Fordham University
Department of Economics
Discussion Paper Series

Heteroskedasticity and Autocorrelation Efficient (HAE) Estimation and Pivots for Jointly Evolving Series

Hrishikesh D. Vinod
Fordham University, Department of Economics

Discussion Paper No: 2008-15
September 2008

Department of Economics
Fordham University
441 E Fordham Rd, Dealy Hall
Bronx, NY 10458
(718) 817-4048

Heteroscedasticity and Autocorrelation Efficient (HAE) Estimation and Pivots for Jointly Evolving Series

H. D. Vinod *Professor of Economics,
Fordham University, Bronx, New York 10458

Abstract

A new two-way map between time domain and numerical magnitudes or values domain (v-dom) provides a new solution to heteroscedasticity. Since sorted logs of squared fitted residuals are monotonic in the v-dom, we obtain a parsimonious fit there. Two theorems prove consistency, asymptotic normality, efficiency and specification-robustness, supplemented by a simulation. Since Dufour's (1997) impossibility theorems show how confidence intervals from Wald-type tests can have zero coverage, I suggest Godambe pivot functions (GPF) with good finite sample coverage and distribution-free robustness. I use the Frisch-Waugh theorem and the scalar GPF to construct new confidence intervals for regression parameters and apply Vinod's (2004, 2006) maximum entropy bootstrap. I use Irving Fisher's model for interest rates and Keynesian consumption function for illustration.

1 Introduction

Since this paper is prepared for a conference in honor of my own teacher Professor Dhrymes, I attempt to build upon his work. Dhrymes (1998, Ch.6)

*Presented at an International Conference of Econometricians in honor of Prof. Phoebus Dhrymes at Paphos, Cyprus, on June 3, 2007. I thank B. D. McCullough and the conference participants including: Steven Durlauf, David Hendry, Soren Johansen, Peter Phillips and Aris Spanos for helpful comments. E-Mail: Vinod@fordham.edu

proposed a novel Wald-type conformity cointegration test (CCT) for testing the null hypothesis of no cointegration using $\Pi(L)$ matrices of coefficients of lagged levels in a typical vector error correction model (ECM). Since the CCT has not received wide following by econometricians, I have included its discussion in the hope that it might. Inspired by CCT but going beyond testing, this paper discusses efficient estimation of coefficients and confidence intervals. We shall see that this paper does not deal at all with the ECM as such, except in the context of CCT and is mainly about HAE estimation. The paper offers following novelty items:

(i) Since inverting Wald-type tests to obtain confidence intervals is problematic due to Dufour's (1997) impossibility theorems, replacing Fisher's pivot by a more robust Godambe's pivot function (GPF) is desirable as explained in Vinod (1998, 2000). This paper exploits the Frisch-Waugh theorem to construct new and much simpler confidence intervals for deep parameters.

(ii) Vinod (2004, 2006) provides a maximum entropy bootstrap for evolutionary series while avoiding the differencing used in CCT. This paper recognizes for the first time that the underlying sorting algorithm yields two-way maps between the time domain and numerical magnitudes or values domain, denoted by (t-dom $\leftrightarrow$ v-dom).

(iii) Since efficient estimation correcting for autocorrelation involves off-diagonal elements of the covariance matrix it is more practical to first correct for it before turning to heteroscedasticity. Using the (t-dom $\leftrightarrow$ v-dom) map this paper suggests a new HAE method for solving heteroscedasticity while avoiding incidental parameters problem. We prove theorems and include a simulation and examples.

Now we describe the motivation behind these novelty items with the help of an example. An economic equilibrium often relates one variable to a set of variables on the right hand side (RHS). For example, Irving Fisher proposed a version of the following equilibrium relation between tax adjusted nominal interest r_t and unobservable expected inflation Ep_t at time t as:

$$r_t = RHS_t + \epsilon_t, \text{ where } RHS_t = \beta_0 + \beta_1 Ep_t. \tag{1}$$

If due to a rise in expected overall inflation the tax adjusted nominal interest falls short, $r_t < RHS_t$, then moneylenders will demand higher interest payments.

Dhrymes' CCT is meant for testing the weak form of Fisher's hypothesis, which states that there is cointegration among the variables. This paper provides confidence intervals for testing the strong form that $\beta_1 = 1$ in (1).

We use (1) to illustrate typical need to estimate equilibrium relations among jointly evolving series, while overcoming a long list of interdependent problems including: autocorrelated and or heteroscedastic errors, endogeneity, identification, spurious regression, and incidental parameters.

1.1 Motivation for the GPF of Novelty Item (i)

If the RHS variables are correlated with the errors, the ordinary least squares (OLS) estimator is inconsistent needing instrumental variables, which are often obtained from additional equations. For example, one can replace Ep_t by a stable stationary first order autoregressive, AR(1), process in the following second equation:

$$p_t = \mu_0 + \mu_1 p_{t-1} + \xi_t, \tag{2}$$

where ξ_t denotes error in expectations. Substituting (2) in (1) yields:

$$r_t = \beta_0 + \beta_1[\mu_0 + \mu_1 p_{t-1}] + u_t, \text{ where } u_t = \beta_1 \xi_t + \epsilon_t, \tag{3}$$

where the original error ϵ_t is revised to become u_t after incorporating the error in inflationary expectations. The endogeneity problem is removed to the extent that lagged inflation is fully known and therefore exogenous at time t. Unfortunately, this has created a new identification problem since the original deep parameter β_1 is replaced by the product $(\beta_1\mu_1)$.

Dufour (1997) proves 'impossibility theorems' and explains why confidence sets for such 'locally almost unidentifiable' parameters can have zero coverage probability, blaming the 'fundamentally flawed' intervals obtained from inverting Wald-type tests. Moreover, the spurious regression problem, generally solved by differencing, is also related to the Wald-type t test. A pivotal quantities lemma (Spanos, 1999, p.726) shows that properly avoiding Wald-type intervals would need a new pivot. Vinod (1998, 2000) uses the GPF in conjunction with an older independent and identically distributed (iid) bootstrap shuffle and double bootstrap. This paper suggests a radically simplified scalar GPF for confidence intervals on all regression coefficients.

1.2 Error Correction Model (ECM) of Co-integration

A vector autoregression of order p, VAR(p) is:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t, \tag{4}$$

with p vectors of lagged values and p matrices A_i of dimensions $K \times K$ for $i = 1, \ldots, p$, and where $u_t \sim WN(0, \Sigma)$, a K-dimensional white noise process with Σ denoting a time invariant positive definite covariance matrix.

The ECM writes Δy_t as a function of lagged levels y_{t-1} and lagged Δy_t, where lagged Δy_t can be viewed as past equilibrium errors:

$$\Delta y_t = \Pi\ y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + ... + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \tag{5}$$

where Π and $\Gamma_j (j = 1, \ldots, p-1)$ are all $K \times K$ matrices. Let I_K denote the identity matrix of dimension K. There are only $p-1$ matrices Γ_j used in (5) to facilitate the following link with the coefficient matrices of VAR(p) of (4).

$$\Pi = -I_K + A_1 + A_2 + \ldots + A_p, \quad \Gamma_i = -(A_{i+1} + \ldots + A_p). \tag{6}$$

If all individual variables are $I(1)$ or $I(0)$, the VAR(p) is said to be cointegrated, if the rank of Π is deficient, that is, if $Rank(\Pi) < K$. If the equilibrium relation is statistically significant, there will be at least one cointegrating regression relation with $I(0)$ errors, even if the K individual variables are all $I(1)$. One solves the spurious regression problem by testing for the presence of cointegrating relations, instead of relying on biased t statistics from OLS. The following remark argues that random walks are unrealistic for many economic series and perhaps rather rare.

