

Non-Parametric Estimation and the Risk Premium

Pagan, Adrian and Y. Hong

Working Paper No. 135
May 1988

University of
Rochester

Non-Parametric Estimation and the Risk Premium¹

A. Pagan and Y. Hong

University of Rochester

April, 1988

1. Introduction

In the last decade there has been an apparent increase in the volatility of many financial variables, with quite dramatic movements in some series such as Treasury Bill rates and equity prices. Since many models predict that economic behavior would be modified in response to such volatility, a lot of attention has been paid to the measurement of the associated risk premium. This literature is vast, but samples of it would be Merton (1980), Pindyck (1984) and Summers and Poterba (1986).

No consensus has emerged on the best way to measure what is fundamentally an unobserved variable, with many approaches to be found in the literature, some of which have been critically surveyed by Pagan and Ullah (1988). Nevertheless, there has been a noticeable tendency in the last five years to opt for the parametric formulation set out in Engle (1982), known as the ARCH model, and, if there is a "standard", this would be it. One cannot be entirely comfortable with this direction, as parameterization in this context does not seem as convincing as in other areas, mainly because there is no optimizing theory to indicate exactly what function the risk premium should be of the information available to agents. In fact, the nature of the dependence is essentially a characteristic of the data generating process.

¹Research supported by NSF grant number SES-8719520

This paper eschews parametric modeling of the risk premium in favor of a non-parametric approach originally suggested in Pagan and Ullah (1988). Section 2 describes the issues and argues that past research and theorizing from simple models points to the need for flexibility when relating risk to agents' information sets. Section 3 details the estimation techniques utilized in the remaining sections. In sections 4, 5 and 6 attention is paid to three applied studies that have utilized the ARCH technology above - these being French et al (1987) on the equity premium, Engle et al (1987) on the excess holding yield for Treasury Bills, and Hodrick (1987) on volatility in output. We ask if the non-parametric approach provides a superior method for the modeling of risk, concluding that there is strong evidence that this is so. Certainly, the non-parametric methods point out deficiencies in the existing studies. Section 7 then sets out the implications of the paper.

2. The Shape of the Risk Premium

Many models in which risk is important ultimately reduce to a linear relation between some variables y_t, x_t and a variable σ_t^2 that reflects the level of risk associated with decisions about y_t . Equation (1) captures such a relation

$$y_t = x_t \beta + \sigma_t^2 \delta + e_t, \quad (1)$$

where e_t will be taken as n.i.d. $(0, \sigma^2)$ and x_t will be weakly exogenous for β and δ . Of course the assumption about e_t can be relaxed. The variable σ_t^2

typically represents the moments of a variable ψ_t , conditional upon some information set \mathcal{F}_t . Sometimes it is the conditional variance of ψ_t , sometimes a covariance, e.g. in Hansen and Hodrick (1983) there are three " σ_t^2 's" involving the variance of the change in the spot rate and the covariances of that variable with the inflation and endowment growth rates. In the linear format of (1), the parameters of interest are δ and β . This may be either because of interest in the magnitude of these coefficients or perhaps because of a desire to test if they equal specified values such as zero. Moreover, any measure of the impact of a risky environment upon y_t would be concerned with the size of $\delta E(\sigma_t^2)$, so that measuring δ is important. Thus this paper has as its central premise the fact that we wish to estimate δ .

It has to be admitted that a linear model such as (1) is not always the outcome of some optimizing model. In fact such an outcome requires fairly strong assumptions to get (1) e.g. the CRRA utility functions allied with log normality of errors as in Hansen and Hodrick, or mean/variance portfolio analysis with fixed portfolio shares. More generally we might just have some unknown function of the information set in place of the linear form above. In this case it might be the derivative of y_t with respect to the information variables which should be computed. However, we ignore that case at the moment, in the spirit of trying to study the estimation problems posed by the presence of risk in as simple an environment as possible. Thus our aim is to try to improve upon some current procedures for allowing for risk rather than getting immersed in deeper issues.

The main difficulty in estimating (1) is to know exactly how σ_t^2 varies with the elements in \mathcal{F}_t , i.e. what the "shape" of the risk premium is. Most

existing research has been concerned with the "level" of σ_t^2 . Mehra and Prescott (1985) in their famous "equity premium puzzle" paper found that the magnitude of the risk premium in a simple asset pricing model was far below what had been observed historically (.33% versus 6.3%), and later research has been focused upon the robustness of this conclusion to changes in the specification of their model. Reitz (1987) for example managed to increase the magnitude of the premium substantially by allowing for a "crash state". This state occurred with low probability but was accompanied by such severe output contraction that it made a very important contribution to the risk premium.

One way to examine the "shape" of the risk premium is to follow the lead of this literature and to simulate artificial economies wherein the risk premium can be determined exactly. An interesting paper along these lines is Backus, Gregory and Zin (1986). These authors construct a model with an endowment exhibiting a growth rate λ_t that features two states ($i = 1, 2$), "low" and "high", with the probability of a transition from state i to state j being p_{ij} ($i = 1, 2; j = 1, 2$). Using standard asset pricing theory they determine the price of an n -period risk free (discount) bond which delivers one unit of the commodity at $(t+n)$ in all states as

$$q_{t+n,t} = E_t \left[\prod_{i=1}^n m_{t+i} \right] \quad (2)$$

where $m_{t+i} = \gamma U'(O_{t+i})/U'(O_t)$, O_t is the endowment level, γ the discount rate and $u(\cdot)$ is a utility function. From (2) they define a one-period forward price $f_{t+1,t} = q_{t+2,t}/q_{t+1,t}$, arriving at a risk premium $\sigma_t^2 = E_t(q_{t+2,t+1}) -$

$f_{t+1,t}$. This last relation can be cast in the form of (1) by defining $y_t = q_{t+2,t+1}$, $x_t = f_{t+1,t}$, $\beta = 1$, $\delta = 1$. σ_t^2 is in fact $\text{cov}(m_{t+1}q_{t+2,t+1})/q_{t+1,t}$.

Adopting a CRRA utility function $U(0_t) = 0_t^{1-\alpha}/(1-\alpha)$, the determinants of the risk premium are the rates of growth of the endowment, the transition probability matrix, the degree of dependence in endowment growth rates, and the risk aversion coefficient (α). After these values are set it is possible to simulate the risk premium in this economy. For a two state economy with the unconditional probability of being in either state being 0.5 and conditional probabilities $p_{ij} = (1-\theta)0.5 + \theta\delta_{ij}$, the risk premium is derived by Backus et al to be proportional to $\theta(1-\theta^2)b^2$, where b is the standard deviation of $m_t = \gamma U'(0_{t+1})/U'(0_t)$. θ is an index here of the degree of persistence in the series, and when it equals zero states are independent over time making the risk premium zero.

Choosing the same values of the parameters as Backus et al ($\theta = -0.3$, $\alpha = 8$), but with eight rather than two states, we have computed the shape of the risk premium as a function of the state variable ($\lambda_t = 0_{t+1}/0_t$). Fig. 1 shows that it is well described as a quadratic function of the state, and this conclusion proved to be remarkably resilient to variations in the values of parameters of the model. To emphasize this, the regression of the risk premium against a quadratic polynomial in λ_t always gave an R^2 around .98 to .99. This is a striking result which highlights the importance of allowing for a quadratic term in any models of the risk premium. Unfortunately, it is a result that is of more limited value than one might suppose, since the risk premium only exhibits variation if there is some dependence within the endowment growth process, and as observed by Backus et al there is little evidence of that in actual data.

An alternative viewpoint espoused here is that these models are not rich enough to capture the essence of a risk premium, as their structure is incompatible with the presence of any conditional heteroskedasticity in the endowment growth process, thereby forcing the relation between the risk premium and state variables to depend solely upon the degree of dependence in the first moment. To avoid this problem consider the two state case with endowment growth rates of λ_1 and λ_2 in the low and high states. Suppose that λ_j is uniformly distributed over $(a_j, 3a_j)$, making the mean of λ_j be $2a_j$. Using the CRRA utility function, and amending (2) to allow for the fact that λ_j is a random variable, means that $q_{t+1,t} = \sum_j \gamma p_{ij} E(\lambda_j^{-\alpha})$. When a_j is fixed $q_{t+1,t}$ is a constant and one obtains the same conclusion as Backus et al that $\theta = 0$ implies a zero risk premium. However, if a_j depends upon the state, then $q_{t+1,t}$ will vary with the state and so will the risk premium. The logic of this re-specification is that the states are "low" and "high", but the magnitudes of growth in each state might vary. On average one might expect that growth in (say) the "high" state would be lower if the previous state was "low" than if it was "high". Such a formulation, a_j depending upon the states, appears to be a more realistic description of actual data, and it induces a mapping between the premium and the states even when $\theta = 0$. Moreover, the exact mapping will depend on the relation of a_j and the states, and this is *exogenous* to the model. Because of this exogeneity, little can be learned from these artificial economies about the structure of the risk premium, although, as Backus et al demonstrate, they can be very useful for understanding econometric results.

An alternative strategy is to infer a relation between σ_t^2 and \mathcal{F}_t based upon past empirical work. By far the most popular mapping has been that of the ARCH process pioneered by Engle (1982), of which a simple variant is $\sigma_t^2 = E[(\psi_t - E(\psi_t | \mathcal{F}_t))^2 | \mathcal{F}_t] = \alpha_0 + \alpha_1(\psi_{t-1} - w_{t-1}\gamma)^2$ where $E(\psi_t | \mathcal{F}_t) = w_t\gamma$. More lags in the error process $(\psi_{t-1} - w_{t-1}\gamma)$ can be added either explicitly or implicitly through the GARCH process in Bollerslev (1986). If \mathcal{F}_t is defined to include ψ_{t-1} and w_{t-1} , the ARCH model clearly makes the risk premium a quadratic function of the state, and it might be expected therefore that it would form the basis of a reasonable approximation to many σ_t^2 . As formulated above, and as usually implemented in practice, it has the disadvantage of assuming that $E(\psi_t | \mathcal{F}_t)$ is linear in the w_t . As there is very little reason to think that agents actually do process information in such a restricted way, it seems inconsistent to be looking for non-linearities in the mapping between σ_t^2 and \mathcal{F}_t but to assume that the first moment is linear in w_t . Close attention must be paid to the modeling of the conditional mean since any failure to specify it correctly means that there will be more variation attributed to the conditional variance than it actually exhibits.

ARCH processes have been fitted to many series and used to model risk in quite a few contexts e.g. Engle et al (1987), French et al (1987), Hodrick (1987). Very few of these papers have attempted to determine if the risk premium actually does follow a quadratic function of the type implied by ARCH, although some diagnostic tests of this reported in French et al (1987) and Pagan and Sabau (1987) suggest that the ARCH model does not capture the risk premium completely and that a more flexible approach to the modeling of the relation between σ_t^2 and \mathcal{F}_t is called for.

If the quadratic form is thought deficient it is important to ask what type of features might need to be built into any assumed mapping between σ_t^2 and \mathcal{F}_t . One suggestion is to seek guidance from the "stylized facts". Nelson (1987) has done this arguing "that the most serious limitation of ARCH models of asset pricing is the assumption that only the size and not the sign of excess returns determine future σ_t^2 ". He cites evidence from Black (1976) and Christie (1982) that the volatility of stock returns tends to fall when prices are rising and to rise when prices are falling. Hence σ_t^2 should be a function of the sign (ψ_{t-1}) and he assumes a non-linear relation of the form (in the context of (1))

$$\sigma_t^2 = \exp[\alpha_0 + \alpha_1 \epsilon_{t-1} + \alpha_2 |\epsilon_{t-1} - (2/\pi)^{1/2}|] \quad (3)$$

where $\epsilon_t = \sigma_t^{-1} e_t$ is $\mathcal{N}(0,1)$ conditional upon \mathcal{F}_t (in fact he has a much more complex distributed lag, but our aim here is to emphasize the asymmetry).

All of the above emphasizes the need, when modeling the risk premium, to entertain a wide variety of functional forms when considering how σ_t^2 is connected to \mathcal{F}_t . Parametric methods such as ARCH imply a particular type of relationship, and one must be careful about how good the approximation they represent is for any data set. In the following sections we investigate the ability of non-parametric methods to aid in the modeling of risk.

3. Estimation Methods for the Risk Premium

Section 2 recounted the difficulties in deciding how σ_t^2 varies with \mathcal{F}_t and the range of proposals should alert us to the need for exercising caution

when estimating the parameters in (1). Importantly, estimators must be chosen which are robust to this uncertainty. Despite this fact most work reported in the literature has parameterized σ_t^2 and proceeded to estimate β and δ by M.L., for example Engle et al (1987). The choice of the M.L. estimator is unfortunate because it will certainly not be robust to errors made in specifying a functional form for σ_t^2 , as σ_t^2 is a constituent part of the conditional mean of y_t .

The issue was discussed in some detail in Pagan and Ullah (1988), ending with the recommendation that instrumental variables should be the preferred estimator. The basic idea advanced in that paper was to replace σ_t^2 by $\phi_t^2 = E(\psi_t - E(\psi_t | \mathcal{F}_t))^2$, and to then consistently estimate β and δ by applying IV with instruments constructed from \mathcal{F}_t (including x_t as its own instrument). Of course instrument construction needs to be done carefully, since they must be as highly correlated with ϕ_t^2 as possible. In fact, because $E(\phi_t^2 | \mathcal{F}_t) = \sigma_t^2$, σ_t^2 appeals as a good instrument, making it desirable to estimate σ_t^2 for this purpose, independently of any interest one might have in σ_t^2 *per se*. An important facet of the IV strategy is that it only requires $m_t = E(\psi_t | \mathcal{F}_t)$ to be estimated accurately; truncation of the instrument set to (say) $\bar{\sigma}_t^2 = E(\phi_t^2 | \mathcal{F}'_t)$, where $\mathcal{F}'_t \in \mathcal{F}_t$, does not affect the consistency of the estimators of β and δ , although efficiency might be affected if $\bar{\sigma}_t^2$ has only a weak correlation with ϕ_t^2 . Overall, there is a trade-off between maximizing instrument correlation and restricting \mathcal{F}'_t to be small enough to get estimates $\bar{\sigma}_t^2$ that are not "too noisy".

