat{\phi}_t^2$ based upon $(\Psi_t - E(\Psi_t|\mathcal{G}_t))^2$ would be inconsistent.

Faced with this difficulty it is absolutely imperative that some indication of whether the ARCH assumption is valid be obtained. A specification test of the Hausman (1978) type may be implemented in the following way. Let $\hat{e}_t$ be the fitted residuals from the ARCH process and $\hat{\sigma}_t^2$ be the estimated value of σ_t^2 implied by the MLE. Form IV estimators of δ and γ in (3) by using $\hat{\sigma}_t^2$ as an instrument for $\hat{e}_t^2$ ($\hat{e}_t = \hat{\phi}_t$ in this case). When the σ_t^2 do follow an ARCH process it is obvious that this IV estimator is consistent, since $\hat{\sigma}_t^2$ is a function of past information and this has zero correlation with $e_t^2 - \sigma_t^2$. In contrast, the most efficient estimator of δ and γ under these conditions is the MLE. A comparision of these two estimators is therefore the suggested specification error test. Provided the specification is correct both estimators are consistent, whereas under the alternative there is no reason to believe that they will have the same probability limit since the set of first order conditions defining both estimators are quite different. Precisely in what circumstances this test will be powerful is however beyond this paper.

5. Two Applications

5.1 Risk in the Foreign Exchange Market

The presence of a risk premium in foreign exchange markets has been much debated. A number of studies have concluded that there is a risk premium, but that the evidence for it is fairly weak, e.g. Hansen and Hodrick (1983) and Domowitz and Hakkio (1985). Many studies of the efficiency of the foreign exchange market actually proceed as if there were no risk premium, and a good example of this is the research reported by Longworth (1981) on the Canadian-U.S. exchange market over the period 1970/7 to 1978/12. As the data used in that study was available to us, we undertook to examine the evidence for a risk premium in that market over the sample period.

The data for the model constituted the log of the spot rate (S_t) and the log of the 30-day forward rate F_{t-1}. Longworth fitted the standard model (δ=0)

$$S_t = a+b\, F_{t-1} + \delta\sigma_t^2 \qquad (14)$$

and tested market efficiency with the null hypothesis a=0, b=1. In the first column of Table 1 we present results from this regression; estimates of a and b are close to their theoretical values. However, a time series analysis of

S_t and F_t, shows that both series are ARIMA(0,1,0) with no evidence of any drift, and therefore the "t statistics" in Table 1 cannot be taken to be asymphotetically normally distributed (Phillips and Durlauf (1986)). A close examination of the residuals of the estimated model points to

TABLE 1: Estimated Model of the \$US/\$C*

	(1)	(2)	(3)
a	.00036	-.0013	-.0008
	(.29)	(1.19)	(.65)
b	1.0229	1.0291	1.0276
	(43.29)	(52.90)	(51.72)
δ		19.311	13.975
		(6.71)	(2.20)

*The estimated model is $S_t = a + bF_{t-1} + \delta\sigma_t^2$. In (2) the proxy for σ_t^2 is $\hat{\phi}_t^2$, where $\hat{\phi}_t$ is generated non-parametrically with conditioning set F_t = $[S_{t-1}-F_{t-2}, S_{t-2}-F_{t-3}]$. In (3) estimation is by generalized instrumental variables with the non-parametric estimator of σ^2 used as instrument for $\hat{\phi}_t^2$. Non-parametric estimation was performed as described in the text. Absolute "t-values" are in parentheses. All computations were done in the DFIT micro-computar package written by B. and H. Pesaran on a COMPAQ Deskpro 286.

potential weaknesses in it. In particular there is evidence of non-linear effects, with the t-statistic for the regression of the squared residuals against their lagged value being 2.05.