REMARK 1: (Economic equilibria, unpredictability and random walks). Arbitrage ensures that aggregate market equilibrium errors ϵ_t should be unpredictable, in the sense (Samuelson's) that 'properly anticipated prices fluctuate randomly.' Since $y_t = y_{t-1} + \epsilon_t = \sum_{j=-\infty}^{t} \epsilon_j$, 'integrating' unpredictable ϵ_j values, upon assuming infinite memory and zero discounting, can build an economic series as $y_t \sim I(1)$, or a random walk. Humans can only have finite memory and elementary economics teaches us that discounting is ubiquitous. Although the market as a collection of individuals need not fully share individual frailties, infinite memory and zero discounting are too strong and unrealistic assumptions. Even if unit root tests show many economic series are $I(1)$, Maddala and Kim's (1998) survey mentions results that these tests have low power and size distortion. Far from $I(1)$, economic series depend on initial (resource) endowments, are subject to irreversible changes (SEC, FCC, Oil shock) and their content (names of 30 stocks in DJIA) and definitions (GDP deflator) change over time. Equilibrium errors among such series can still be unpredictable, without being subject to the ECM straight jacket.

Dhrymes' CCT, discussed in Section 2 below, assumes that the ECM representation (5) is valid. Noriega and Ventosa-Santaulria (2006) show that spurious regression arises even without $I(1)$ variables, such as when the variables have mixed $I(d)$ features with distinct values of d (including fractional and zero) along with deterministic trends and structural breaks. For example, Phillips (2005, p.153) uses fractional d estimates to conclude that "unit root nonstationarity and short memory are both clearly rejected." for the Fisher equation data studied by him. Hence there is a need for an alternative to the CCT for these general situations. However, if any such extension continues to use differencing or de-trending to achieve stationary data, it is most likely to exacerbate the identification problem, quite similar to Section 1.1 above. After all, an economic equilibrium relation between y_t and x_t with a deep parameter β_1 (say) is much different from a relation between changes Δy_t and Δx_t. The identification problem is that the deep parameters of the original specification might be difficult to recover.

1.3 Motivation for the t-dom v-dom Maps of Novelty Item (ii)

Consider a toy example with $x_t = (4, 12, 36, 20, 8)$ values in the time domain (t-dom). Consider a matrix having (1,2,...,5) in the first column and x_t in the second. Now the order statistics $x_{(t)} = (4, 8, 12, 20, 36)$ are said to belong to the (ordered numerical) values domain (v-dom), are obtained by sorting on the second column while carrying the first column along. The sorted first column yields the reverse map v-dom $\rightarrow$ t-dom.

In the v-dom Vinod's (2004, 2006) maximum entropy bootstrap fits a 'mean preserving' maximum entropy (ME) density and creates J (=999, say) iid resamples. The reverse map (t-dom $\leftarrow$ v-dom) is particularly useful for developing a bootstrap which can re-capture the time subscript and thereby avoid the assumption that $x_t \sim I(1)$. The algorithm with examples and full details is now freely available as the 'meboot' package in R(2006) maintained by J. Lopez-de-Lacalle.

The 'mean preserving' property of the ME density in the values domain becomes the ergodic theorem in the time domain. Using the meboot package, it is trivial to create J reincarnations (resamples) of evolving economic time series. Vinod (2006) suggests using the resamples to estimate what might happen to deep parameters and to construct approximate confidence intervals, while avoiding differencing and de-trending. Although fundamental problems with OLS (e.g., autocorrelation, heteroscedasticity, overestimation of Student's t statistic from OLS applied to trending series) are not removed by simply considering J coefficient estimates, certain confidence intervals do become easier to construct.

1.4 Novelty in HAE Estimation, Item (iii)

Efficient estimation despite autocorrelated and /or heteroscedastic errors are two old problem for which separate tools are already available. Newey (1993) describes efficient estimation under heteroscedasticity of the unknown form from the viewpoint of generalized method of moments (GMM) and instrumental variables, showing how to add moment conditions to improve the asymptotic efficiency. The novelty here is that my moment condition involves a whirlwind tour through the (t-dom $\leftrightarrow$ v-dom) map.

Auxiliary variables needed for efficient estimation are traditionally constructed from the matrix of regressors X wedded to the time domain. For example, Hall (1980) and Cragg (1983) use squares and cross-products of columns of X, whereas Robinson (1987) uses a k-nearest-neighbors (kNN) where nearness refers to nearness in time. When Cragg constructs $\delta_t = \sigma_t^2 - \bar{\sigma}^2$ based on the deviation of residual variances from their average, he is implicitly recognizing that what matters for heteroscedasticity is their numerical magnitude. After showing that magnitudes are best studied in the values domain, this paper suggests a new nonstochastic auxiliary variable constructed from a simple sequence of numbers (1,2,...).

We have seen in Section 1.1 that columns of X often depend on errors, making Cragg's (See his footnote 2) reliance on squares and cross products problematic. Similarly, Robinson's iid assumption is problematic in the time domain, but might have worked better in the values domain. Moreover, my method for sequentially correcting for autocorrelation first and heteroscedasticity next is new, to the best of my knowledge.

1.4.1 Avoiding the Novelty in the Choice of the Minimand

Since heteroscedasticity is a very old problem some authors have attempted to solve it by modifying the least squares objective function itself. Our proposal does not follow this route of changing the problem to find a solution.

Professor C. R. Rao suggested variance components model where there are fewer components than T and a quadratic form $y'Ay$ to estimate them. He minimizes the Euclidian norm of the difference between the true quadratic form and its estimator and called it minimum norm quadratic estimation (MINQUE). The large literature inspired by MINQUE is reviewed by Rao (1977).

Nonparametric generalization of the regression model contains R-estimates based on ranks of residuals minimizing the dispersion of residuals defined by $\sum_{t=1}^{T} w(t)R(\hat{u}_t)(\hat{u}_t)$, where $w(t)$ represents suitably defined weights (often Wilcoxon scores) and $R(.)$ represents ranks. For example, Dixon and McKean (1996) analyze the heteroskedastic linear model using rank-based statistics for both scale and regression coefficients. Even though our proposal uses order statistics, we do not modify the basic minimand of least squares estimation under heteroscedasticity.