Basically the IV approach requires the computation of m_t and some estimate of σ_t^2 (although as noted above it is not crucial that the latter be

accurate). When $\psi_t \neq y_t$, m_t might be linear in the members of \mathcal{F}_t , and this is a frequent assumption in VAR modeling. However, it seems inconsistent to restrict the conditional mean to be linear in \mathcal{F}_t but to make the conditional variance a non-linear function of \mathcal{F}_t , and therefore it is desirable to allow for some non-linearity in the mapping. When $\psi_t = y_t$, m_t will normally be a non-linear function of \mathcal{F}_t as it is determined by σ_t^2 (see (1)). Hence, it is advisable to allow for a general dependence of m_t upon \mathcal{F}_t by employing non-parametric estimation methods to determine m_t . With such an estimate \hat{m}_t , $\hat{\phi}_t^2 = (\psi_t - \hat{m}_t)^2$ replaces σ_t^2 and an instrument is used for $\hat{\phi}_t^2$. In later sections we use as instruments the non-parametric estimate of the conditional variance of ψ_t given \mathcal{F}'_t , sometimes making $\mathcal{F}'_t = \mathcal{F}_t$.

Now although neither \hat{m}_t nor $\hat{\sigma}_t^2$ is likely to be estimated very precisely unless $\dim(\mathcal{F}_t)$ is small or the number of observations is large, these quantities are only being used as "regressors" in (1), and so the IV estimator of β and δ should be root-N consistent and asymptotically normal. To get this result, following Pagan and Ullah(1988) it is necessary to show that $N^{-1/2} \sum z_t (\hat{\phi}_t^2 - \phi_t^2) \xrightarrow{P} 0$, where z_t are the chosen instruments. When the conditional mean is parameterized as $w_t \gamma$ this is true, as discussed in Pagan(1984) and exploited in tests for heteroskedasticity. Moreover, if $\psi_t \neq y_t$, and the conditioning elements in $E(\psi_t / \mathcal{F}_t)$ are strictly exogenous, it also holds in this case, provided the non-parametric estimator of the conditional mean is uniformly consistent and ϕ_t, z_t obeys some mixing conditions. But when $\psi_t = y_t$ we have not yet been able to construct a complete proof of the result, and hence just conjecture that it is true generally. An alternative estimator, if $\mathcal{F}'_t = \mathcal{F}_t$, might be to regress y_t against x_t and $\hat{\sigma}_t^2$, as this would give a consistent

estimator of δ , β , but the standard errors would need to be adjusted because of a "generated regressor" bias arising from the use of $\hat{\sigma}_t^2$ rather than σ_t^2 . Just as in the case when $\hat{\sigma}_t^2$ is constructed parametrically from \mathcal{F}_t , Pagan (1984), the appropriate adjustment may be difficult to do directly, but is easily done with the IV procedure. Hence, estimation and inference is naturally performed within the IV framework.

The final issue to be addressed concerns which non-parametric method should be used to construct \hat{m}_t and $\hat{\sigma}_t^2$. In later sections we employ both the kernel based approach and the Fourier method of Gallant (1982). Both are easy to apply and have advantages and disadvantages. It should be noted that for the Fourier technique, $\hat{\sigma}_t^2$ is estimated as the predictions obtained by fitting a Fourier approximation to $(\psi_t - \hat{m}_t)^2$, since $E(\psi_t - m_t)^2 = \sigma_t^2$ is the basis of such a regression.

4. Risk and the Stock Market

There has been a lot of debate over the causes of the poor performance of the share market in the 1970's and a number of explanations have been proposed. Perhaps the explanation that has generated more controversy than any other has been the contention that the stock market changes reflected an increase in the variability of the marginal returns to capital. Pindyck (1984) claimed that this latter quantity could be inferred from the variances of the observed stock price and price level changes, and that it was the right order of magnitude to explain what had occurred.

Pindyck argued that for the parameters of his pricing model most of the variance of the returns to capital was reflected in the variance of nominal stock returns, and his graphs point to a strong correlation. Subsequently, a literature developed in which the stock premium was related to an index of the volatility of these returns (σ_t^2). Pindyck (1984) estimated σ_t^2 by taking a twelve month moving average of the squared changes in the equity premium and motivated his paper by regressing returns against this estimate. As argued in Pagan (1984) this must yield an inconsistent estimator of the effects of volatility since the regressor must be correlated with the error term producing an errors in variables bias. Summers and Porterba (1986) also use observed data to construct $\hat{\sigma}_t^2$. In their case they average the daily changes in the stock price to give a monthly series for σ_t^2 . This also has an errors in variables problem but, as pointed out by French et al (1987), it is of small magnitude. Whether the daily volatility is a suitable proxy for monthly volatility is, however, much more questionable.

Rather than attempt to measure σ_t^2 directly other authors have tried to parameterize it. Gennotte and Marsh (1985) made σ_t^2 a function of the squared change in the stock price while French et al write the equity premium ($R_m - R_f$), where R_m is the market and R_f is the risk free rate of return, as²

$$R_m - R_f = \delta \sigma_t^2 + e_t \quad (4)$$

²In fact e_t is an MA(1) in their case but as we will be using monthly data over a period for which their estimated MA coefficient was small and insignificant we have deleted it.

and make σ_t^2 follow Bollerslev's (1986) GARCH process

$$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2. \quad (5)$$

To see the implication of this assumption set $\beta_0 = 0$, $\beta_1 = 0$, $\alpha_2 = 0$, $y_t = R_m - R_f$ and substitute (5) into (4) to get

$$y_t = \delta \alpha_1 e_{t-1}^2 + e_t \quad (6)$$

$$= \delta \alpha_1 (y_{t-1} - \delta \sigma_{t-1}^2)^2 + e_t \quad (7)$$

$$= \delta \alpha_1 y_{t-1}^2 - 2\delta^2 \alpha_1 \sigma_{t-1}^2 y_{t-1} + \delta^3 \alpha_1 \sigma_{t-1}^4 + e_t \quad (8)$$

and after continued substitution for σ_{t-1}^2 from (5) etc.

$$y_t = \delta \alpha_1 y_{t-1}^2 - 2\delta^2 \alpha_1^2 y_{t-2}^2 y_{t-1} + 4\delta^3 \alpha_1^2 \sigma_{t-2}^2 y_{t-2} y_{t-1} + \dots, \quad (9)$$

from which it is clear that σ_t^2 is a complicated non-linear function of an infinite set of lagged values of y_t , albeit with declining weights. It is interesting to observe that the presence of terms such as y_{t-1} implies that σ_t^2 responds to the sign of the lagged premium as well as the size, and so it goes some way toward meeting the stylized facts identified earlier with Christie and Black. Given the complexity of the non-linear AR implied by σ_t^2 following a GARCH process, one might expect that a sufficiently good approximation to σ_t^2 would be found by relating σ_t^2 to a finite number of lags of y_t . Much depends

however upon the size of $(\delta\alpha_1)$ and β_1 . A high β_1 will impose a lot of dependence upon σ_t^2 and hence a large number of elements would need to enter the conditioning function.

Now the GARCH specification for σ_t^2 is a complex one, but there is no reason to think that the precise non-linear relation between y_t and $\{y_{t-1}, y_{t-2}, \dots\}$ that it imposes upon the data is correct. If incorrect a specification error in σ_t^2 would result in an inconsistent estimator of δ if δ is found by regressing y_t against x_t and $\hat{\sigma}_t^2$. Consequently, a non-parametric estimator of σ_t^2 appeals as a solution to this dilemma, although the fact that it operates with only a finite number of conditioning elements makes it unable to explicitly handle a GARCH type process. Nevertheless, by allowing the information set to expand with the sample size one might be able to establish consistency of the estimator. As far as we are aware, however, there are no current theorems that would justify such a conjecture.

Choosing q lags of y_t (the equity premium) as the conditioning elements, $\hat{\sigma}_t^2$ was computed as the conditional variance of y_t using a Gaussian kernel with diagonal covariance matrix given by the sample variances of each of the variables multiplied by $T^{-1/(4+q)}$. Hence the bandwidth is proportional to $T^{-1/(4+q)}$. Data matches that in French et al (1987), with y being defined as the value weighted yield on equities in the S&P composite portfolio including dividend on the New York Stock Exchange less the 3 month Treasury Bill rate.³ Initially q was set to 4 as an attempt to capture the GARCH type process used by French et al.

³We are grateful to William Schwert for providing this data and for his advice at various times during work on this data.

Once $\hat{\sigma}_t^2$ is estimated δ could be estimated in two ways. First, y_t can be regressed against a constant and $\hat{\sigma}_t^2$. Second, letting \hat{m}_t be the non-parametric conditional mean estimator, $\hat{e}_t = y_t - \hat{m}_t$, and we can estimate an equation $y_t = \gamma + \delta \hat{e}_t^2 + u_t$, with $\hat{\sigma}_t^2$ as an instrument for \hat{e}_t^2 . The rationale for this latter estimator is given in Pagan and Ullah (1988). Table 1 gives the estimates of γ and δ for monthly data for 1953/1-1984/12, so being directly comparable to French et al's Table 6.a (p. 19).

Table 1

Estimate of Effect of Volatility,
Non-parametric Estimation of Volatility

	<u>OLS</u>		<u>IV</u>	
	γ	δ	γ	δ
Estimate	.0062	-.73	.0079	-1.42
abs.t.	(1.56)	(.30)	(.86)	(.30)
Hetero.Cons.t	(1.48)	(.25)	(.75)	(.25)

The standard errors used in Table 1 from the least squares regression are not accurate since both $\hat{\sigma}_t^2$ and \hat{m}_t are "generated regressors", but they are probably correct for the instrumental variable estimator under conditions that are likely to be satisfied here.

What is most striking about Table 1 is the negative effect of volatility about the equity premium, which is a contrast to the positive effects claimed by Pindyck and found by French et al. Based on Merton (1980), the coefficient

δ would be the coefficient of risk aversion and so a positive value is expected. Some explanations of this difference seem desirable and we examine three of these here.

First, as observed above, it is possible that σ_t^2 is really GARCH and so four conditioning variables are insufficient. If this is true it will be hard for the non-parametric estimator to estimate $\hat{\sigma}_t^2$ effectively as the bandwidth will become too large. However, the reason for the GARCH choice by French et al was because the squared equity premium exhibited an autocorrelation function which died away slowly. Since σ_t^2 enters the conditional mean of y_t , y_t^2 cannot be used to make inferences about the nature of the conditional variance of e_t , and it is possible that the autocorrelation function of y_t^2 dies away slowly since it has not accounted for the presence of $\delta\sigma_t^2$. In fact, our prior belief was that it was unlikely that more than one or two lags of y_t would influence σ_t^2 . Hence, we took the opposite extreme to French et al and set $q = 1$. Figures 2 and 3 show the non-parametric mean and variance of y_t plotted against y_{t-1} . When interpreting these and later figures it needs to be borne in mind that at the end points there will be an understatement of estimates since data is only available on one side of the point. Nevertheless the results are very interesting, showing that the mean tends to be slightly non-linear with y_{t-1} , whereas the variance exhibits a non-linearity that would not be well described by the quadratic format implicit in GARCH. In fact, as evident from fig. 3, σ_t^2 is larger when the premium is negative than when it is positive, and this is evidence of the asymmetry noted earlier from the work of Black and Christie. Table 2 gives the first sixteen autocorrelation coefficients of the non-parametric residuals \hat{e}_t^2 (when $q = 1$) along with their 't statistics'.

Table 2
Autocorrelation Function of Squared Non-parametric
Residuals (q = 1)*

<u>Lag</u>	<u>a.c.f.</u>	<u> t </u>	<u>Het. t.</u>	<u>Lag</u>	<u>a.c.f.</u>	<u> t.</u>	<u>Het. t.</u>
1	.0005	.009	.008	9	.0625	1.19	1.51
2	.1038	2.03	1.61	10	-.0240	.46	.54
3	.1556	3.06	1.74	11	.0550	1.04	.64
4	.0765	1.47	1.31	12	.1040	1.97	1.36
5	.0413	0.79	0.91	13	-.0269	0.51	0.80
6	.1019	1.95	1.72	14	.0292	.55	.35
7	-.0331	0.63	1.09	15	.0328	.62	.41
8	.0658	1.25	1.26	16	-.0307	.58	.62

*The estimates of the autocorrelation coefficients (a.c.f.) were obtained by regressing \hat{e}_t^2 against a constant and \hat{e}_{t-j}^2 ($j = 1, 2, \dots, 16$) one at a time. The t-statistics are from these regressions as are the heteroskedastic consistent versions.

There is very little evidence in Table 2 of the GARCH effects observed by French et al in y_t^2 and we are led to the conclusion that long lags in y_t are unnecessary for the modeling of σ_t^2 . Putting the $\hat{\sigma}_t^2$ constructed when $q = 1$ into the regression produces virtually identical results to the $q = 4$ case. Since we are entering $\hat{\sigma}_t^2$ into a regression the rate of convergence should be

as $T^{1/2}$, and this illustrates that the much more inaccurate estimate of σ_t^2 obtained if q is set to 4 when it is actually unity (owing to the slower rate of convergence of $\hat{\sigma}_t^2$ to σ_t^2) gets "smoothed over" in the regression. We generally found this to be true in all the following analyses. What was more important, however, was whether in computing \hat{m}_t one should drop the t 'th observation from the formula for computing the non-parametric mean. Because the kernel gives relatively large weight to that point when the corresponding y_t is an outlier, the estimate of \hat{m}_t lies close to y_t . Hence the residual \hat{e}_t is small and there is much less volatility predicted than when the t 'th observation is excluded (because \hat{e}_t is smaller). In all empirical work below the t 'th observation was always dropped. It might be mentioned, however, that for this application the conclusion about the sign and significance of $\hat{\delta}$ is invariant to the inclusion of this observation.