Our first task is to check for the *existence* of a risk premium, i.e. to test if δ=0 in (14). As observed in footnote 3, *any* proxy for σ_t^2 could be used for that purpose. Moreover, OLS is a perfectly appropriate estimator of

$\delta=0$ since, under the null hypothesis, the correlation of ϕ_t^2 with $\delta(\sigma_t^2 - \phi_t^2)$ is zero. The best proxy to use would be σ_t^2. On the basis of theoretical models such as Stockman (1978), measures of σ_t^2 should derive from the conditional variances of domestic and foreign money supplies. Instead we follow Domowitz and Hakkio (1985) who make $\phi_t = e_t = S_t - F_{t-1} - \delta\sigma_t^2 = \Psi_t - E(\Psi_t|F_t)$, where $\Psi_t = S_t - F_{t-1}$.[9] F_t was then defined as $)\Psi_{t-1}, \Psi_{t-2})$ and a non-parametric estimate of ϕ_t was found. The normal kernal, $\kappa(\cdot) = (2\Pi)^{-1/2} \exp\{-(2\gamma_T^2 s_\Psi^2)^{-1}[(\Psi_{t-1}-\bar{\Psi}_{t-1})^2 + (\Psi_{t-2}-\bar{\Psi}_{t-2})^2]\}$, where s_Ψ is tje sample standard deviation of Ψ_t, $\bar{\Psi}_{t-j}$ is the observed value of Ψ_{t-j} and $\gamma_t = T^{-1/6}$, was used to compute $E(\Psi_t|\bar{\Psi}_{t-1},\bar{\Psi}_{t-2})$ (arguments for the use of s_Ψ and γ_T of this form are given in Singh and Ullah (1985)). $\hat{\phi}_t^2$ was then used as a proxy for σ_t^2, and the resulting regression is given in column 2 of Table 1. Proposition 6 covers this case.

From the results in Table 1 there is very strong evidence that the basic "efficiency" model is not a complete representation of the data. Of course, rejection of the null hypothesis $\delta=0$ need not lead to acceptance of the alternative that the missing factor is a risk premium, but the fact that theoretical models accord it a role in the determination of the discount $S_t - F_{t-1}$ is suggestive. It is of some interest to see how σ_t^2 varied over the sample period. Accordingly, σ_t^2 was estimated non-parametrically in the same

[9]Actually these authors define $\phi_t = S_t - E(S_t|F_t)$, where the conditional expectation is $a+bF_{t-1}+\delta\sigma_t^2$ and σ_t^2 is defined as an ARCH process in terms of ϕ_{t-j}^2. Because S_t and F_t are ARIMA(0,1,0), we assume that the co-integrating vector is (1 - 1) as given by theory. This must also be true for Domowitz and Hakkio, as the asymptotic theory for ARCH processes requires ϕ_t to be stationary and this could not be so if $b\neq1$.