It would be an interesting reseach project to compare all these strands under various choices of the minimand in future work. The outline of the remaining paper is as follows. Section 2 discusses an application of Dhrymes' procedure to Irving Fisher's hypothesis. Section 3 discusses a new sequential approach to heteroscedasticity cum heterogeneity and autocorrelation efficient (HAE) estimation. Section 4 describes a Monte Carlo simulation. Section 5 reviews estimating functions (EFs) at the root of Godambe pivot functions (GPFs) with a subsection 5.1 describing how the Frisch-Waugh theorem helps convert a vector pivot to scalar in regression problems. Section 6 describes the scalar GPF bootstrap. Section 7 has a summary and our final remarks.

2 Conformity Cointegration Test (CCT)

Assuming r_0 cointegrating relations among K variables using T data points, Dhrymes' CCT tests whether the rank of a $K \times K$ matrix $\hat{M}$, illustrated in (8) below as a covariance matrix among K variables sandwiched between Π matrices, is less than the 'maximal possible number'. Dhrymes suggests estimating the eigenvalues λ_i of $\hat{M}$. If there is only one cointegrating vector, $(r_0 = 1)$, Dhrymes' null hypothesis of no cointegration for his ' trace' statistic is that the last (smallest) eigenvalue $\lambda_K = 0$. His test statistic is the sample size T times the observed smallest eigenvalue. If the Wald type test statistic is smaller than critical values tabulated by Dhrymes, we accept the null of no cointegration.

Now we test the weak form of Irving Fisher's model of (1) by using the data set called 'Mpyr' available in the R software package called 'Ecdat.' It has annual time series of p_t (logs of net national product price deflator) to measure inflation and r_t (log of commercial paper interest rate in annualized percent rate units) from 1900 to 1989 ($T = 90$). We write the 2×2 variance-covariance matrix of p_t and r_t, vcov(p, r), from Var(p)= 0.6049009, cov$(p, r) = 1.353613$, Var$(r) = 8.407432$. The $\Pi(1)$ matrix is estimated from the VAR estimation of (4) using the R package called 'vars'. Next we write

$$\Pi(1) = A_1 + A_2 - I_2, \tag{7}$$

as the sum of three matrices upon using (6). Now the $\hat{M}$ matrix in Dhrymes' notation is obtained from the matrix multiplication:

$$\hat{M} = \Pi(1)\ \text{vcov}(p, r)\ [\Pi(1)]'. \tag{8}$$

The eigenvalues of vcov(p, r) matrix are: $\lambda_1 = 8.6355903$ and $\lambda_2 = 0.3767427$, respectively. The eigenvalues of $\hat{M}$ are $\lambda_1 = 0.1491416$, and $\lambda_2 = 0.00001179838$. For testing one cointegrating relation, ($r_0 = 1$ in Dhrymes' notation), Dhrymes' test statistic is $T\lambda_2 = 0.001061854$, which is smaller than his tabulated critical value 8.125 at 95% for $K = 2$ (K is q in Dhrymes' notation). Hence the null of zero eigenvalue implying the presence of a cointegrating relation is not rejected, supporting the weak form of Fisher hypothesis, that moneylenders are sensitive to general inflation and are not fooled by nominal returns. Phillips (2005) and many authors in that journal issue have recently revisited the Fisher model without mentioning CCT.

3 Heteroscedasticity and Autocorrelation Efficient (HAE) Estimation.

The usual linear regression model with a general covariance matrix is:

$$y = X\beta + \epsilon, \quad E\epsilon = 0, \quad E\epsilon\epsilon' = \sigma^2\Omega, \tag{9}$$

where X is $T \times p$, β is $p \times 1$, y and ϵ are $T \times 1$. In practice, the large $T \times T$ Ω matrix is rarely, if ever, known. The quasi score function (QSF) represents p "normal equations." Write the QSF at time t assuming $\Omega =$

I_T as: $S_t = X_t'(y_t - X_t\beta)$. The OLS estimator $b = (X'X)^{-1}X'y$, has the covariance matrix $\text{Cov}(b) = \sigma^2(X'X)^{-1}$, where σ^2 is commonly estimated by $s^2 = (y - Xb)'(y - Xb)/(T - p)$. The observable score $\hat{S}_t$ at t is obtained by replacing β by b. This type of estimators using observable scores are, of course, conditional on specification of the original model. Ignoring the Ω matrix makes OLS estimates of β inefficient and standard errors $SE(b)$ potentially misleading. However, Remark 2 explains why one writes (Greene, 2000, p.470) $\Omega(\phi)$ as a function of few parameters in the vector ϕ. Our consistent estimates $\hat{\phi}$ will allow full asymptotic efficiency. For brevity, we often write $\Omega(\hat{\phi}) = \hat{\Omega}(\phi)$ as simply $\hat{\Omega}$.

REMARK 2: (parsimony rule of thumb): The incidental parameters (inconsistency) problem arises whenever the number of unknown parameters to be estimated increases with T. With only T data points, a rule of thumb suggests that we should not have to estimate more than $T/5$ parameters. Since the regression model has p coefficients to be estimated, the number of elements in the ϕ vector of $\Omega(\phi)$ should be less than $(T/5) - p$. That is, it is desirable that: $dim(\phi) < (T/5) - p$.

Assuming a consistent estimate $\hat{\Omega}$ defined above is available, the feasible generalized least squares (FGLS) estimator of β is given by solving $\sum_{t=1}^{T} S_t = 0$ for β as:

$$b_{FGLS} = [X'\hat{\Omega}^{-1}X]^{-1}X'\hat{\Omega}^{-1}y, \text{ with } \text{Cov}(b_{FGLS}) = s^2[X'\hat{\Omega}^{-1}X]^{-1}. \quad (10)$$

The considerable literature dealing with heteroscedasticity and autocorrelation consistent (HAC) estimation helps practitioners who want to use the OLS estimator b but want a better estimate of standard errors from diagonal square roots of a sandwich estimator:

$$\text{Cov}_{HAC}(b) = (X'X)^{-1}[X'\hat{\Omega}X](X'X)^{-1}, \quad (11)$$

where the large $\hat{\Omega}$ matrix enters the expression only through a $p \times p$ matrix: $[X'\hat{\Omega}^{-1}X]$. Zeileis (2007) offers a brief discussion of important references including details on how to construct autocorrelation consistent $\hat{\Omega}$ from

$$\hat{\Omega} = (1/T) \sum_{i,j}^{T} w_{|i-j|} \, \hat{S} \, \hat{S}', \quad (12)$$

where $\hat{S}$ is a $T \times 1$ vector of observed scores $\hat{S}_t = X_t'(y_t - X_t b)$. The lag is represented by $|i - j|$ and weights $w_{|i-j|}$ decrease as the lag increases. Zeileis discusses many weight functions in the literature, including Bartlett, Parzen, Tukey-Hanning, Quadratic-Spectral, with good examples.

Denote the square root matrix of the inverse of either Ω or its estimated $\hat{\Omega}$ version (depending on the context) as $V = \Omega^{-1/2}$ or $\hat{\Omega}^{-1/2}$. Sometimes we write V_r instead of V with $r = (a, h)$, where the subscript 'a' means $\hat{\Omega}$ involves autocorrelation-correction alone and subscript 'h' means heteroscedasticity-correction alone is made. Now premultiply (9) by V and write it as:

$$Vy = VX\beta + V\epsilon, \quad EV\epsilon = 0, \quad E[V\epsilon\epsilon'V'] = \sigma^2 I_T. \tag{13}$$

Let us denote by: $y_r = Vy$ and $X_r = VX$, the revised versions of y and X, respectively. The revised estimator of β is obtained by OLS on the revised model which is equivalent to FGLS of (10).