A second possible explanation for why we should get a negative estimate of δ when in fact it is positive is that our non-parametric approach gives poor estimates, especially because of its use of a finite number of conditioning variables. As French et al estimated a GARCH model we decided to check this hypothesis by conducting a Monte Carlo simulation of the non-parametric estimator of the model in (4) and (5) utilizing the estimated parameter values given by French et al for 1953(1)-1984(12).

($\gamma = -.0064$, $\delta = 7.81$, $\beta_0 = .000167$, $\beta_1 = .751$, $\alpha_1 = -0.019$, $\alpha_2 = .168$).

Estimated over 50 replications the mean of $\hat{\delta} = 4.12$ (OLS), 5.28 (IV)

highlights a severe downward bias in the estimators and emphasizes how hard it will be to exploit non-parametric procedures for modeling GARCH processes.

However, what is more relevant to the current situation is that in no

replication was an estimated value of δ negative, so it is very hard to ascribe a negative $\hat{\delta}$ to sampling problems with the kernel estimator.

A final answer is that the specification for σ_t^2 or y_t employed by French et al is incorrect. This point has already been alluded to above in our comment that there does not seem to be a need for long lags in y_t for modeling σ_t^2 , but it is possible to make a more direct test of the validity of their model. In Pagan and Sabau (1987) it was argued that specification tests of the Newey (1985) and Tauchen (1985) variety could be usefully applied to these models. In this case part of the maintained hypothesis is that $E(\sigma_t^2(e_t^2 - \sigma_t^2)) = 0$, since σ_t^2 is a function of past information and $e_t^2 - \sigma_t^2$ should be a martingale difference with respect to the sigma field generating it. Substituting the GARCH residuals (\tilde{e}_t) and conditional variance ($\tilde{\sigma}_t^2$) we can test that the conditional moment restriction above holds by regressing $\tilde{e}_t^2 - \tilde{\sigma}_t^2$ against $\tilde{\sigma}_t^2$ and a constant and testing if the coefficient on $\tilde{\sigma}_t^2$ is zero. In fact the t-statistic was 16.42, and after a correction for heteroskedasticity it is 8.77, indicating substantial mis-specification.

The source of all of these problems is located in the specification of the conditional mean of the equity premium. Computed decade means reveal that this variable has steadily declined since the 1950's. In contrast $\hat{\sigma}_t^2$ by French et al's measure and also our non-parametric one has risen. Figure 3 shows why this is so. Since the equity premium has become smaller over time volatility has risen. Consequently if we regress the equity premium against $\hat{\sigma}_t^2$ we must get a negative coefficient for $\hat{\sigma}_t^2$. Of course this need not be true for French et al as they are effectively regressing the equity premium weighted by $\hat{\sigma}_t^{-2}$ against various weighted lags of the premium, and the behavior

of these variables can be quite different. French et al found that when they resorted to a regression type approach negative values of $\hat{\delta}$ emerged, and this points to the fact that employing an MLE estimator can be a very poor strategy unless very close attention is paid to the detection of specification errors. Exactly what variable should be added to the equity premium equation remains to be seen, although there are many suggestions in the literature. One has been to add the Treasury Bill rate to the equation as Fama and Schwert (1977) found a negative correlation between the equity premium and this rate. There are conceptual difficulties with doing this, since tests show that the equity premium seems an I(0) process whereas the Treasury Bill rate is I(1). However, even when we add it to our regressions σ_t^2 still has a negative sign.

5. The Excess Holding Yield on Treasury Bills

A number of authors have concentrated upon an explanation of the excess holding yield (y_t) on 3 month Treasury bills (r_S) relative to 6 month bills (r_L). Mankiw and Summers (1984) demonstrated that the excess yield was a function of the yield differential, and this dependence would seem to contradict the efficient markets hypothesis. An alternative viewpoint might be that the yield differential was representing a risk premium. Engle,

Lillien and Robins (ELR)(1987) therefore made y_t a function of the yield differential and a term capturing volatility σ_t^2 .⁴ Their model was

$$y_t = \beta_0 + \beta_1(r_L - r_S) + \delta\sigma_t^2 + e_t \quad (10)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sum_{j=1}^4 w_j e_{t-j}^2 \quad (11)$$

where $w_j = (5-j/10)$, and they applied ML, finding that σ_t^2 was an important variable in explaining y_t , although the yield differential retained its significance.

In Pagan and Sabau (1987) the diagnostic test reported in the preceding section during the discussion of French et al was applied to the ELR formulation, resulting in a rejection of it at a very low level of significance. Thus the specification is inadequate. As in the equity premium models one possible explanation is that the ARCH model is an inappropriate one, and it is this theme which is pursued below.

Because it seems clear from past studies that the yield differential $x_t = (r_L - r_S)$ has an important role in accounting for the excess holding yield, we make it the primary factor upon which the conditional variance σ_t^2 depends.

⁴In their preferred model they have $\log \sigma_t^2$ as the regressor. Use of log here has the unfortunate effect of inducing a negative sign to the risk premium whenever $\sigma_t^2 < 1$. Moreover as $\sigma_t^2 \rightarrow 0$ the effect on y_t would be infinite. As there was little to choose between their models with σ_t^2 and $\log \sigma_t^2$ we work with what seems to be a more acceptable model, except when it comes to model comparisons.

Because there are only one hundred observations, keeping the number of conditioning elements low is vital if non-parametric procedures are to work well. However, provided the conditional mean (m_t) is accurately estimated, the instrumental variable estimator in Pagan and Ullah (1988), wherein σ_t^2 is replaced by $\hat{\phi}_t^2 = (y_t - \hat{m}_t)^2$ and is then instrumented with $\hat{\sigma}_t^2$, will be consistent even if the conditional variance has been mis-specified. Another advantage of having σ_t^2 depend on a single variable is that the mapping between σ_t^2 and the information set is easy to show graphically, and this is important for understanding our results.

The conditional mean was made to depend upon y_{t-1} and x_t and a Gaussian kernel was applied to generate \hat{m}_t and $\hat{\sigma}_t^2$. Figures 4 and 5 plot these against the yield differential (x_t).⁵ There is a very clear non-linear pattern in all the diagrams with high values of the differential tending to be associated with large values of σ_t^2 . Eventually a peak effect is reached, although there are few observations on yield differentials of this magnitude and the peak may be spurious. More important than the peak, however, is that whilst the yield differential would seem to have a quadratic element to it, the long tail with an associated sharp upward movement makes it hard to represent the pattern with a pure quadratic.

⁵Since Fourier series approximations are extensively employed in this section it is necessary that data be scaled over $(0, 2\pi)$. Accordingly, the following transformations were employed to ensure this - $y_t = 2(.8 + y'_t)$, $x_t = 1000(.00016 + x'_t)$, where the primes indicate the original data. Note also that we plot $\hat{\sigma}_t^2$ and \hat{m}_t against the *observed* values of x_t as this makes the mapping clearer for those values of x_t when there are few observations determining the conditional moments.

To emphasize this last point $\hat{\sigma}_t^2$ was approximated by fitting a Fourier series approximation as in Gallant (1982) i.e.

$$\hat{\sigma}_t^2 = \gamma_0 + \gamma_1 x_t + \gamma_2 x_t^2 + \sum_{j=1}^3 [\psi_j \cos(jx_t) + \theta_j \sin(jx_t)]. \quad (12)$$

The resulting regression had an R^2 of .93. Dropping the trigonometric terms so as to just retain a quadratic yields an R^2 of only .32. Graphically this effect is seen in figures 6 and 7; the first has the predictions from the quadratic. It is very clear from this that the quadratic is just not capable of picking up the sharp turning point found for $\hat{\sigma}_t^2$ and this suggests that procedures which highlight quadratic approximations, such as ARCH, will be an inadequate representation of σ_t^2 .⁶

The next step was to estimate (10) with $\hat{\sigma}_t^2$ replacing σ_t^2 (OLS) or with $\hat{\phi}_t^2$ replacing σ_t^2 and being instrumented with $\hat{\sigma}_t^2$ (IV). Examination of results from this regression showed that y_{t-1}^2 seemed to be needed as an extra regressor in (10) so that Table 3 gives parameter estimates for

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1}^2 + \delta \sigma_t^2 \quad , \quad (13)$$

where the observations are quarterly from 1960/3 to 1984/2, and data was obtained directly from the authors.

⁶In fact ARCH has a quadratic function of x_{t-1}^2 and not x_t^2 , and this dating is an unsatisfactory characteristic of ARCH models. ELR (p403) observed that "a useful extension would be to allow the yield differential to directly influence the variance", and the superiority of the non-parametric formulation over ARCH documented later is partly explained by this different dating.

Table 3

Parameter Estimates For Equation (13)*Kernel Based Conditional Moments

	<u>OLS</u>	<u>IV</u>
β_0	.960 (4.22)	1.062 (5.19)
β_1	.433 (3.95)	.393 (3.96)
β_2	-.045 (2.84)	-.054 (4.31)
δ	.215 (1.86)	.249 (2.33)

*Absolute value of t-statistics in parentheses

The non-parametric methodology seems to have generated reasonable results. However, since the Fourier approximation was previously very successful in capturing $\hat{\sigma}_t^2$, it was decided to see if that approach might yield equivalent results. Consequently the regressors in (12), augmented with y_{t-1} and y_{t-1}^2 , were fitted to y_t to obtain \tilde{m}_t , and then the residuals $(y_t - \tilde{m}_t)^2$ were regressed against the same set of regressors to obtain $\tilde{\sigma}_t^2$. Table 4 then gives estimates of the parameters of (13) but with $\tilde{\sigma}_t^2$ and $(y_t - \tilde{m}_t)^2$ as the non-parametric estimators.

Table 4

Parameter Estimates for Equation (13)*
Fourier Series Based Conditional Moments

	<u>OLS</u>	<u>IV</u>
β_0	1.081 (5.14)	1.081 (5.44)
β_1	.322 (3.12)	.322 (3.31)
β_2	-.033 (2.12)	-.033 (2.25)
δ	.636 (3.84)	.636 (4.07)

*Absolute values of t-statistics in parentheses.

The conclusions from Table 4 are much the same as in Table 3, although there is a much greater value of δ estimated from the Fourier-based approach. Diagnostic tests applied after the IV estimation give test statistic values of .006 (serial correlation), .65 (RESET) and 6.61 (heteroskedasticity), all of which should be referred to a $\chi^2(1)$ (the tests are described in Pesaran and Pesaran (1987)). Thus there is some evidence of heteroskedasticity in the residuals, and after a White adjustment the corresponding t-statistics in Table 4 are 4.29, 3.31, 1.8, and 2.86.

Because the model is the same in both situations discussed above, differing solely in choice of non-parametric estimator of conditional moments, non-nested tests might be usefully applied to determine the quality of each approximation. Since the "generated regressors correction" is small in both cases, we compared the OLS estimates and these are given in Table 5.

Table 5

Non-Nested Comparisons of Kernel (K) and
Fourier Series (F) Estimators*

	<u>F vs. K</u>	<u>K vs. F</u>
Cox	.31	-7.46
J-Test	-.32	3.30
JA-Test	-.32	3.30
F	.1	10.88

*The Cox test is Cox (1961) as adapted by Pesaran (1974); the J test is Davidson and Mackinnon's (1981); the JA is Fisher and Mc Aleer (1983); the F test adds the regressors from one model to the other. All tests except the last can be referred to the "t tables", although this is exact only for the JA test. The F should be referred to $F_{1,93}$.

It is clear from Table 5 that the Fourier based approximation yields a superior explanation of the behavior of the excess holding yield. Figure 8 plots $\hat{\sigma}_t^2$ against x_t . It is noticeable that there are a few small negative values for $\hat{\sigma}_t^2$ since no non-negativity constraint was imposed in estimation.

The largest of these occurs for the largest value of x_t in the data and is almost certainly sampling error. Obviously, it is worth exploring how one might impose non-negativity here, since one of the advantages of the kernel-based procedure is that this is easily done.⁷

A comparison of this non-parametric approach with ELR's ARCH method can also be effected through non-nested model comparisons. To this end equation (13), with $\tilde{\sigma}_t^2$ replacing σ_t^2 , is compared to an equation with the same regressors but with σ_t^2 replaced by $\log (\sigma_t^*)^2$, where $(\sigma_t^*)^2$ is ELR's ARCH estimate of the conditional variance. This comparison is a more favorable one than if it had been made directly with ELR as y_{t-1}^2 is an important contributor to the explanatory power of this modified ELR model. Table 6 gives the same non-nested tests as in Table 5.

Table 6
Non-Nested Comparisons of Fourier Series (F)
and ARCH Estimators (ELR)

	<u>F vs. ELR</u>	<u>ELR vs. F</u>
Cox	-.30	-11.77
J	.27	3.42
JA	.27	3.42
F	.07	11.67

⁷Although when conditioning upon different elements in the mean (Z_1) and variance (Z_2) earlier it is necessary to compute $E_{Z_2} [(y_t - E_{Z_1}(y_t))^2]$, and so two applications of the kernel are needed. When $Z_1 = Z_2$ the formula simplifies to $E_{Z_2}(y_t^2) - [E_{Z_2}(y_t)]^2$.