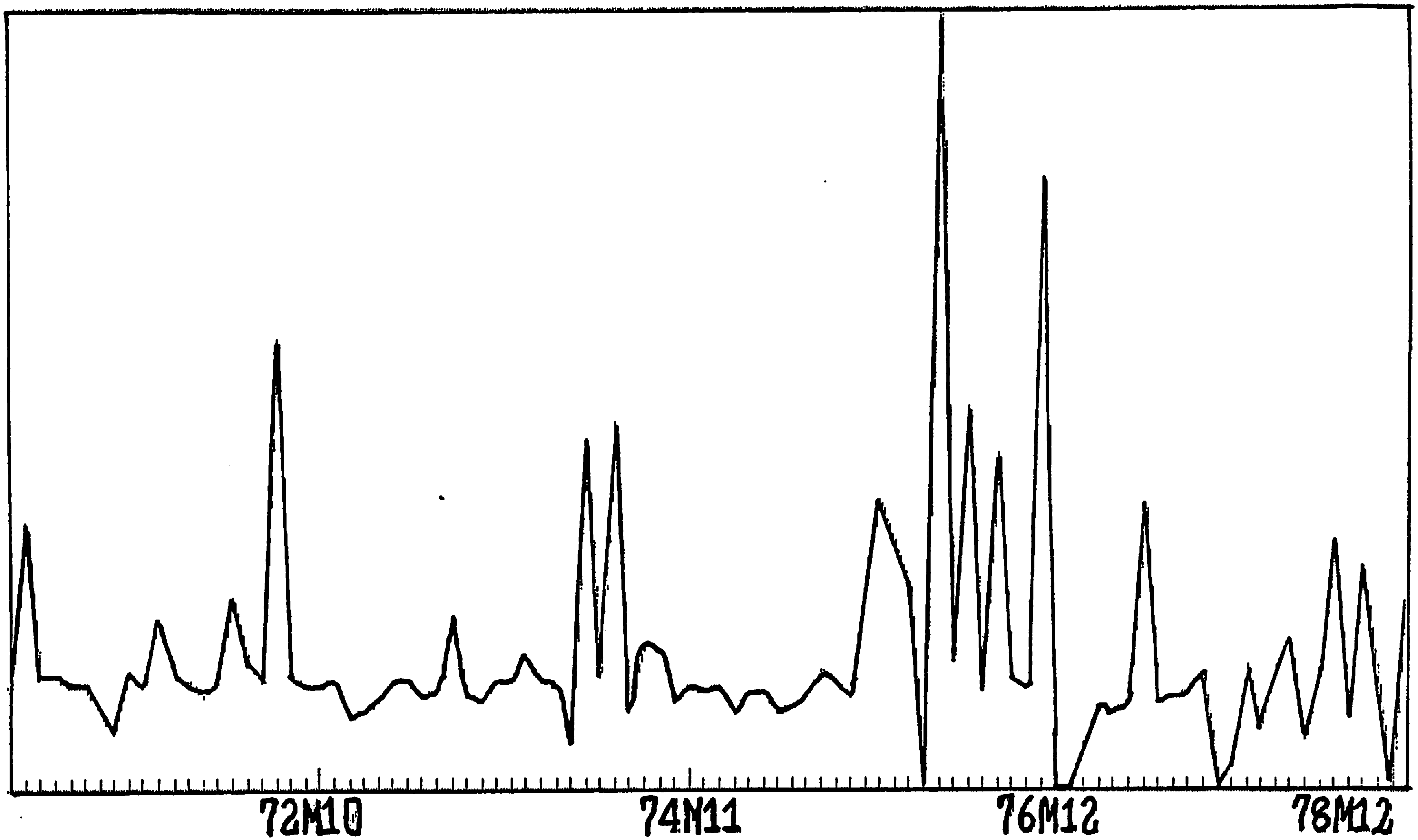
way as ϕ_t was, and figure 1 provides a graph of that series. What is most evident is the rise in σ_t^2 that occurs in 1976; the most likely explanation being the increase in uncertainty over this period due to the election of the Parti Quebeiois, whose platform emphasized the separation of Quebec from Canada.

Column three of Table 1 estimates δ by the instrumental variables estimator of Proposition 5, where the instrument for $\hat{\phi}_t^2$ is effectively $\hat{\sigma}_t^2$. In fact, Proposition 5 showed that the estimator of δ would be inconsistent, although no limiting distribution could be found. What is of most interest in this column is that the OLS estimate of δ actually exceeds the IV estimate. Although the OLS estimator inconsistency should always be negative because of the "errors in variables" problem, in this instance there is another source of bias arising from the fact that $\hat{\phi}_t^2 = \hat{e}_t^2$. If the e_t in (2) are not symmetrically distributed, the OLS estimator of δ would be inconsistent due to correlation of regressor with disturbance. The inconsistency in the OLS estimator due to this effect is positive if $E(e_t^3)$ is positive, and can easily outweigh the errors in variables bias if $E(e_t^3)-E(e_t^4) + (E(e_t^2))^2$ exceeds zero, since the sign of the inconsistency depends on that quantity. In fact, the non-parametric residuals $\hat{\phi}_t$ have a positive third moment that is some ten times greater than the fourth moment , and this seems to be the likely explanation for the relative magnitudes of the OLS and IV estimators of δ in Table 1.