We suggest making the autocorrelation correction before any heteroscedasticity correction, since certain analytical tools are available for estimation of V_a, which cannot then be combined with estimates of V_h.

If the errors in (9) are AR(1), the order of dynamics of an underlying stochastic difference equation is also 1, and the covariance matrix $\sigma^2\Omega$ is analytically known. For further generalization, it is convenient to insert a subscript i to represent the order of dynamics and write the covariance matrix as:

$$\mathbf{W_L} = \begin{pmatrix} 1 & \lambda_i & \lambda_i^2 & \dots & \lambda_i^{T-1} \\ \lambda_i & 1 & \lambda_i & \dots & \lambda_i^{T-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \lambda_i^{T-1} & \lambda_i^{T-2} & \lambda_i^{T-3} & \dots & 1 \end{pmatrix}$$

$$\Omega_i = d_i \mathbf{W_L}, \text{ where } d_i = (1 - \lambda_i^2)^{-1} \tag{14}$$

Note that this is constructed with only one parameter $\rho = \lambda_i$, which is the regression coefficient of the autoregression and also the root of the characteristic polynomial of AR(1) model given by $(1 - \lambda_i L)$. We can devise a similar matrix for higher order dynamics, and write $\Omega = \sum_{i=1}^{q} \Omega_i$ following Vinod (1985). He gives an explicit expression for $q = 2$, and explains how to write the Ω for ARMA$(q, q-1)$ regression errors when $q > 2$. Since inversion of large matrices may involve numerical inaccuracies, Vinod (1985) suggests analytical approximations for Ω^{-1} matrices needed to develop V_a matrices here.

In the AR(1) case we need not use (14) at all, since V_a becomes a simple bi-diagonal matrix (Davidson and MacKinnon, 2004, p.286). This V_a has ones along the main diagonal, except for $(1-\rho^2)^{1/2}$ in the top left corner, and all sub-diagonal elements are equal to $-\rho$. For $q > 2$, the V_a matrix has to be computed by numerical methods. We suggest eigenvalue decomposition of the symmetric positive definite matrix from ARMA$(q, q-1)$ model for errors as: $\hat{\Omega} = G\Lambda G'$, with G as the matrix of eigenvectors and Λ as the diagonal matrix of eigenvalues. Next compute $V_a = G\Lambda^{-1/2}G'$, which is known to be a numerically reliable inversion and square root operation. Next, we obtain FGLS estimates by OLS on the revised model of (13), and corresponding residuals $(\hat{u}_a)$. The autocorrelation correction using V_a matrix should not be made, if formal testing rejects autocorrelated errors. We do not mean to suggest the FGLS is always preferred to OLS. Vinod (1976) discusses the case where it might not be, since FGLS are based on stronger assumptions.

3.1 Efficient estimator of β using a new V_h to correct for heteroscedasticity

In this subsection, we denote $(\hat{u}_a)$ as $\hat{u}$. Let $H = X(X'X)^{-1}X'$, be the usual hat matrix and let the diagonals be denoted by the subscript 'tt'. It is well known that $var(\hat{u}_t) = E\hat{u}_t^2 = \Omega_{tt}(1 - H_{tt})$. Efficient estimation under heteroscedasticity tries to give a larger weight to observations with a lower $var(\hat{u}_t)$ than higher variance. Accordingly, the following estimates of Ω_{tt} are found in the literature:

(HC0): $s_{tt,0} = \hat{u}_t^2 = [y_t - E(y_t|X)]^2$,

(HC1): $s_{tt,1} = T\hat{u}_t^2/(T-p)$,

(HC2): $s_{tt,2} = \hat{u}_t^2/(1 - H_{tt})$,

(HC3): $s_{tt,3} = \hat{u}_t^2/(1 - H_{tt})^2$ and

(HC4): $s_{tt,4} = \hat{u}_t^2/(1 - H_{tt})^{\delta tt}$, where $\delta tt = min(4, H_{tt}/mean(H_{tt}))$.

Omitting the comma and second subscript yields the generic estimate s_{tt}. Note that HC0 is a proxy for the conditional scale of y_t, which is a non-normal non-negative random variable. Davidson and MacKinnon (2004, p.200) define HC1 to HC3 and suggest a slight preference for HC3 based on the jackknife. Cribari-Neto (2004) suggests HC4.

Although HC0 to HC4 are used in consistent estimation of $SE(b)$, these cannot be used for 'efficient' estimation of β (our HAE) unless we solve the ‘incidental parameters’ problem (See Remark 2 above). Also, we should use formal tests to make sure we have heteroscedasticity before proceeding to correct for it. Cook and Weisberg’s (1983) score test is attractive here because: (i) It allows for dependence of the variance on an arbitrary set of variables, rendering many heteroscedasticity tests as special cases; (ii) It can directly test the null hypothesis in the form of a model proposed in this paper (given later in eq. 18) and is readily implemented as ‘ncv.test’ in the ‘car’ package of R.

If max $(\hat{u}^2)$ is a lot larger (>>) than min $(\hat{u}^2)$, it stands out better in the v-domain. In general, non-constant diagonal *values* of Ω become readily noticeable upon reordering them. Our solution to the incidental parameters problem achieves parsimony by a regression of $log(s_{tt})$ on the sequence $(1, 2, \ldots)$ in the the v-domain.

Before we proceed, note that we cannot rule out a locally perfect fit with $\hat{u}_t = 0$. Assume that T' of the $\hat{u}_t^2$ values are zero. This creates a practical problem that their $log(s_{tt})$ becomes $-\infty$. McCullough and Vinod (2003) show that it is wrong to replace the T' zeros with suitably small numbers close to zero. Hence, let us work with the remaining ‘good’ observations, $T_g = T - T'$, while temporarily excluding the troublesome T' components from the following heteroscedasticity correction algorithm.

3.1.1 Map t-domain to v-domain for heteroscedasticity correction

We construct a $T \times 2$ matrix W, having the first column containing the set $\tau' = 1, 2, ..., T$, and the second column containing s_{tt}. Next, we sort the W matrix on the second column, ordering its elements from the smallest to the largest, while carrying the first column along during the sorting process. This finds the monotonic order statistics $s_{(tt)}$ belonging to the v-domain in the second column, where the first T' elements will be already made zero in the sorting process. Now, the remaining T_g (good) elements will be nonzero, with well-defined logarithms. Use a subscript ‘s’ to denote the sorted version of the W matrix: $W_s = W_{s,i,j}$, where its elements for row i and column j bear the subscript (s,i,j). Denote its columns 1 and 2 as $\tau'_s = W_{s,.,1}$ and $s_{(tt)} = W_{s,.,2}$, respectively, replacing the ‘i’ by a dot, using Professor Dhrymes’ notational convention.

Clearly, efficient estimation problem cannot be solved without some assumptions about the process generating Ω_{tt}.