Table 6 shows the clear superiority of a non-parametric approach to the estimation of the excess holding yield on Treasury bills over the parametric method exemplified by ARCH. Our argument, based on earlier graphs, is that part of this reason is due to the greater responsiveness of the non-parametric estimator to the yield differential. Since the yield differential tended to change most after October 1979, it might be surmised that the improved fit here is due to an ability of the non-parametric estimators to capture the experience of that era better. A dramatic illustration of this is possible by plotting the predictions from equation (13) with $\tilde{\sigma}_t^2$ replacing σ_t^2 , the predictions from ELR (this does not include y_{t-1}^2 in the conditional mean and so corresponds to their figure 1), and the excess holding yield. This is done in Figure 9. The ARCH model just does not capture enough of the post October 1979 variation.⁸

6. Modeling Output Fluctuations

Many models involving risk require the derivation of a series for the volatility of output or consumption. Hodrick (1987) for example, working with a cash-in-advance model to determine the exchange rate, finds the solution to

⁸Indeed this explains why the diagnostic test in Pagan and Sabau (1987) is so significant. That test regresses the squared prediction errors less $(\sigma_t^*)^2$ against $(\sigma_t^*)^2$ and, as the squared errors rise more than $(\sigma_t^*)^2$ after October 1979, it is not surprising to find a large value of the test statistic.

incorporate a risk premium that is a function of the conditional variance of the growth in endowments (among other things). He estimated this quantity by choosing as ψ_t the rate of growth of industrial production in the U.S. and fitting the model

$$\psi_t = \mu + \gamma_1 \psi_{t-1} + \gamma_2 \psi_{t-2} + \phi_t \quad (14)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \phi_{t-1}^2 \quad (15)$$

where ϕ_t was assumed conditionally normal with variance σ_t^2 . This methodology, wherein an autoregression represents the conditional mean while an ARCH process represents the conditional variance, is a very common one in applied work. Consequently, it is worth investigating the application of non-parametric techniques to the estimation of σ_t^2 , since the assumption of linearity in the mean and a quadratic in the variance may be too restrictive.

As argued earlier any non-linearity in the conditional mean would affect the estimates of σ_t^2 , because residuals would be larger than the true innovations. Therefore the estimated α_1 may well be larger than it should be, as the ARCH model essentially estimates that quantity by a regression of the residuals squared against the lagged residuals squared. In the first two columns of Table 7 OLS estimates of γ_1 and γ_2 using monthly data over the period 1973(5) - 1986(10) are given. In the final two columns a simple test of non-linearity in the conditional mean is performed by augmenting (14) with the variable $\psi_{t-1} \hat{\phi}_{t-1}$, where $\hat{\phi}_t$ are the residuals from OLS on (14). This test is for a bilinear effect $\psi_{t-1} \phi_{t-1}$, and constitutes the LM test for that

effect. What is clear from this table is that there is strong evidence of a non-linearity in the conditional mean, and it is interesting to observe that the second lag in ψ_t , which is not significant in the linear regression, seems much more important once a non-linearity is allowed for.

Table 7
OLS Estimates of (14)

	<u>Basic Model</u>		<u>Augmented Model</u>	
	Est.	t-value	Est.	t-value
μ	1.63	5.28	.86	2.28
γ_1	.51	6.36	1.16	5.68
γ_2	.09	1.12	-.35	-2.35
coef. $\psi_{t-1}\phi_{t-1}$			-.18	-3.44

Because of this possible non-linearity both kernel and Fourier estimates of the non-parametric mean were found, being implemented exactly as described in the preceding sections, but with ψ_{t-1} and ψ_{t-2} as conditioning variables. Figure 10 graphs the estimated conditional means against ψ_{t-1} , the variable which seems to have the major influence on the mean. Of course the graph is not very smooth owing to the presence of ψ_{t-2} in the mean, but the pattern is quite distinct. There are only a few observations for low growth rates, and therefore the estimates at the left hand end are unlikely to be very reliable. Nevertheless, both methods give very similar estimates of the conditional mean and non-linearity does seem to be present. Further evidence for such non-linearity is the fact that the standard deviation of the (kernel-based)

non-parametric residuals is .2054 versus .3219 for the OLS residuals from (14) (the non-parametric residuals being defined as the difference between ψ_t and the kernel-based estimate of the conditional mean).

Non-parametric estimates of σ_t^2 were then obtained and these are plotted in Figure 11, again as a function of ψ_{t-1} . There is a fairly close correspondence between the two estimates of σ_t^2 , although the negative estimate from the Fourier approximation, as well as the fact that it is not quite as smooth a function, makes us prefer the kernel estimator. The correlation between the two estimates is .76. An "ARCH estimate" of σ_t^2 can be found by regressing the squared OLS residuals from (14) against a constant and the lagged squared residuals to get α_0 and α_1 , and then σ_t^2 is estimated as the predictions from this regression. Although the values of α_0 , α_1 , γ_1 and γ_2 determined by this strategy are not the MLE's, they are actually very close to what Hodrick reports. ($\hat{\alpha}_1$ being .31 with a t-value of 4.1.) The correlation of this ARCH estimate of σ_t^2 with the non-parametric estimate of σ_t^2 is -.05 (kernel) and .20 (Fourier), showing that major differences in estimates of σ_t^2 would be obtained from the two approaches.

To emphasise this point we performed the regression described above but with kernel-based non-parametric residuals, getting an $\hat{\alpha}_1$ of .008 and a t-value of .01, making this a dramatic illustration of the fact that, although σ_t^2 certainly varies with ψ_{t-1} and ψ_{t-2} , it does not seem to do so in the particular way described by an ARCH process. Consequently, it is likely that the presence of ARCH in Hodrick's work arises largely because of his mis-specification of the conditional mean, and his estimates of the volatility of output are exaggerated. There is in fact much less volatility in the

non-parametric estimate of σ_t^2 , with its standard deviation being .20 versus .32 for the ARCH estimate. Figure 12 shows the greater smoothness of the non-parametric estimate by plotting both against time.

7. Conclusion

This paper has argued that our lack of knowledge about the nature of the risk premium demands that considerable flexibility be employed when modeling series in which risk is likely to be an influential element. Early approaches concentrated upon describing the risk premium as a parametric function of information available to agents, and quite restrictive parameterizations have been employed. This paper has substituted a non-parametric for a parametric analysis. From the examples of the paper there seemed to be considerable gains from doing so. It has to be admitted though that the examples used were rather special, in that the number of conditioning variables taken to be important for the risk premium was small. For the financial variables of this paper such an assumption does not seem too unrealistic, and it is one generally employed in the parametric approach as well. Nevertheless, it is not clear how useful the methods discussed here will be once one leaves the realm of financial variables and considers other series in which risk is important e.g. the effects of commodity price movements upon demand and supply for these commodities. Moreover, financial modeling is generally blessed with the luxury of large data sets, which non-parametric methods may well need for success.

More research is therefore needed before one might confidently recommend non-parametric approaches as a standard tool of investigation where risk premia are involved. It may be that the function of non-parametric analysis is to suggest how current parametric approaches could be modified to obtain better results. In this regard the examples of sections 4-6 all point to the need to allow for a significant "tail" component to any mapping between the risk premium and the information set. Issues such as these and the performance of non-parametric methods in relatively small data sizes and with larger numbers of conditioning variables should obviously constitute important ingredients in any future research agenda.

References

- Backus, D.K., A. W. Gregory and S. E. Zin (1986), "Risk Premiums in the Term Structure: Evidence from Artificial Economies", *Discussion Paper No. 665*, Queen's University.
- Black, F. (1976), "Studies of Stock Price Volatility Changes," *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association*, 177-181.
- Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-328.
- Christie, A.A. (1982), "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, 10, 407-432.
- Cox, D.R. (1961), "Tests of Separate Families of Hypotheses," in *Proceedings of the Fourth Berkeley Symposium*, 1, 105-123 (University of California Press).
- Davidson, R. and J. G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49, 781-793.
- Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.
- Engle, R.F., D. M. Lillien and R. P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", *Econometrica*, 55, 391-407.
- Fama, E.F. and G. W. Schwert (1977), "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115-146.
- Fisher, G.R. and M. McAleer (1981), "Alternative Procedures and Associated Tests of Significance for Non-Nested Hypotheses", *Journal of Econometrics*, 16, 103-119.
- French, K.R., G. W. Schwert and R. F. Stambaugh (1987), "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3-29.
- Gallant, R. (1982), "Unbiased Determination of Production Technologies," *Journal of Econometrics*, 20, 285-323.
- Genotte, G. and T.A. Marsh (1985), "Variations in Ex-Ante Risk Premiums on Capital Assets," (mimeo, M.I.T.).
- 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. Frankel (ed.) *Exchange Rates and International Macroeconomics* (University of Chicago Press for the N.B.E.R., Chicago).

- Hodrick, R. (1987), "Risk, Uncertainty and Exchange Rates", (forthcoming, *Journal of Monetary Economics*)
- Mankiw, N.G. and L.H. Summers (1984), "Do Long-Term Interest Rates Overreact to Short-Term Interest Rates?", *Brookings Papers on Economic Activity*, 223-242.
- Mehra, R. and E. Prescott (1985), "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145-161.
- Merton, R.C. (1980), "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323-361.
- Nelson, D.B. (1987), "Conditional Heteroskedasticity in Asset Returns: A New Approach" (mimeo, M.I.T.).
- Newey, W.D. (1985), "Maximum Likelihood Specification Testing and Conditional Moments," *Econometrica*, 53, 1047-1070.
- Pagan, A.R. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221-247.
- Pagan, A.R. and H. Sabau (1987), "Consistency Tests for Heteroskedastic and Risk Models," (mimeo, University of Rochester).
- Pagan, A.R. and A. Ullah (1988), "The Econometric Analysis of Models with Risk Terms," *Journal of Applied Econometrics*,
- Pesaran, M.H. (1974), "On the General Problem of Model Selection," *Review of Economic Studies*, 41, 153-171.
- Pesaran, M.H. and B. Pesaran (1987), *Data-Fit: An Interactive Econometric Software Package* (Oxford University Press, Oxford).
- Pindyck, R.S. (1984). "Risk, Inflation and the Stock Market", *American Economic Review*, 74, 335-351.
- Reitz, T.A. (1987), "The Equity Risk Premium: A solution," *Journal of Monetary Economics* (forthcoming).
- Summers, L.H. and J.M. Poterba (1986), "The Persistence of Volatility and Stock Market Fluctuations," *American Economic Review*, 76, 1142-1151.
- Tauchen, G. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415-443.