5.2 Inflation Volatility and Real Effects

The real effects of high and variable inflation were a much debated issue in the early 1980's and a number of papers investigated the quantitative impact of a volatile inflation rate upon output growth, employment,

Fig.1: Conditional Variance,$US/$C,70(9)-78(12)



investment, etc. Coulson and Robins (1985) is a good representative of such studies. They argued that the rate of unemployment (R_t) could be dynamically related to the unanticipated inflation rate (ϕ_t) and an index of the volatility of inflation (σ_t^2). Specifically, they estimate the equation

$$R_t = a+\delta\sigma_t^2 +\sum_{j=1}^{8} c_j R_{t-j} +\sum_{j=0}^{6} d_j \phi_{t-j}, \qquad (15)$$

restricting c_5, c_6 and c_7 to be zero. (15) can be re-parameterized to give $\gamma_0 = \sum_{j=1}^{8} c_j$ and $\gamma_1 = \sum_{j=0}^{6} d_j$ as coefficients, and with series on σ_t^2 and ϕ_t (15) can be estimated with data over the period 1951/1 to 1979/4.

Coulson and Robins followed Engle (1983) in assuming that the inflation rate (ψ_t) was a function of a number of variables such as the money growth rate (w_t), while σ_t^2 followed an ARCH process. Estimates of ϕ_t and σ_t^2 were then made from the residuals, $\hat{\phi}_t$, and estimated variances, $\hat{\sigma}_t^2$, generated after maximum likelihood estimation of the ARCH process.

Coulson and Robins regressed R_t against a constant, $\hat{\sigma}_t^2$, R_{t-j} and $\hat{\phi}_{t-j}$ and found that δ was -.698 with a t-statistic of -3.772. To parallel Coulson and Robins' work we took the civilian unemployment series from Gordon (1984) as R_t and values for $\hat{\phi}_t$ and $\hat{\sigma}_t^2$ from Engle (1983). Column 1 of Table 2 gives "Coulson-Robins" estimates. It is apparent that, although the estimate of δ is different, the same conclusions about the impact of σ_t^2 upon R_t would be reached as in the earlier study.[10]

[10] Actually, because $\hat{\sigma}_t^2$ is a function of $\hat{\theta}$, the standard errors of $\hat{\delta}$ from a regression package will be biased - Pagan (1984, p.241) - unless account is taken of the fact that θ is estimated. However, we could make no adjustment for this as the original data Engle used could not be located.

Column 2 of Table 2 shows what would happen if $\hat{\phi}_t^2$ was used as a regressor proxying for σ_t^2. In fact, it is not unusual in investigations of this sort for the $\hat{\phi}_t$ to be the residuals from a fitted ARMA model. As predicted, the resulting estimator of δ is severely biased, and an investigator doing this regression would inevitably conclude that inflation volatility had negligible effects upon unemployment. Finally, column 3 estimates the equation of column 2 with Sargan's generalized IV estimator, where the list of instruments is the regressors with $\hat{\phi}_t^2$ replaced by $\hat{\sigma}_t^2$. Now there is once again evidence of an effect of σ_t^2 upon unemployment, emphasizing the importance of instrumenting $\hat{\phi}_t^2$ when attempting to estimate δ

<u>Table 2</u>

Estimates of the Parameters of Equation (15)*

	(1)	(2)	(3)
a	.3760	.2943	.2882
	(2.60)	(2.00)	(1.44)
δ	-.5340	-.0680	-.9194
	(2.91)	(.75)	(2.06)
γ_0	.9542	.9458	.9848
	(35.3)	(33.72)	(22.85)
γ_1	.6420	.3017	1.1527
	(3.01)	(1.59)	(2.3)

*Absolute values of t-statistics in parantheses $\gamma_0 = \Sigma c_j$, $\gamma_1 = \Sigma d_j$. Data on R_t is from Gordon (1984), $\hat{\phi}_t$ and $\hat{\sigma}_t^2$ from Engle (1983). Column (1) is OLS using $\hat{\sigma}_t^2$ in place of σ_t^2. Column (2) is OLS using $\hat{\phi}_t^2$ in place of σ_t^2. Column (3) is GIVE with $\hat{\sigma}_t^2$ replacing $\hat{\phi}_t^2$ in the list of instruments.

6. Conclusion

In this paper we have attempted to provide an integrated approach to the estimation of models with risk terms. In sections two and three it was argued that there exists orthogonality conditions between variables in the information set and higher order moments of the unanticipated variable density. These could be exploited to provide consistent estimators of the parameters associated with the risk term. Specifically, it was recommended that an IV estimator should be applied, with instruments constructed from the information set. Viewed in this way, our analysis represents an extension of current methods for the estimation of models featuring anticipations. Differences largely stem from the fact that higher order moments are involved, but these are sufficiently distinctive to justify a separate treatment.