ASSUMPTION A1 (smoothness of heteroscedasticity): Denote the order statistics of true Ω_{tt} by $\Omega_{(tt)}$ after mapping them into the v-domain. We assume that all Ω_{tt} lie on a piecewise smooth function $f(\Omega_{(tt)})$ in the (numerical values) v-domain.

The intuition behind the assumption is simply that conditional scaling of $E(y_t|X)$ is fundamentally related to the numerical values (magnitudes of y and X) in the v-domain, rather than their timing. In practice, the assumed smooth function $f(\Omega_{(tt)})$ in the v-domain is unknown and kernel smoothing which allows smooth free-shaped curves is a possibility. Instead, we use polynomial smoothing by exploiting the monotonicity of $s_{(tt)}$ in the v-domain. Since variances must be positive, we use an exponential link function between observable $s_{(tt)}$ and the unknown $f(\Omega_{(tt)})$. That is, we first note that log is a monotonic transformation and seek a smooth intermediate function $g(\log s_{(tt)}, z)$, where z denotes a suitable set of exogenous variables. Applying the exponential link, $exp(g) > 0$, yields the smooth function $f(\Omega_{(tt)})$ as a function of z in the v-domain.

Any set of monotonic variables can be our exogenous z. For simplicity we define $\tau = T' + 1, T' + 2, \ldots, T' + T_g$, and propose the non-decreasing set $\tau^k (k = 0, 1, .., h)$. Now estimate g by a polynomial regression of $\log s_{(tt)}$ on (h+1) column vectors τ^k, of dimension $T_g \times 1$:

$$g(\log s_{(tt)}, \tau) = \sum_{k=0}^{h} \phi_k \tau^k + \epsilon_s. \tag{15}$$

We choose the h with some trial and error, perhaps starting with a quintic ($h = 5$) and using the R^2 of (15) adjusted for degrees of freedom as a guide. We also recommend graphs to assess the shapes. The final choice of h should try to satisfy the rule of thumb of Remark 2 above, now modified to be $(T/5) - p - h - 1 > 0$. Let (15) represent a model after the suitable h is found.

Let X^τ denote the matrix of regressors in (15). The normal equations to obtain the OLS estimates of ϕ must be solved simultaneously and represent the following 'moment condition' (in the GMM literature terminology) imposed in the values domain. It is expected to yield improved efficiency, Newey (1993).

$$E\left[X^{\tau\prime}(g(\log s_{(tt)} - \tau) - \sum_{j=0}^{h} \phi_j \tau^j)\right] = 0. \tag{16}$$

ASSUMPTION A2: The polynomial regression in (15) satisfies the usual assumptions: validity of (15), $E(\epsilon_s) = 0$, the matrix of regressors is full rank and satisfies the Grenander conditions (Greene, 2000, p. 354) for convergence, regressors are uncorrelated with errors (exogenous), and $E\epsilon_s\epsilon_s' = \sigma_s^2 I$.

Since all variables in (15) are monotonic, we expect a good fit, but need to guard against collinearity. This is why the full rank part of Assumption A2 is needed. Other parts of Assumption A2 ensure that the following well-known Lemma holds, where the proof can be omitted since exogenous regressors readily satisfy Grenander's conditions:

LEMMA 1: Let $\hat{\phi}$ denote the $(h+1) \times 1$ vector of ordinary least squares estimates of coefficients ϕ in (15). Let $\Rightarrow$ denote convergence as $T \to \infty$. We have $\hat{\phi} \Rightarrow \phi$ in the values domain.

Using the Lemma we estimate the smooth function $f(\Omega_{(tt)})$ as $exp(\hat{g})$, where $\hat{g}$ is:

$$\hat{g}(\log s_{(tt)}, \tau) = \sum_{k=0}^{h} \hat{\phi}_k \tau^k, \tag{17}$$

By properties of OLS, $\hat{g}$ obtained by a form of trend-fitting provides an unbiased (consistent) estimator of $g, \hat{g} = g + residuals$. If some of these residuals are larger than a tolerance constant, corresponding observations are 'outliers,' for our smooth function. Next, we can apply a truncation window such that we simply delete additional T'' observations as outliers. The beauty of the v-domain is that we can omit $T' + T''$ elements, truncating the left hand side of (15) with impunity. We can simply use $\tau' = 1, 2, ..., T$, instead of $\tau = T' + 1, T' + 2, ..., T' + T_g$, on the right hand side of (17) to get the right number of estimates in the time domain. Of course, it needs the reverse map from values to time domain, described next.

3.1.2 The Reverse Map: t-dom $\leftarrow$ v-dom.

The true unknown smooth function $f(\Omega_{(tt)})$ is approximated by $exp(\hat{g})$ in the v-domain. We still need to map this into the time domain to get the diagonals Ω_{tt} of Ω representing heteroscedasticity. Substituting the $T \times 1$ vector τ' on right side of (17) yields a $T \times 1$ vector $\hat{g}$, as desired. Note that the initial T' components of $\hat{g}$ are most likely to be negative and large (but not $-\infty$). The T'' outliers are likely to be scattered in the sample. Still, we can replace the second column of the W_s matrix by $exp(\hat{g})$, and sort on the first column till it has elements equaling τ', yielding a doubly sorted $T \times 2$ matrix denoted as W_{ss}, where the subscript 'ss' suggests double sorting. Let the individual elements in the second column of W_{ss} be denoted by $W_{ss,t,2}$, which are time domain $\hat{\Omega}_{tt}$ quantities. Finally, our proposed correction for heteroscedasticity uses the transformation matrix:

$$V_h = diag(1/W_{ss,t,2})^{1/2}. \tag{18}$$

LEMMA 2. The sorting map from the time domain to the values domain and the reverse map from the values domain to the time domain are linear (matrix) operations.

PROOF: Let the t-domain have $x_t = (4, 12, 36, 20, 8)$, whose map to the v-domain converts first column $\tau' = 1 : 5$ to $\tau'_s = (1, 5, 2, 4, 3)$. Start with I_5 (identity matrix) and rearrange the diagonal ones to positions given by τ'_s to create a O^{rev} matrix, as exhibited at Vinod (2006, p.964). Verify that the reverse map amounts to pre-multiplication by O^{rev}. Using the inverse matrix $(O^{rev})^{-1}$ note that both maps in (t-dom $\leftrightarrow$ v-dom) are linear operations. Using induction generalizes this to any T.

3.1.3 Cook-Weisberg Heteroscedasticity Testing.

Now we turn to the Cook-Weisberg test using z_{ij} matrix of proxy data for heteroscedasticity. The (j+1)-st column of z_{ij} contains j-th power of sorted τ'_s for $j = 0, 1, \ldots, h$. The true slope coefficients ϕ_j for $j = 1, \ldots, h$ of the polynomial of order h are all zero, under the null of homoscedasticity. The Cook-Weisberg model is:

$$\Omega_{tt} = exp(\sum_{j=0}^{h} \phi_j z_{ij}), \tag{19}$$

where z_{ij} is a $T \times (h + 1)$ known matrix of arbitrary known quantities, which may or may not be related to one or more of the columns of X in their framework. White (1980) considers a similar z_{ij} from 'all second order products and cross products of original regressors.' However, Greene (2000, p. 509) criticizes that White's test is 'nonconstructive,' since upon rejecting the null of homoscedasticity it fails to suggest a remedy. The algorithm proposed in this paper avoids such criticism. Conditional on our choice of the z_{ij} matrix, the Cook-Weisberg method tests the null of homoscedasticity. The validity of our choice of z_{ij} in (19) depends on our theorem proved in the sequel.