Fig.1 Risk Premium as a Function of
the State for Backus et al Model

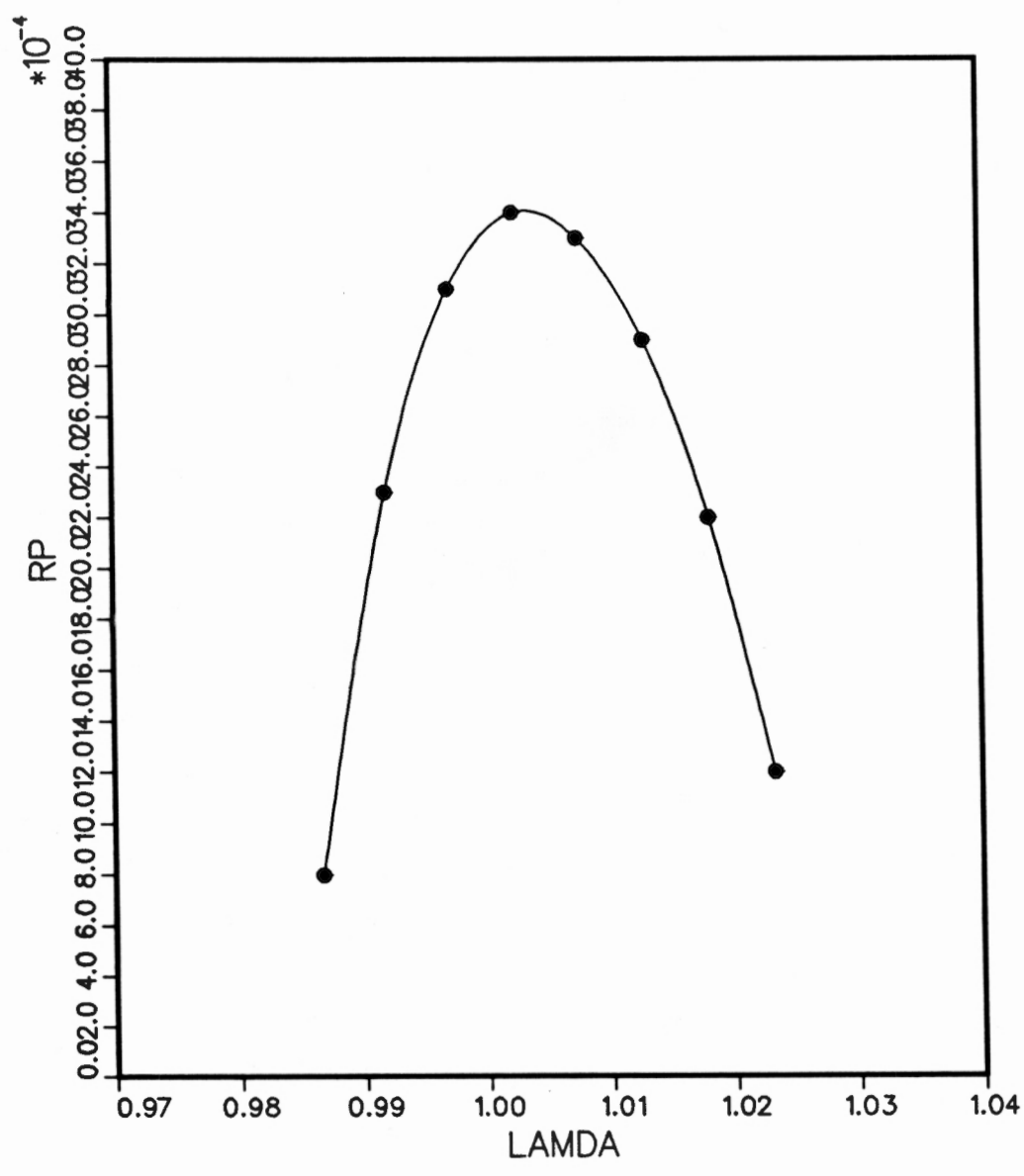


Fig. 2 Non-parametric Mean Equity Yield Premium : Kernel Based

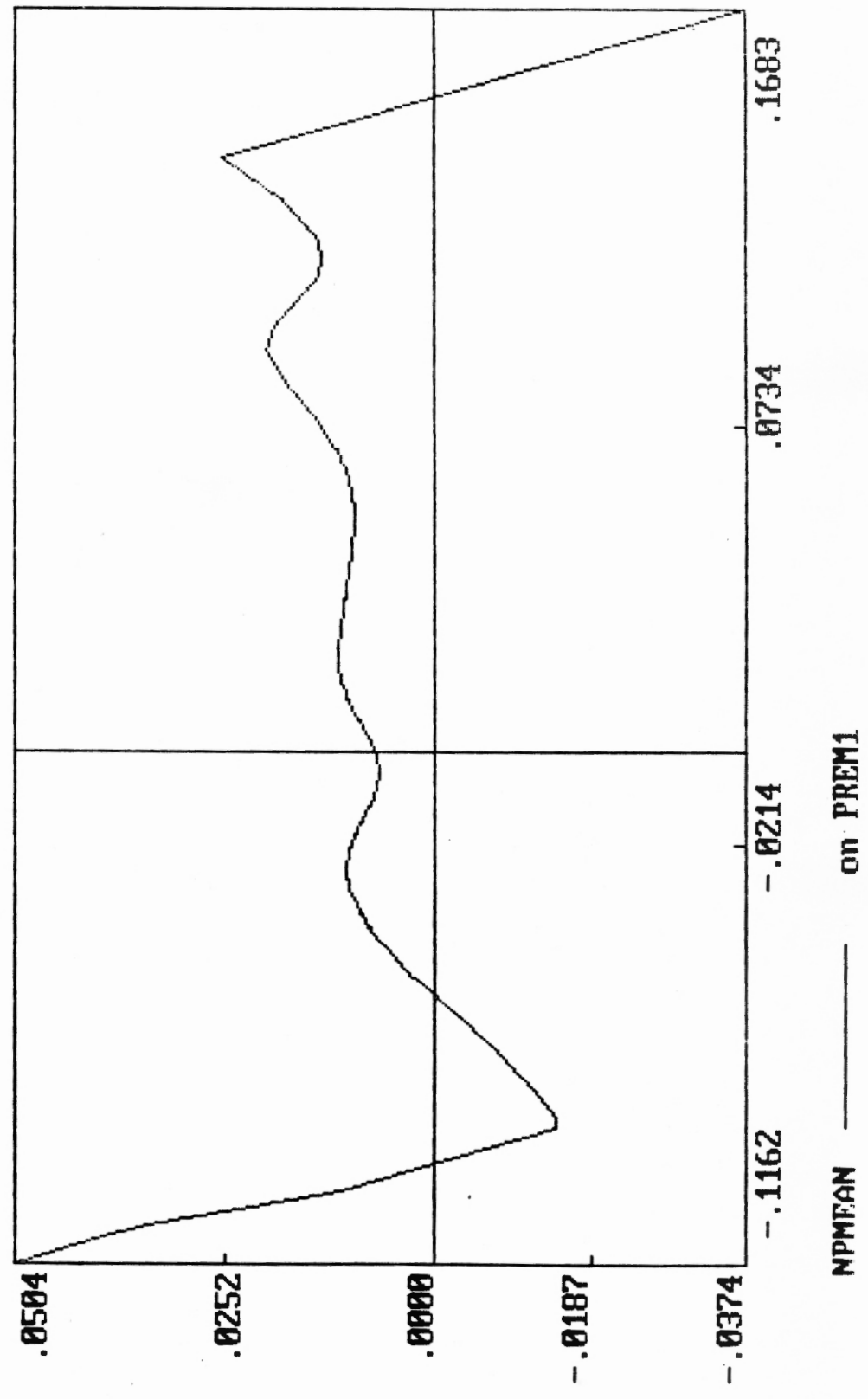


Fig. 3 Nonparametric Variance of Equity Yield Premium : Kernel Based

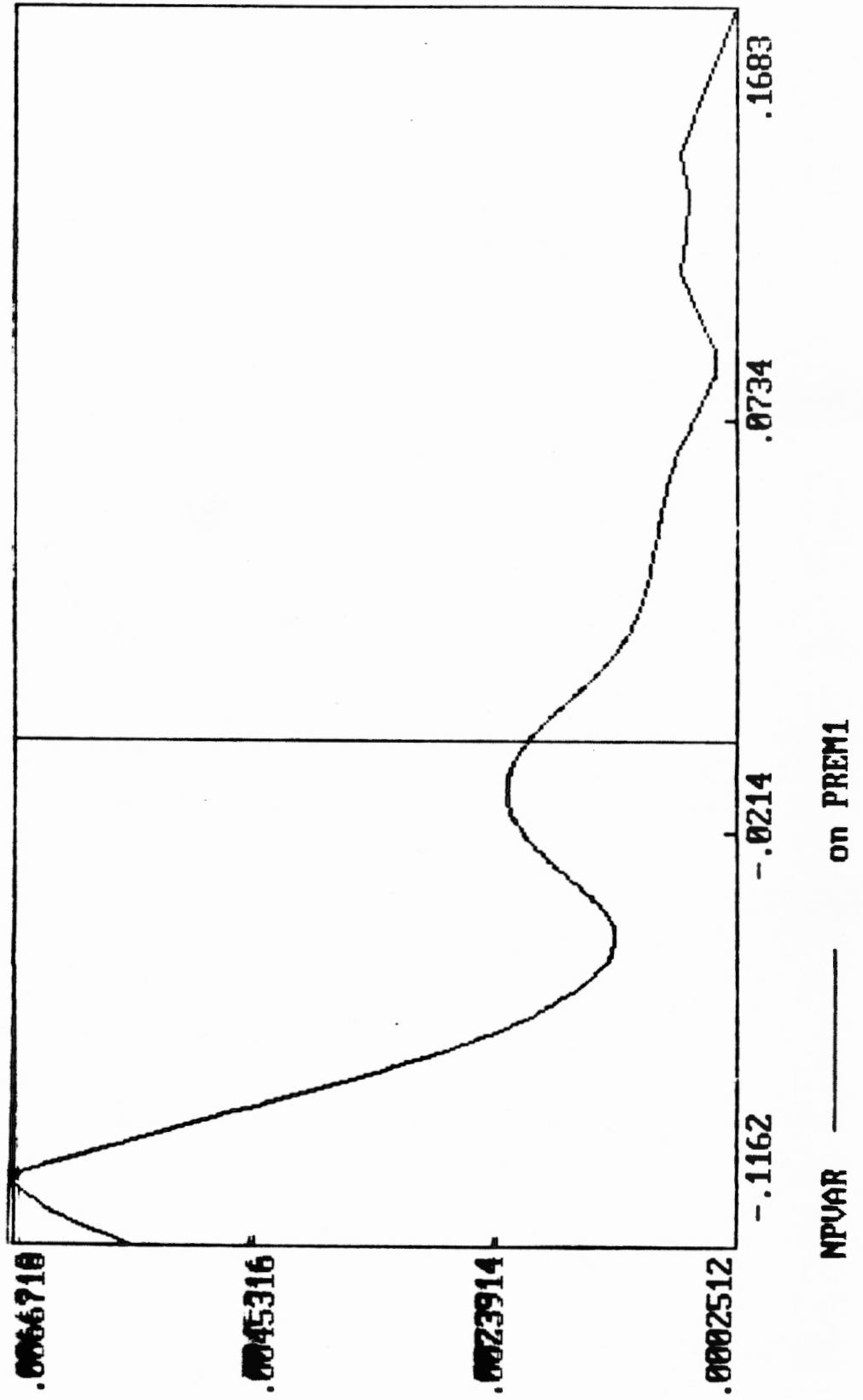


Fig. 4 Non-parametric Mean, T.Bills Holding Yield, Kernel Based

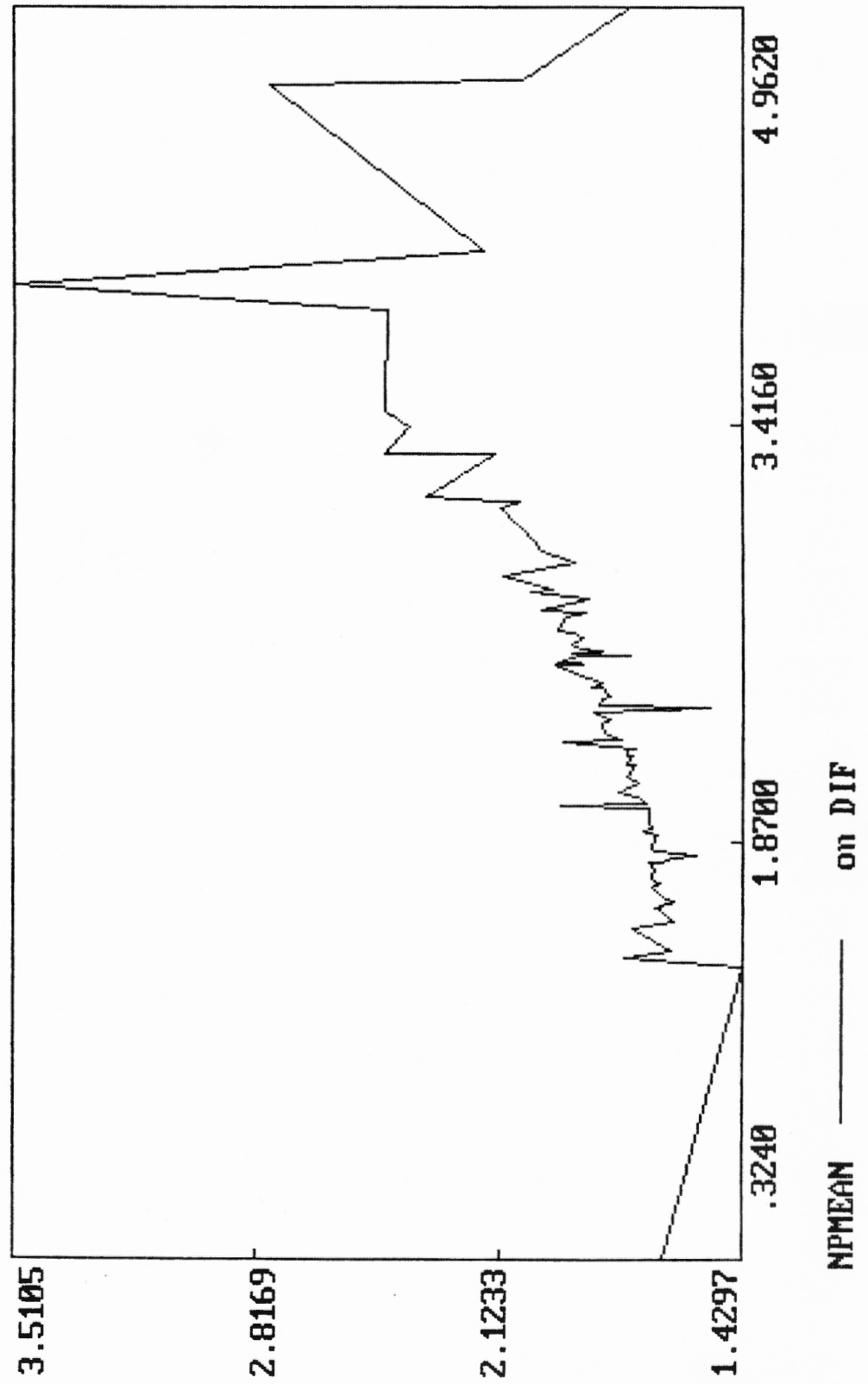


Fig.5 **Kernal Based Non-parametric Variance : Excess Holding Yield,T.Bills**