Section 4 analysed four existing methods commonly used to estimate models with risk terms. These involved the construction of risk moving averages of time series, risk measures from relative prices and survey data, and direct parameterisation. Various problems with the use of each method were identified. Finally, section 5 used the theory of earlier sections to re-examine a few applied studies in the literature. These involved the presence of a risk term in the $US/$C exchange market, and the effects of price uncertainty upon production.

REFERENCES

Bierens, H. (1983), "Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions," *Journal of American Statistical Association*, 78, 699-707.

Bierens, H. (1985), "Kernal Estimators of Regression Functions", (paper presented at the Fifth World Congress of the Econometric Society, Boston).

Carroll, R. J. (1982), "Adapting for Heteroscedasticity in Linear Models", *Annals of Statistics*, 10, 1224-1233.

Coulson, N. E. and R. P. Robins (1985), "Aggregate Economic Activity and the Variability of Inflation: Another Look", *Economics Letters*, 17, 71-75.

Cukierman, A. and P. Wachtel (1979), "Differential Inflationary Expectations and the Variability of the Rate of Inflation: Theories and Evidence", *American Economic Review*, 69, 595-609.

Domowitz, I. and C. S. Hakkio (1985), "Conditional Variance and the Risk Premium in the Foreign Exchange Market", *Journal of International Economics*, 19, 47-66.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50, 987-1008.

Engle, R. F. (1983), "Estimates of the Variance of U.S. Inflation Based Upon the ARCH Model", *Journal of Money, Credit and Banking*, 15, 286-301.

Engle, R. F., D. M. Lillien and R. P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-Model", *Econometrica*, 55, 391-407.

Engle, R. F. and C. W. J. Granger (1987), "Dynamic Model Specification with Equilibrium Constraints: Co-Integration and Error Correction", *Econometrica*, 55, 251-276.

Frankel, J. (1985), "Portfolio Crowding - Out Empirically Estimated", *Quarterly Journal of Economics*, Vol C, Supplement, 1041-1065.

French, K.R., G. W. Schwert and R.F. Stambaugh (1987), "Expected Stock Returns and Volatility", *Managerial Economics Research Center Discussion Paper* 85-10, University of Rochester.

Friedman, M. (1977), "Nobel Lecture: Inflation and Unemployment", *Journal of Political Economy*, 85, 451-1065.

Gylfason, T. (1981), "Interest Rates, Inflation and the Aggregate Consumption Function", *Review of Economics and Statistics*, 63, 233-245.

Gordon, R. J. (1984), *Macroeconomics* (Little, Brown and Co., Boston).

Hansen, L. P. and R. J. Hodrick (1983), "Risk Averse Speculation in the Forward Foreign Exchange Market: An Econometric Analysis of Linear Models", in J. A. Frenkel (ed.), *Exchange Rates and International Macroeconomics* (Chicago: University of Chicago Press for the National Bureau of Economic Research).

Hausman, J. (1978), "Specification Tests in Econometrics", *Econometrica*, 46, 1251-1272.

Huizinga, J. and F. S. Mishkin (1985), "Monetary Policy Regime Shifts and the Unusual Behavior of Real Interest Rates", paper given to the Carnegie-Rochester Conference on Public Policy.

Hercowitz, Z. (1981), "Money and the Dispersion of Relative Prices", *Journal of Political Economy*, 89, 328-356.

Ibrahim, I.B. and R. Williams (1978), "Price Unpredictability and Monetary Standards: A Comment on Klein's Measure of Price Uncertainty", *Economic Inquiry*, 16, 431-437.

Klein, B. (1977), "The Demand for Quality-Adjusted Cash Balances: Price Uncertainty in the U.S. Demand for Money Function", *Journal of Political Economy*, 85, 691-715.

Lawrence, C. (1983), "Rational Expectations, Supply Shocks and the Stability of the Inflation-Output Tradeoff", *Journal of Monetary Economics*, 11, 225-245.

Leland, H. E. (1968), "Saving and Uncertainty: The Precautionary Demand for Saving", *Quarterly Journal of Economics*, 82, 465-473.