3.1.4 Analogy with Spectral Analysis.

Our v-domain is somewhat analogous to the frequency domain of spectral analysis. Priestley (1981, p.432) states that the variance of sample periodogram does not to tend to zero as $T \to \infty$, because it has "too many" sample autocovariances. Our incidental parameters problem is almost the same. Two Fourier integrals, Priestley (1981, p.201), allow two-way maps between the time and frequency domains. Our double sorting is simpler than Fourier integrals and allows similar two-way mappings between the time and v-domains. The spectral kernel smoothers omit (down-weight) sample periodogram values outside a 'window' to satisfy a technical assumption similar to our A3 below. We also omit T'' additional 'outliers' failing to satisfy the smoothness assumption. Similar to the following Theorem, Priestley (p.464) proves consistency results in spectral analysis and notes the simplifying value of linearity.

ASSUMPTION A3 (truncation of s_{tt}): As the sample size increases we omit a certain number of s_{tt} values outside a truncation window and keep only T^{α} with $(0 < \alpha < 1)$, so that $(T^{\alpha}/T) \to 0$, as both T and $T^{\alpha} \to \infty$.

Since we are keeping only $T^{\alpha} = T - T' - T''$ observations, we can satisfy Assumption A3 provided $T' + T'' > 0$.

THEOREM 1: Under Assumptions A1 to A3, the $W_{ss,t,2}$ in (18) yields consistent estimates of Ω_{tt} in the time domain, implying that $\Omega(\hat{\phi}) \Rightarrow \Omega(\phi)$.

PROOF: Although the variance of each s_{tt} is $O(1/T)$, the variance of the vector s_{tt} is $O(1)$. In the v-domain s_{tt} become $s_{(tt)}$, the order statistics. There we omit $T'+T''$ observations to satisfy Assumption A3, so that $var(s_{(tt)}) \Rightarrow 0$ for the entire set. Now use Lemma 1, Assumption A1 and Slutsky's theorem to verify that $\hat{g}(\log s_{(tt)}, \tau) \Rightarrow f(\Omega_{(tt)})$, the smooth function linking the diagonals of Ω in the v-domain. Upon substituting in (17) and again using Slutsky's theorem, $exp(\hat{g}) \Rightarrow \Omega_{(tt)}$, still in the v-domain. Now $W_{ss,t,2}$ is a t-domain image of $exp(\hat{g})$ and Ω_{tt} is a t-domain image of $\Omega_{(tt)}$. Since the map from v-domain to time domain is linear by Lemma 2, $W_{ss,t,2} \Rightarrow \Omega_{tt}$. QED.

ASSUMPTION A4: The alternative to $\Omega(\phi)$ of (9) is contiguous satisfying: $\Omega_{(tt)} = [1 + 2BT^{-1/2} f_t(X, \beta, \phi)]\Omega_{(tt)}$, where f_t is an arbitrary unknown function satisfying: $T^{-1}\Sigma_{t=1}^{T} f_t^2 \to \mu$, $(0 < \mu < \infty)$, and B is an arbitrary scalar. The FGLS uses preliminary OLS estimate b, satisfying $T^{1/2}(b-\beta) = O_p(1)$, to compute the residuals and $\hat{\phi}$ satisfies $T^{1/2}(\hat{\phi}-\phi) = O_p(1)$. Also suppose that for some positive definite matrix S_{pd}, we have $T^{-1}[X'\Omega^{-1}X] \Rightarrow S_{pd}$, and errors ϵ in (9) are normally distributed.

ASSUMPTION A5: $\text{plim}_{T\to\infty} T^{-1}[X'\hat{\Omega}^{-1}X] = \text{plim}_{T\to\infty} T^{-1}[X'\Omega^{-1}X]$, and $\text{plim}_{T\to\infty} T^{-1/2}[X'\hat{\Omega}^{-1}\epsilon] = \text{plim}_{T\to\infty} T^{-1/2}[X'\Omega^{-1}\epsilon]$.

THEOREM 2: Assuming A1 to A5 the FGLS of (10) is efficient and robust. The asymptotic distribution of $T^{-1/2}$ $[b_{FGLS} - \beta]$ is normal, $N(0, S_{pd}^{-1})$, under either the original $\Omega(\phi)$ or its contiguous alternative specifications.

PROOF: The efficiency of FGLS has been established in the literature, Greene (2000, sec.11.4), provided A5 holds and provided $\Omega(\hat{\phi}) \Rightarrow \Omega(\phi)$, which is proved in Theorem 1. Assumption A4 and a proof of robustness under contiguous specification alternatives are in Carroll and Ruppert (1982).

The familiar generalized autoregressive conditional heteroscedasticity (G-ARCH) models fit an ARMA model in the t-domain, where the conditioning is on the past values of observable $s_{t-1,t-1}$ (volatility). Any relationship between conditional and v-domain heteroscedasticity is hinted by a time series plot of sorted τ_s' against τ' and by the autocorrelation function for τ_s'. Denote the Pearson correlation coefficient between *lagged* $\tau'[1 : T-1] = 1, 2, \ldots, T-1$ and $\tau_s'[2 : T]$ by $\text{COR}(L\tau', \tau_s')$. If the samples follow independent normal distributions, (which they do not) COR follows the Student's t distribution (with $T-2$ degrees of freedom). In our case, the p-value for testing the null hypothesis of zero correlation is a heuristic. If conditional heteroscedasticity is present, the COR will be 'significantly' nonzero by our heuristic. In general, monotonic patches among the τ_s' suggest GARCH effects and a further study might reveal interesting insights.

This completes our discussion of estimation of the square root of the inverse of $\hat{\Omega}$ matrix, denoted as V, to sequentially correct for both autocorrelation and heteroscedasticity, in that order, and only as needed in a given problem. Our new HAE algorithm is implemented via these V matrices using the standard OLS software on (13). The choice of the dynamic order q for autocorrelation and polynomial of order h for heteroscedasticity should try to satisfy the parsimony rule of thumb: $(T/5) - p - h - 3 - 2q > 0$.

3.2 An Example of Heteroscedasticity Correction

Since the heteroscedasticity correction (18) is new, it is important to discuss its operation with the help of an example. Instead of Irving Fisher's model with only one regressor, it is perhaps more useful to illustrate the HAE algorithm with a model having two regressors. Let us use a Keynesian consumption function designed for testing Friedman's permanent income hypothesis. We can be extremely brief here, since the model and data are discussed in Murray's (2006) textbook and also in a vignette available on the Internet as a part of my 'meboot' package. Let C =log (personal consumption), Y =log (disposable income). The consumption function model is:

$$C_t = \beta_1 + \beta_2 C_{t-1} + \beta_3 Y_{t-1} + \epsilon_t. \tag{20}$$

Under rational expectations, the regressor C_{t-1} has all relevant information regarding current consumption. Hence, if Friedman's permanent income hypothesis is supported, the coefficient of lagged income should be insignificant. This is indeed the case based on OLS estimation of (20) available in Murray (2006, p.800) and omitted here for brevity. The standard error and the t-statistic for $\hat{\beta}_3$ are respectively: (0.14389, 0.187). If we use the heteroscedasticy corrected (HC) standard errors (using R software package 'sandwich') these become (0.1613667, 0.1670287), respectively, without changing the conclusion.