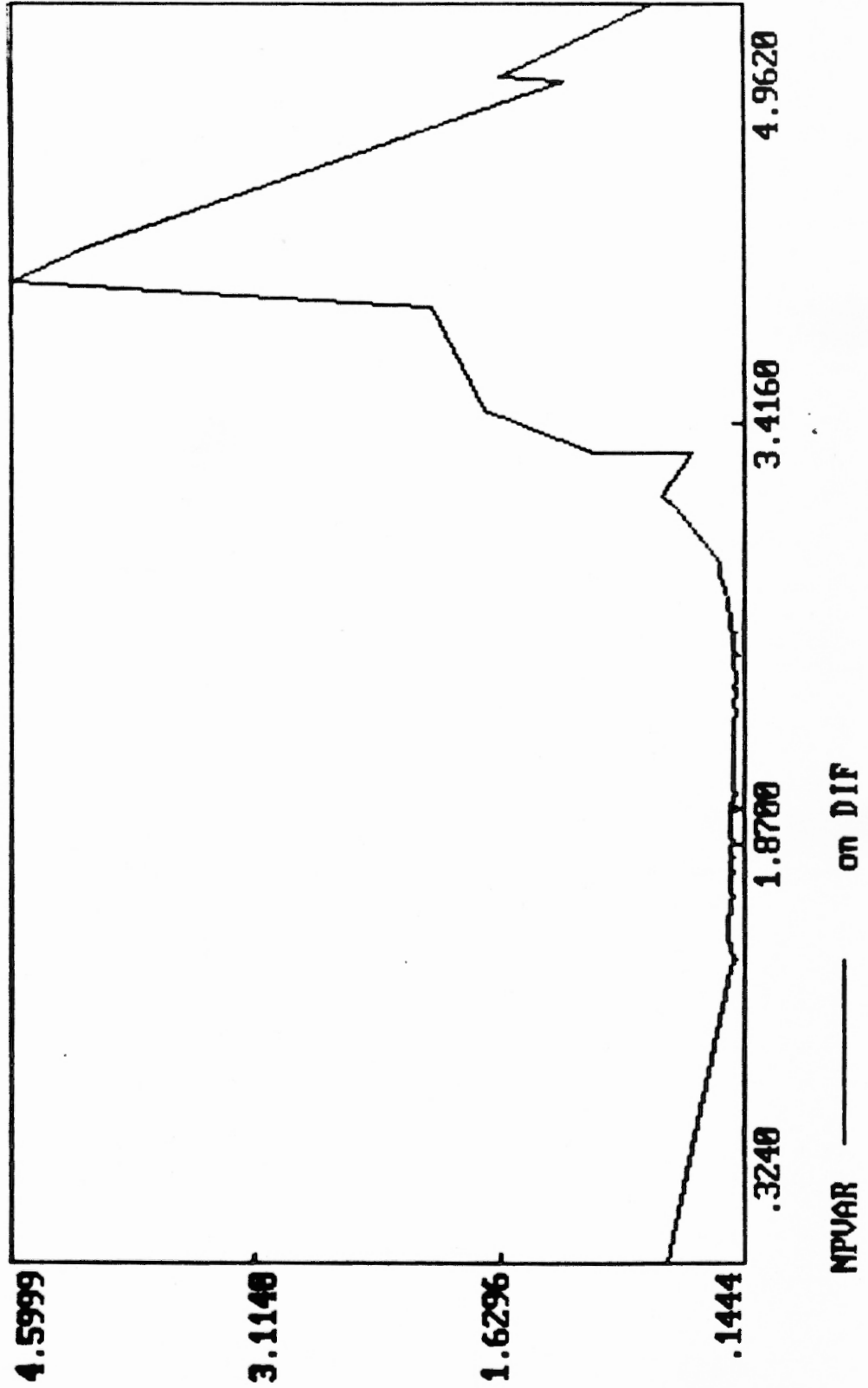


Fig.6: Quadratic Approximation to Non-parametric Variance

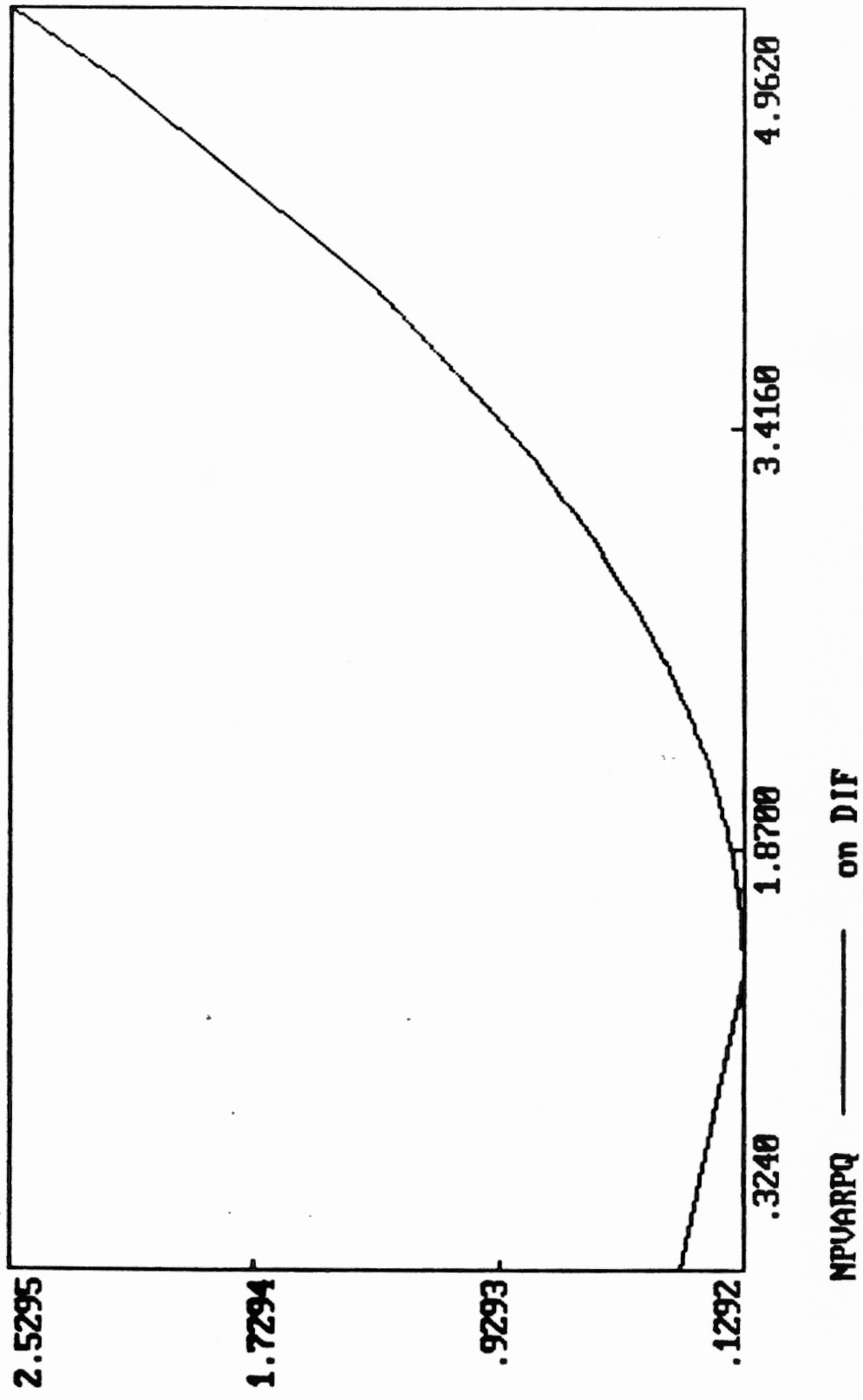


Fig. 7: **Fourier Approximation to Non-parametric Variance**

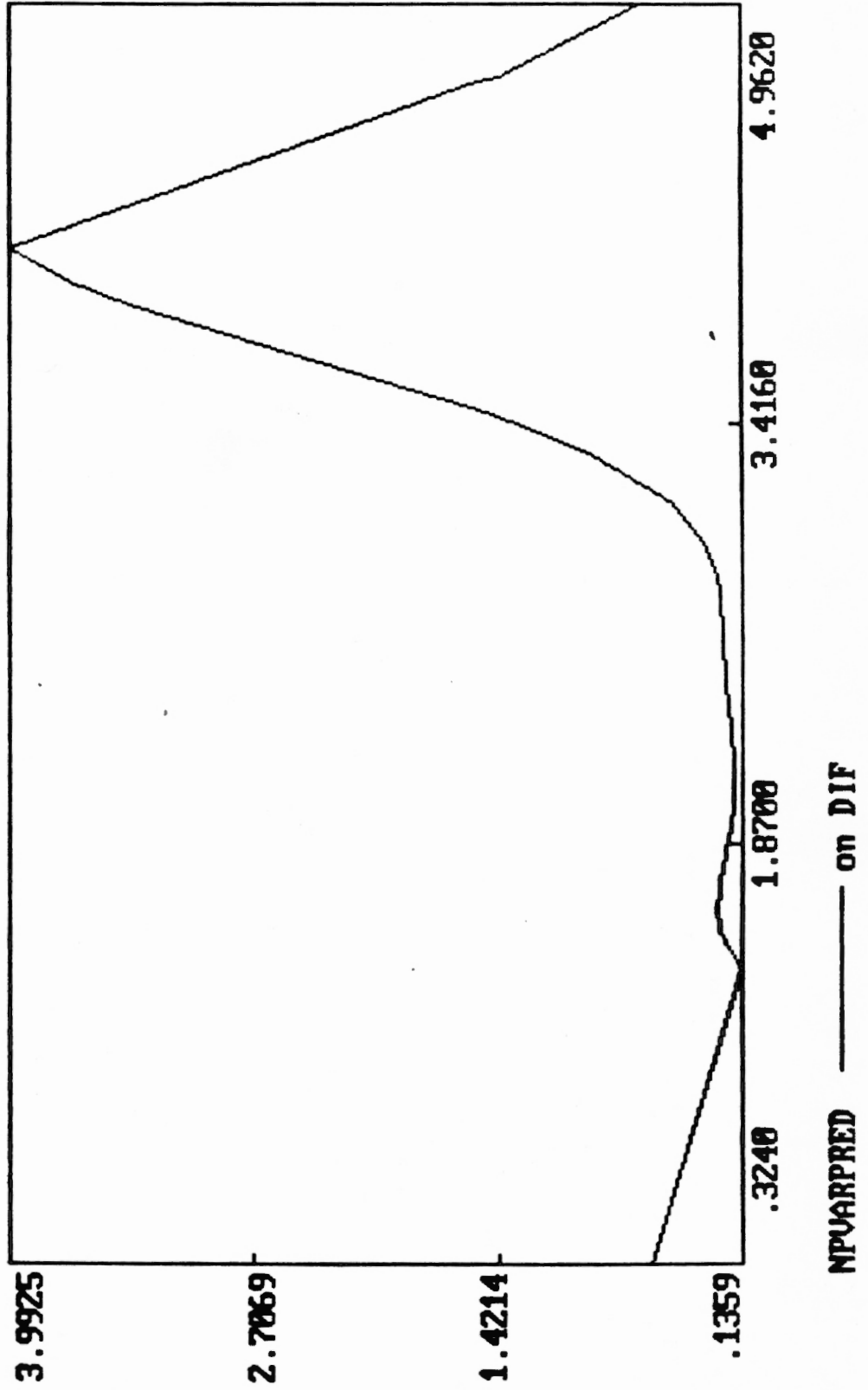
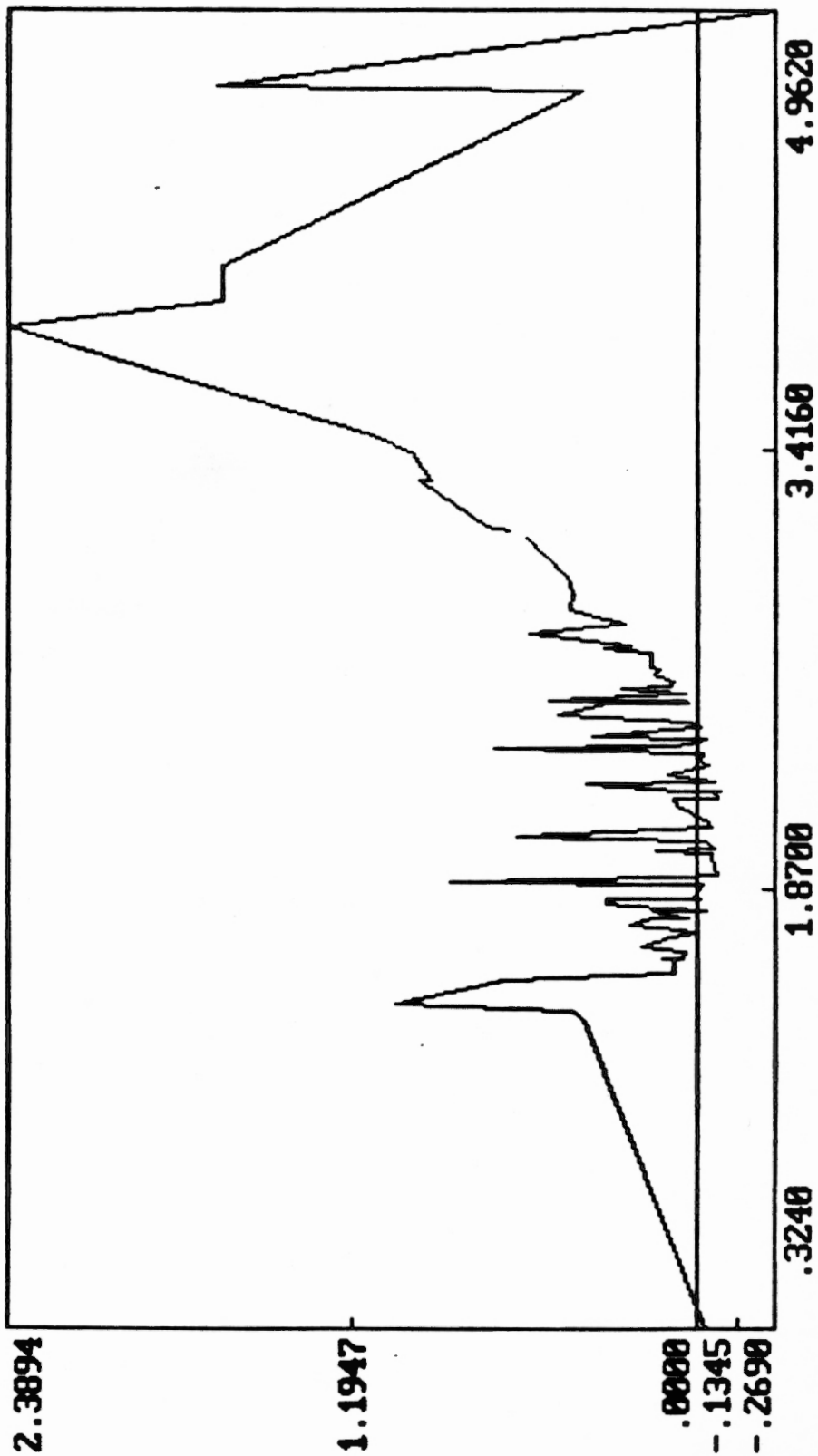


Fig 8: **Fourier Based Non-parametric Variance**



NPVARG ——— on DIF

FIG.9 Excess Holding Yield, T Bills : Predictions from Fourier Based Non-parametric Variance and Engle/Lillien/Robins Equation