Longworth, D. (1981), "Testing the Efficiency of the Canadian-U.S. Exchange Market Under the Assumption of No Risk Premium", *Journal of Finance*, 36, 43-49.

Levi, M. and J. Makin (1979), "Fisher, Phillips, Friedman and the Measured Impact of Inflation on Interest", *Journal of Finance*, 34, 35-52.

Lucas, R. E. (1973), "Some International Evidence on Output-Inflation Tradeoffs", *American Economic Review*, 63, 326-334.

Malkiel, B. G. (1979), "The Capital Formation Problem in the United States", *Journal of Finance*, 34, 291-306.

Mascaro, A. and A. H. Meltzer (1983), "Long and Short-Term Interest Rates in a Risky World", *Journal of Monetary Economics*, 12, 485-518.

Nadaraya, E. A. (1964), "On Estimating Regression", *Theory of Probability and its Applications*, 9, 141-142.

Nelson, C. R. (1975), "Rational Expectations and the Estimation of Econometric Models", *International Economic Review*, 16, 555-561.

Neuey, W. and K. West (1986), "A Simple Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *N.B.E.R. Technical Paper No.* 55.

Pagan, A. R. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors", *International Economic Review*, 25, 221-247.

Pagan, A. R., A. D. Hall and P. K. Trivedi (1983), "Assessing the Variability of Inflation", *Review of Economic Studies*, 50, 585-596.

Pindyck, R. S. (1984), "Risk, Inflation and the Stock Market", *American Economic Review*, 74, 335-351.

Phillips, P. C. B. (1985), "Time Series Regression with Unit Roots", *Cowles Foundation Discussion Paper No. 740, Yale University*.

Phillips, P.C.B. and S.N. Durlauf (1986), "Multiple Time Series with Integrated Variables", *Review of Economic Studies*, 53, 473-496.

Poterba, J. M. and L. H. Summers (1986), "The Persistence of Volatility and Stock Market Fluctuations", *American Economic Review*, 76, 1142-1151.

Robinson, P. M. (1983), "Non-parametric Estimators for Time Series", *Journal of Time Series Analysis*, 4, 185-207.

Robinson, P. M. (1986a), "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form" (Mimeo, London School of Economics).

Robinson, P. M. (1986b), "Nonparametric Methods in Specification," *Economic Journal*, supplement, 96, 134-141.

Sargan, J. D. (1958), "The Estimation of Relationships Using Instrumental Variables", *Econometrica*, 26, 393-415.

Shiller, R. J. (1981), "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?", *American Economic Review*, 71, 421-436.

Sims, C. A., J. H. Stock and M. W. Watson (1986), "Inference in Linear Time Series Models with Some Unit Roots", (Mimeo, Stanford University).

Singh, R. S. and A. Ullah (1985), "Nonparametric Time Series Estimation of Joint DGP, Conditional DGP and Vector Autoregression", *Econometric Theory*, 1, 27-52.

Singh,R.S. and A.Ullah (1987),"Corrections",(mimeo,University of Western Ontario)

Stock, J.H. (1984), "Asymptotic Properties of Least Squares Estimators of Co-Integrating Vectors", mimeo, Harvard University.

Stockman, A. (1978), "Risk Information and Forward Exchange Rates", in J. A. Frenkel and H. G. Johnson (eds.) *The Economics of Exchange Rates*, Reading, Massachusetts: Addison-Wesley, 1978.

Ullah, A. (1986), "Non-Parametric Estimation of Econometric Functionals", (Mimeo, University of Western Ontario).

Vanderhoff, J. (1983), "Support for Rational Expectations Models with U.S. Data", *Journal of Monetary Economics*, 12, 297-308

Watson, G. S. (1964), "Smooth Regression Analysis", *Sankhya*, Series A, 26, 359-372.

White, H. (1980), "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity", *Econometrica*, 48, 817-838.

White H. (1984), *Asymptotic Theory for Econometricians* (New York: Academic Press).

Wickens, M. R. (1982), "The Efficient Estimation of Econometric Models with Rational Expectations", *Review of Economic Studies*, 49, 55-68.