Upon fitting (20) to data, none of the OLS residuals is zero ($T' = 0$ here). The residual autocorrelations for lags 0 to 6 are respectively: 1, 0.146, -0.035, -0.088, -0.088, -0.058, and -0.104. Their plot remains inside the confidence band. Also, the Breusch-Godfrey test statistic 1.0816 with 1 degree of freedom has the p-value of 0.2983, supporting the null of no autocorrelation at the 5% level. For alternatives of orders 2 to 4 the respective p-values are: (0.5366, 0.6711, 0.7704). This means we do not need to make any correction for autocorrelation with a V_a matrix here. Even if we did, we can effectively work with: $y = X\beta + u$ as the model, before turning to the new heteroscedasticity correction.

Table 1 lists key results for 20 choices of HCj for j=0,...4 and τ^h for $h = 1, \ldots 4$. The first two columns identify the HCj and τ^h. Although the v-domain is clearly unsuitable for studying GARCH effects, Column 3 of Table 1 contains heuristic p-values for a t test on the simple correlation coefficient COR$(L\tau', \tau'_s)$, where small p-values suggest rejection of independence and support presence of GARCH effects. Column 4 has p-values for Cook-

Weisberg score test for the presence of heteroscedasticity, based on our new choice for their artificial matrix of z_{ij}. The $(j+1)$-st column of z_{ij} contains j-th power of sorted (τ_s') for $j = 0, 1, \ldots, h$. If there is no heteroscedasticity, all the coefficients ϕ in (19) will be insignificant. Since the p-values in column 4 are near zero, homoscedasticity is rejected for all 20 choices.

Column 5 of Table 1 has the adjusted R^2 for the regression of logs of sorted $(\hat{u})^2$ on powers of τ in (17). Column 7 has the t-statistic on β_3 in (20) given in column 6 associated with Y_{t-1} after the pre-multiplication by V_h based on the particular choice of HCj and polynomial power h. We choose two rows marked with a (*) for further analysis: HC0 with linear polynomial with the lowest t- statistic on β_3 and HC3 with quartic ($h = 4$). The starred choices have adjusted R^2 =0.8988 and 0.9902, respectively.

Figure 1 reports graphs for the linear HC0 case. Upper panel plots the order statistics of logs of squared residuals $s_{tt,0}$ as the solid line along with a dashed line for the fitted values from a simple straight line in time, representing our smooth intermediate function $g(\log\, s_{(tt)}, z)$. After computing $exp(g)$ and using the reverse map, we get the second column of doubly sorted W_{ss} matrix. The lower panel plots them in time domain as the dashed line along with the solid line representing original squared residuals over time. It is clear from both panels that with only 2 parameters of ϕ (intercept and coefficients of τ) we are able to get good estimates of heteroscedastic variances, thanks to the double sort.

Figure 2 is similar to Figure 1, except that we have the quartic case with HC3 here, showing that the quartic fit is better. After all, HC3 with the quartic has a high (=0.9902) adjusted R^2 in Table 1. The figures show that rearranging observations in an increasing order of squared residuals can reveal hidden heteroscedasticity with just a few additional parameters in ϕ. The original OLS, as well as all 20 cases of efficient estimation show significant β_3 implying rejection of Friedman's permanent income hypothesis.

Table 2 reports details for the row HC3* of Table 1. It has feasible GLS estimates after heteroscedasticity correction by a quartic under HC3 transformation of squared residuals. The F(2,47) statistic for the overall fit is 5.098e+06, with a near zero p-value. The OLS coefficient of Y_{t-1}, which was statistically insignificant before heteroscedasticity correction, has now become significantly different from zero at the 5% level. Table 2 confidence intervals do not support Friedman's hypothesis.

It is interesting that if we use per capita consumption and disposable income without the log transformation, then the Friedman hypothesis has support. The version in levels needs a first order autocorrelation correction with V_a and summary results are in Table 3. All models in Table 3 yield insignificant 't-stat' values in column 7 for the estimates of β_3.

REMARK 3: (HAE estimation): We have shown how to correct for autocorrelation and heteroscedasticity (in that order) by incorporating them in V_r matrices with $r = (a, h)$, and how the original model becomes: $Ey_r = X_r\beta$. It is shown in (13) that after incorporating V_r, we can set $\Omega = I$ for all practical purposes, without loss of generality.

4 A Monte Carlo Simulation of Efficiency of HAE

Since assumptions A1 to A5 are not always easy to verify, we supplement our theorems with a simulation experiment. Long and Ervin (2000), Godfrey (2006), Davidson and MacKinnon and others have simulated the size and power of HAC estimators of SE(b). Cook and Weisberg's (1983) simulation used cloud seeding data having $T = 24$ observations. Let us inject objectivity into our Monte Carlo design by combining the Monte Carlo designs used by Long-Ervin with that of Cook-Weisberg. Of course, our focus is on efficient estimation of coefficients themselves, not inference. We pick x_1 to x_4 data from a cloud seeding experiment (suitability criterion 'sne,' 'cloudcover,' 'prewetness' and rainfall). These were among those chosen by Cook-Weisberg to allow wide variety of heteroscedasticity possibilities.

Our dependent variable is constructed artificially (as in both designs) by the relation:

$$y = b_0'1 + b_1'x_1 + b_2'x_2 + b_3'x_3 + b_4'x_4 + e, \tag{21}$$

where $b_\ell' = 1$, for all $\ell = 0, 1, \ldots, 4$, as in Long and Ervin (except that their b_4' is zero) and where 'e' represents a vector of T random numbers chosen according to one of the following methods, which are called scedasticity functions by Long and Ervin. They have a far more extensive simulation and their focus is on HAC estimators. As $j = 1, \ldots, J$ (=999) let $\epsilon_{df5,j}$ denote a new vector of Student's t distributed (fat tails) independent pseudo-random numbers with five degrees of freedom.

SC1): $e = \epsilon_{df5,j}$. (no heteroscedasticity)

SC2): $e = (x_1)^{1/2}\epsilon_{df5,j}$. (disallows $x_1 < 0$)

SC3): $e = (x_3 + 1.6)^{1/2}\epsilon_{df5,j}$.

SC4): $e = (x_3)^{1/2}(x_4 + 2.5)^{1/2}\epsilon_{df5,j}$.

SC5): $e = (x_1)^{1/2}(x_2 + 2.5)^{1/2}(x_3)^{1/2}\epsilon_{df5,j}$.