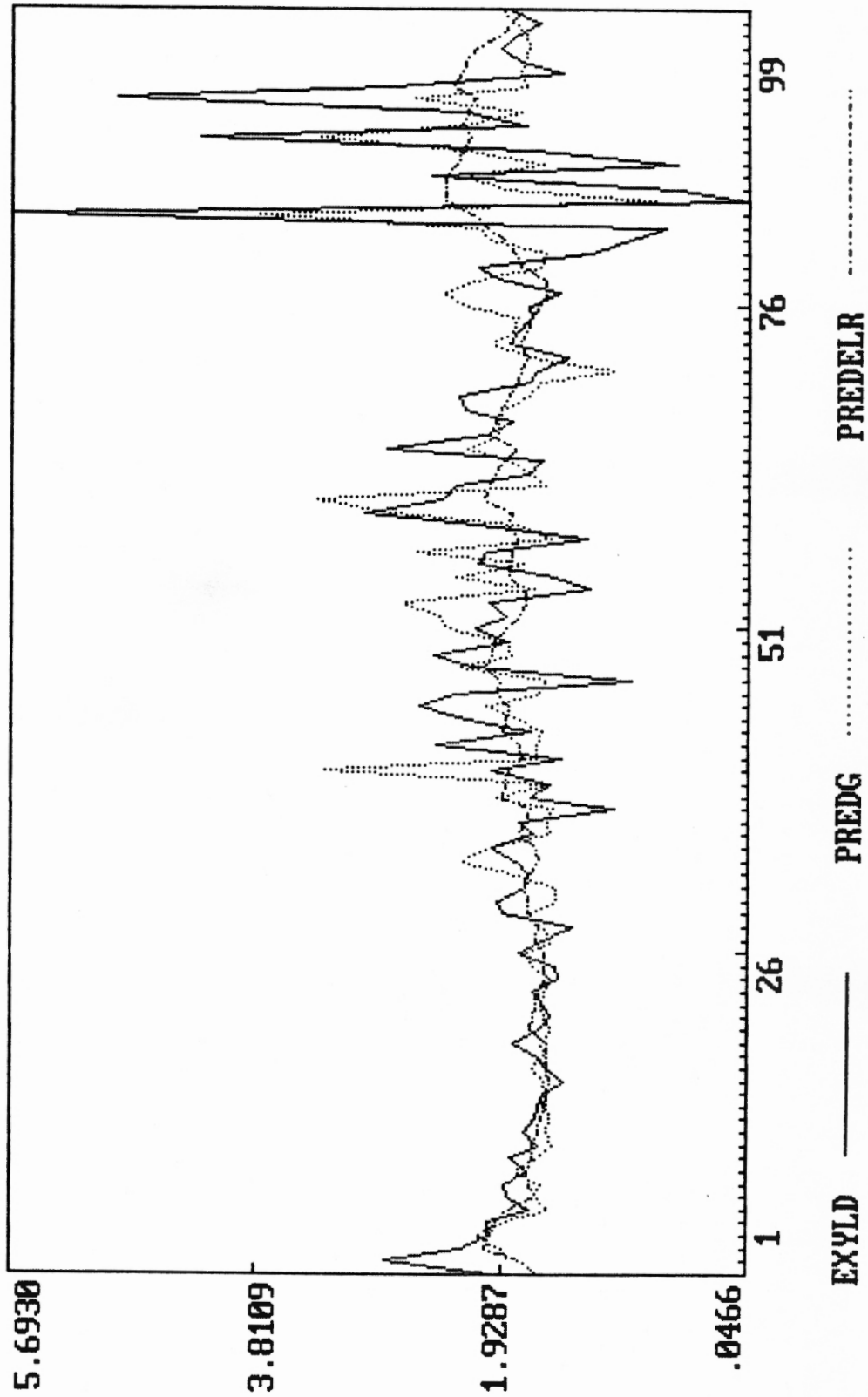


Fig.10 Kernel and Fourier Non-parametric Means, Production Growth

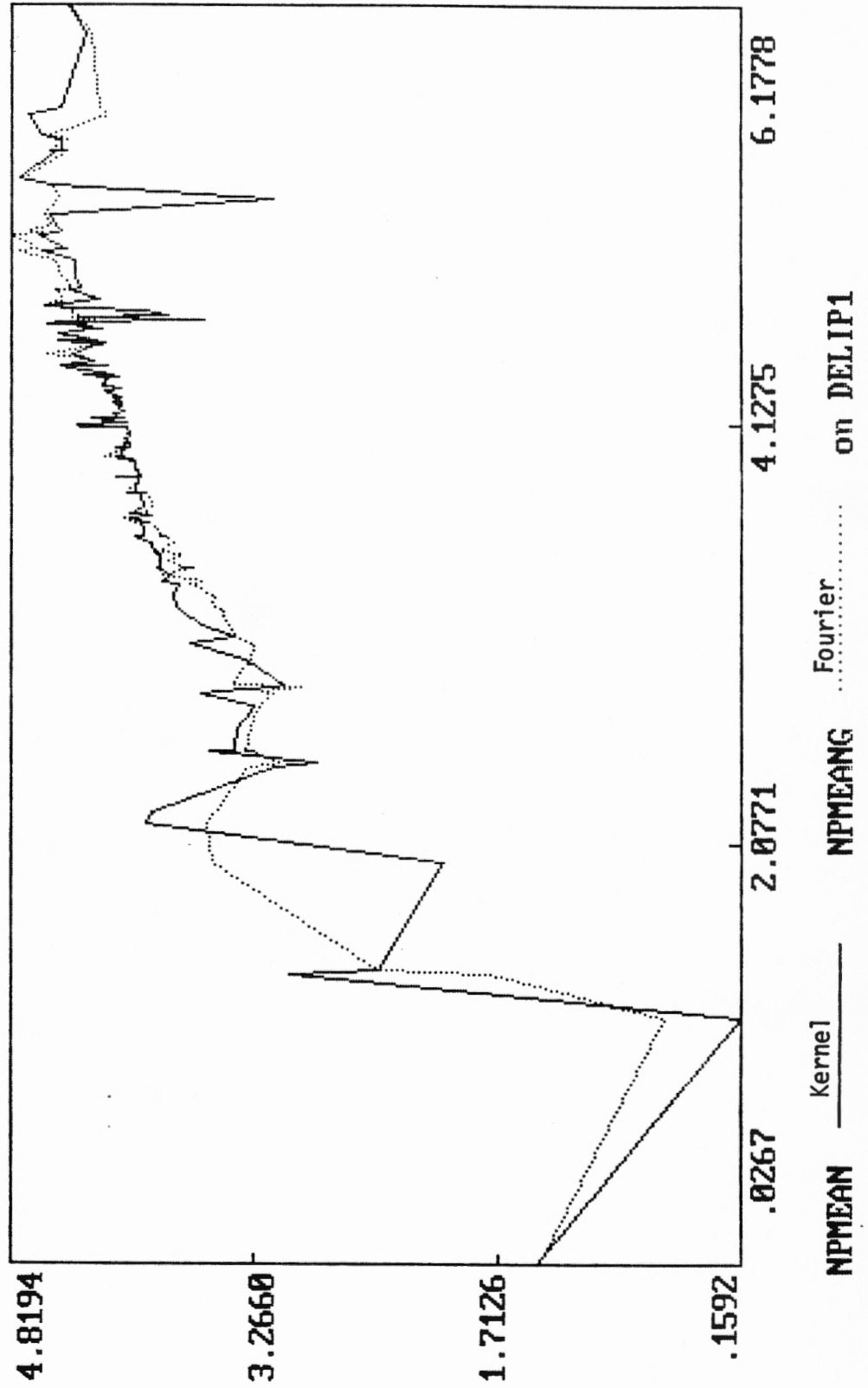


Fig 11 Kernel and Fourier Non-Parametric Variances: Production Growth

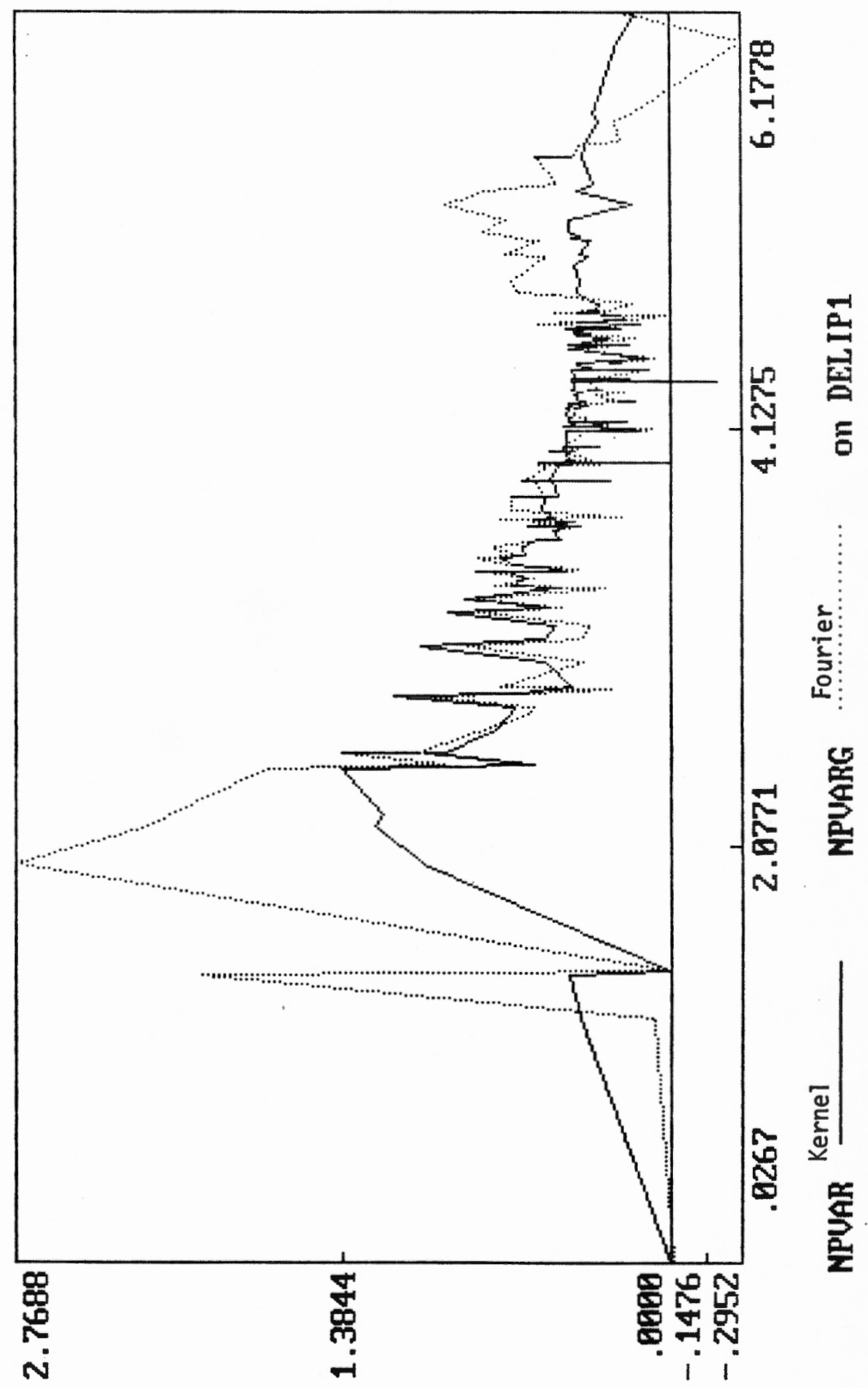
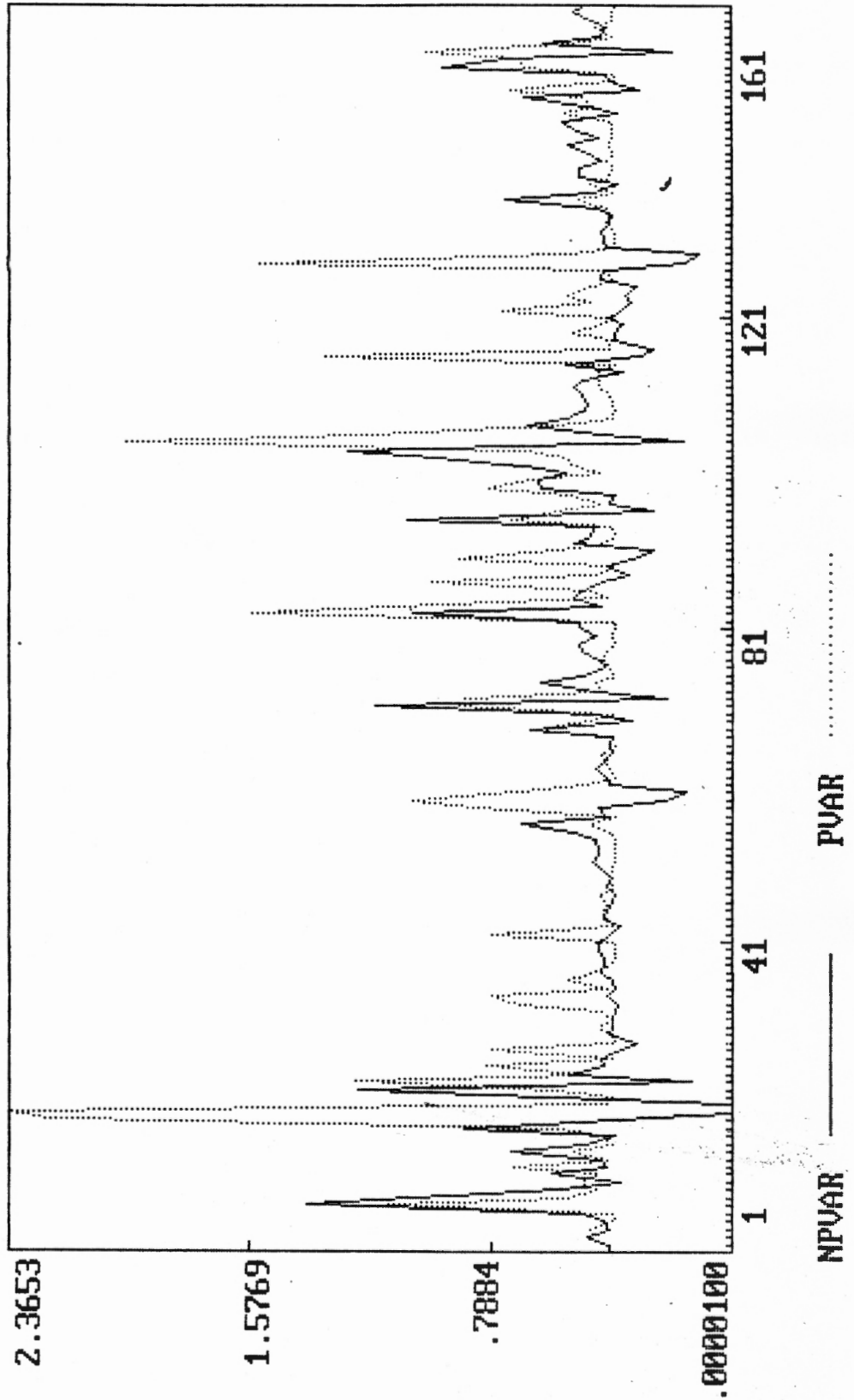


Fig. 12 Non-parametric and ARCH Estimated Conditional Variances for
U.S. Industrial Production Rate of Growth



Rochester Center for Economic Research
University of Rochester
Department of Economics
Rochester, NY 14627

1987-88 DISCUSSION PAPERS

- WP#68 RECURSIVE UTILITY AND OPTIMAL CAPITAL ACCUMULATION, I: EXISTENCE,
by Robert A. Becker, John H. Boyd III, and Bom Yong Sung, January
1987
- WP#69 MONEY AND MARKET INCOMPLETENESS IN OVERLAPPING-GENERATIONS MODELS,
by Marianne Baxter, January 1987
- WP#70 GROWTH BASED ON INCREASING RETURNS DUE TO SPECIALIZATION
by Paul M. Romer, January 1987
- WP#71 WHY A STUBBORN CONSERVATIVE WOULD RUN A DEFICIT: POLICY WITH
TIME-INCONSISTENT PREFERENCES
by Torsten Persson and Lars E.O. Svensson, January 1987
- WP#72 ON THE CONTINUUM APPROACH OF SPATIAL AND SOME LOCAL PUBLIC GOODS OR
PRODUCT DIFFERENTIATION MODELS
by Marcus Berliant and Thijs ten Raa, January 1987
- WP#73 THE QUIT-LAYOFF DISTINCTION: GROWTH EFFECTS
by Kenneth J. McLaughlin, February 1987
- WP#74 SOCIAL SECURITY, LIQUIDITY, AND EARLY RETIREMENT
by James A. Kahn, March 1987
- WP#75 THE PRODUCT CYCLE HYPOTHESIS AND THE HECKSCHER-OHLIN-SAMUELSON THEORY
OF INTERNATIONAL TRADE
by Sugata Marjit, April 1987
- WP#76 NOTIONS OF EQUAL OPPORTUNITIES
by William Thomson, April 1987
- WP#77 BARGAINING PROBLEMS WITH UNCERTAIN DISAGREEMENT POINTS
by Youngsub Chun and William Thomson, April 1987
- WP#78 THE ECONOMICS OF RISING STARS
by Glenn M. MacDonald, April 1987
- WP#79 STOCHASTIC TRENDS AND ECONOMIC FLUCTUATIONS
by Robert King, Charles Plosser, James Stock, and Mark Watson,
April 1987
- WP#80 INTEREST RATE SMOOTHING AND PRICE LEVEL TREND-STATIONARITY
by Marvin Goodfriend, April 1987
- WP#81 THE EQUILIBRIUM APPROACH TO EXCHANGE RATES
by Alan C. Stockman, revised, April 1987

- WP#82 INTEREST-RATE SMOOTHING
by Robert J. Barro, May 1987
- WP#83 CYCLICAL PRICING OF DURABLE LUXURIES
by Mark Bilis, May 1987
- WP#84 EQUILIBRIUM IN COOPERATIVE GAMES OF POLICY FORMULATION
by Thomas F. Cooley and Bruce D. Smith, May 1987
- WP#85 RENT SHARING AND TURNOVER IN A MODEL WITH EFFICIENCY UNITS OF HUMAN CAPITAL
by Kenneth J. McLaughlin, revised, May 1987
- WP#86 THE CYCLICALITY OF LABOR TURNOVER: A JOINT WEALTH MAXIMIZING HYPOTHESIS
by Kenneth J. McLaughlin, revised, May 1987
- WP#87 CAN EVERYONE BENEFIT FROM GROWTH? THREE DIFFICULTIES
by Herve' Moulin and William Thomson, May 1987
- WP#88 TRADE IN RISKY ASSETS
by Lars E.O. Svensson, May 1987
- WP#89 RATIONAL EXPECTATIONS MODELS WITH CENSORED VARIABLES
by Marianne Baxter, June 1987
- WP#90 EMPIRICAL EXAMINATIONS OF THE INFORMATION SETS OF ECONOMIC AGENTS
by Nils Gottfries and Torsten Persson, June 1987
- WP#91 DO WAGES VARY IN CITIES? AN EMPIRICAL STUDY OF URBAN LABOR MARKETS
by Eric A. Hanushek, June 1987
- WP#92 ASPECTS OF TOURNAMENT MODELS: A SURVEY
by Kenneth J. McLaughlin, July 1987
- WP#93 ON MODELLING THE NATURAL RATE OF UNEMPLOYMENT WITH INDIVISIBLE LABOR
by Jeremy Greenwood and Gregory W. Huffman
- WP#94 TWENTY YEARS AFTER: ECONOMETRICS, 1966-1986
by Adrian Pagan, August 1987
- WP#95 ON WELFARE THEORY AND URBAN ECONOMICS
by Marcus Berliant, Yorgos Y. Papageorgiou and Ping Wang,
August 1987
- WP#96 ENDOGENOUS FINANCIAL STRUCTURE IN AN ECONOMY WITH PRIVATE INFORMATION
by James Kahn, August 1987
- WP#97 THE TRADE-OFF BETWEEN CHILD QUANTITY AND QUALITY: SOME EMPIRICAL EVIDENCE
by Eric Hanushek, September 1987

- WP#98 SUPPLY AND EQUILIBRIUM IN AN ECONOMY WITH LAND AND PRODUCTION
by Marcus Berliant and Hou-Wen Jeng, September 1987
- WP#99 AXIOMS CONCERNING UNCERTAIN DISAGREEMENT POINTS FOR 2-PERSON
BARGAINING PROBLEMS
by Youngsub Chun, September 1987
- WP#100 MONEY AND INFLATION IN THE AMERICAN COLONIES: FURTHER EVIDENCE ON
THE FAILURE OF THE QUANTITY THEORY
by Bruce Smith, October 1987
- WP#101 BANK PANICS, SUSPENSIONS, AND GEOGRAPHY: SOME NOTES ON THE
"CONTAGION OF FEAR" IN BANKING
by Bruce Smith, October 1987
- WP#102 LEGAL RESTRICTIONS, "SUNSPOTS", AND CYCLES
by Bruce Smith, October 1987
- WP#103 THE QUIT-LAYOFF DISTINCTION IN A JOINT WEALTH MAXIMIZING APPROACH TO
LABOR TURNOVER
by Kenneth McLaughlin, October 1987
- WP#104 ON THE INCONSISTENCY OF THE MLE IN CERTAIN HETEROSKEDASTIC REGRESSION
MODELS
by Adrian Pagan and H. Sabau, October 1987
- WP#105 RECURRENT ADVERTISING
by Ignatius J. Horstmann and Glenn M. MacDonald, October 1987
- WP#106 PREDICTIVE EFFICIENCY FOR SIMPLE NONLINEAR MODELS
by Thomas F. Cooley, William R. Parke and Siddhartha Chib,
October 1987
- WP#107 CREDIBILITY OF MACROECONOMIC POLICY: AN INTRODUCTION AND A BROAD
SURVEY
by Torsten Persson, November 1987
- WP#108 SOCIAL CONTRACTS AS ASSETS: A POSSIBLE SOLUTION TO THE
TIME-CONSISTENCY PROBLEM
by Laurence Kotlikoff, Torsten Persson and Lars E. O. Svensson,
November 1987
- WP#109 EXCHANGE RATE VARIABILITY AND ASSET TRADE
by Torsten Persson and Lars E. O. Svensson, Novmeber 1987
- WP#110 MICROFOUNDATIONS OF INDIVISIBLE LABOR
by Vittorio Grilli and Richard Rogerson, November 1987
- WP#111 FISCAL POLICIES AND THE DOLLAR/POUND EXCHANGE RATE: 1870-1984
by Vittorio Grilli, November 1987
- WP#112 INFLATION AND STOCK RETURNS WITH COMPLETE MARKETS
by Thomas Cooley and Jon Sonstelie, November 1987

- WP#113 THE ECONOMETRIC ANALYSIS OF MODELS WITH RISK TERMS
by Adrian Pagan and Aman Ullah, December 1987
- WP#114 PROGRAM TARGETING OPTIONS AND THE ELDERLY
by Eric Hanushek and Roberton Williams, December 1987
- WP#115 BARGAINING SOLUTIONS AND STABILITY OF GROUPS
by Youngsub Chun and William Thomson, December 1987
- WP#116 MONOTONIC ALLOCATION MECHANISMS
by William Thomson, December 1987
- WP#117 MONOTONIC ALLOCATION MECHANISMS IN ECONOMIES WITH PUBLIC GOODS
by William Thomson, December 1987
- WP#118 ADVERSE SELECTION, AGGREGATE UNCERTAINTY, AND THE ROLE FOR MUTUAL
INSURANCE COMPANIES
by Bruce D. Smith and Michael J. Stutzer, February 1988
- WP#119 INTEREST ON RESERVES AND SUNSPOT EQUILIBRIA: FRIEDMAN'S PROPOSAL
RECONSIDERED
by Bruce D. Smith, February 1988
- WP#120 INTERNATIONAL FINANCIAL INTERMEDIATION AND AGGREGATE FLUCTUATIONS
UNDER ALTERNATIVE EXCHANGE RATE REGIMES
by Jeremy Greenwood and Stephen D. Williamson, February 1988
- WP#121 FINANCIAL DEREGULATION, MONETARY POLICY, AND CENTRAL BANKING
by Marvin Goodfriend and Robert G. King, February 1988
- WP#122 BANK RUNS IN OPEN ECONOMIES AND THE INTERNATIONAL TRANSMISSION OF
PANICS
by Peter M. Garber and Vittorio U. Grilli, March 1988
- WP#123 CAPITAL ACCUMULATION IN THE THEORY OF LONG RUN GROWTH
by Paul M. Romer, March 1988
- WP#124 FINANCIAL INTERMEDIATION AND ENDOGENOUS GROWTH
by Valerie R. Bencivenga and Bruce D. Smith, March 1988
- WP#125 UNEMPLOYMENT, THE VARIABILITY OF HOURS, AND THE PERSISTENCE OF
"DISTURBANCES": A PRIVATE INFORMATION APPROACH
by Bruce D. Smith, March 1988
- WP#126 WHAT CAN BE DONE WITH BAD SCHOOL PERFORMANCE DATA?
by Eric Hanushek and Lori Taylor, March 1988
- WP#127 EQUILIBRIUM MARKETING STRATEGIES: IS THERE ADVERTISING, IN TRUTH?
by Ignatius Horstmann and Glenn MacDonald, revised, March 1988
- WP#128 REAL EXCHANGE RATE VARIABILITY UNDER PEGGED AND FLOATING NOMINAL
EXCHANGE RATE SYSTEMS: AN EQUILIBRIUM THEORY
by Alan C. Stockman, April 1988

- WP#129 POST-SAMPLE PREDICTION TESTS FOR GENERALIZED METHOD OF MOMENT ESTIMATORS
by Dennis Hoffman and Adrian Pagan, April 1988
- WP#130 GOVERNMENT SPENDING IN A SIMPLE MODEL OF ENDOGENOUS GROWTH
by Robert J. Barro, May 1988
- WP#131 FINANCIAL DEVELOPMENT, GROWTH, AND THE DISTRIBUTION OF INCOME
by Jeremy Greenwood and Boyan Jovanovic, May 1988
- WP#132 EMPLOYMENT AND HOURS OVER THE BUSINESS CYCLE
by Jang-Ok Cho and Thomas F. Cooley, May 1988
- WP#133 A REFINEMENT AND EXTENSION OF THE NO-ENVY CONCEPT
by Dimitrios Diamantaras and William Thomson, May 1988
- WP#134 NASH SOLUTION AND UNCERTAIN DISAGREEMENT POINTS
by Youngsub Chun and William Thomson, May 1988
- WP#135 NON-PARAMETRIC ESTIMATION AND THE RISK PREMIUM
by Adrian Pagan and Y. Hong, May 1988

To order copies of the above papers complete the attached invoice and return to Christine Massaro, W. Allen Wallis Institute of Political Economy, RCER, 109B Harkness Hall, University of Rochester, Rochester, NY 14627. Three (3) papers per year will be provided free of charge as requested below. Each additional paper will require a \$5.00 service fee which must be enclosed with your order. For your convenience an invoice is provided below in order that you may request payment from your institution as necessary. Please make your check payable to the **Rochester Center for Economic Research**. Checks must be drawn from a U.S. bank and in U.S. dollars.

W. Allen Wallis Institute for Political Economy

Rochester Center for Economic Research, Working Paper Series

OFFICIAL INVOICE

Requestor's Name _____

Requestor's Address _____

Please send me the following papers free of charge (**Limit: 3 free per year**).

WP# _____ WP# _____ WP# _____

I understand there is a \$5.00 fee for each additional paper. Enclosed is my check or money order in the amount of \$_____. Please send me the following papers.

WP# _____ WP# _____ WP# _____

WP# _____ WP# _____ WP# _____

WP# _____ WP# _____ WP# _____

WP# _____ WP# _____ WP# _____