Our next task is to use HAE estimators from (18) of the slopes b'_1 to b'_4 in (21) in Monte Carlo simulation experiments with small $T < 100$. If asymptotic results of our theory hold for such T, the efficiency of GLS should be greater than that of OLS.

My simulation program creates a 4 dimensional array with dimensions (4, 6, 999, 5). The first dimension is for the estimates of $p = 4$ slope coefficients. The second dimension with 6 values is for the OLS, and five HAE estimators denoted by HC0 to HC4 and described earlier in our discussion before eq. (15). Of course, the 'C' in HC refers to 'correction' by our eq. (18) not to the usual 'consistency' in the sense used by Long and Ervin. The last dimension is for SC1 to SC5 in increasing order of heteroscedasticity severity.

After computing the standard deviations of 999 coefficient estimates we construct a summary array of dimension (4, 6, 5). It is convenient to suppress the first dimension and average the standard deviations over the 4 coefficients. Next, we divide the standard deviations for HC0 to HC4 by the standard deviation for OLS, reducing it from 6 to 5. The final results are reported in Figure 5, where we look for numbers staying below the OLS vertical value of 1. The numbers 1 to 5 on the horizontal axis refer to HC0 to HC4. In this experiment, the sophisticated HC3 and HC4 corrections do not seem to offer great advantages in terms of efficiency. Since several values are below 1, many of our procedures are indeed more efficient than OLS.

Not surprisingly, the efficiency improvement is generally higher when heteroscedasticity ought to be intuitively more severe (by looking at the complications of the formulas for SC1 to SC5 given above), although the intuition can fail in specific examples. In Figure 5 the efficiency gain is the highest for graph for the most severe SC5 (line marked '5') and lowest for the (SC1 marked as '1') representing homoscedasticity, as might be expected. One can easily guard against SC1 by formal heteroscedasticity testing. Figure 6 is similar to Figure 5, except that here we use $b'_4 = 0$ in (21) as in Long and Ervin. The efficiency gains over OLS continue to be achieved using correction formulas of HC0 to HC4 where all lines are below unity.

In another experiment, I use (21) without the x_4 variable and economic data with non-missing $T = 46$ observations from the 'USfygt' data set of my 'meboot' package. The x_1 to x_4 are: 'fygt1, infl, reallir, and usdef'. Details are omitted to save space. Again, efficiency gains are clear except for SC1. Our experiments support the econometric practice of formally testing for heteroscedasticity before considering any corrections.

It is surprising that Long and Ervin's (2000) large simulation finds that for typical sample sizes in economics ($T < 100$) the commonly used HC0, HC1 and HC2 methods provide inferior inference (in size and power) than the simplest $s^2(X'X)^{-1}$ of OLS. In other words, OLS is hard to beat with $T < 100$. Yet I have chosen $T = 24, 46$ to raise the bar. My HAE method is worthy of study, since it is able to reduce the variance of OLS (over $J = 999$ experimental values) when heteroscedasticity is present in the model. Cook and Weisberg emphasize a need to supplement simulations with graphics. Figures 1 to 3 illustrate some interesting patterns of heteroscedasticity in econometric applications.

5 Godambe Pivot Functions (GPFs)

Having completed a discussion of HAE estimation, we turn to replacing Wald-type pivots by more robust Godambe pivots in our construction of confidence intervals. Estimating functions (EFs) by Godambe (1960) and Durbin (1960) are similar to moment conditions in the GMM literature. Wedderburn (1974) developed quasi maximum likelihood (QML) estimators. The EF theory relies on Wedderburn's lead for robustness and seeks to find optimal EFs satisfying several desirable properties. Most remarkably, when the EF estimators differ from the QML (for certain limited dependent variable problems recognized by Wedderburn) Vinod (2002) proves the superiority of EF estimators over rather time-honored principles of least squares (Gauss) and maximum likelihood (Fisher). Heyde (1997) explains that unbiasedness, efficiency, sufficiency, reaching the Cramer-Rao bound are all properties of the function itself, quite apart from similar properties of the roots (QML estimators). Heyde also discusses small sample advantages of the EF viewpoint.

Assuming that σ^2 and Ω are known, 'normal equations' for (9) are p equations in p unknowns in the β vector. They yield the QSF, mentioned earlier in Section 3, as the following function of data and parameters when $\Omega \neq I_T$:

$$S(y, X, \beta) = X'\Omega^{-1}(y - X\beta) = X'\Omega^{-1}\epsilon. \quad (22)$$

Vinod (1997) notes that this quasi score is an optimal EF. Vinod (1998) explains that GPFs achieve greater robustness by not satisfying the so-called information matrix equality: $I_F = I_{2op} = I_{opg}$, involving matrices for Fisher information, outer product of gradients (opg) and second-order-partials (2op) of the log-likelihood. Note that the equality holds only for a non-robust distribution having zero skewness and zero excess kurtosis.

Denoting the optimal estimating function by g*, Godambe (1985) proposed:

$$GPF = \sum_{t=1}^{T} g_t^* \left[\sum_{t=1}^{T} E(g_t^*)^2 \right]^{-1/2}. \quad (23)$$

Vinod (1998, 2000) extends the GPF to the vector case and applies it to the regression problem (9). He writes it as a sum of T scaled scores $\tilde{S}_t$, with scale factor S_c as:

$$GPF = \sum_{t=1}^{T} S_t/S_c = \sum_{t=1}^{T} \tilde{S}_t, \text{ where } S_c = [\sum_{t=1}^{T} E(S_t)^2]^{1/2}. \quad (24)$$

Vinod's scaled score function of GPF is a $p \times 1$ vector:

$$z_G = [X'\Omega^{-1}X]^{-1/2} X'\Omega^{-1}[(y - X\beta)/\sigma] = \sum_{t=1}^{T} \tilde{S}_t, \quad (25)$$

where the scale factor $S_c = [X'\Omega^{-1}X]^{-1/2}$ does not depend on the unknown parameters β. Write the asymptotic covariance matrix of the GLS estimator as:

$$cov(b_{GLS}) = (A_{SE})^2 = (I_F)^{-1}, \quad (26)$$

where I_F is Fisher information matrix $[X'\Omega^{-1}X]/s^2$, where s^2 is the residual mean square, and where A_{SE} denotes an asymptotic standard error matrix. Hence the vector version of Fisher's pivot function (FPF) used in Wald-type tests (we are trying to avoid) is

$$FPF = (b_{GLS} - \beta)(A_{SE})^{-1}, \quad (27)$$

where b_{GLS} are viewed as roots of $S(y, X, \beta) = 0$. The GPF expression (23) avoids the roots and their sampling distribution. Since we admit $z_G \neq 0$, the division by σ needs to be recognized explicitly, as we do in (25).

Upon writing the GPF as a sum of T scaled scores in (25), which might be dependent in finite samples, Vinod (1998) lists the assumptions for McLeish's (1974) central limit theorem for dependent data. His Proposition 1 proves that $z_G \Rightarrow N(0, I)$, and that GPF are more robust than the FPF of (27), making them attractive in finite samples. For quick verification substitute the matrix $S_c = [X'\Omega^{-1}X]^{-1/2}$ into (25) to yield: $z_G = S_c X'\Omega^{-1}[\epsilon/\sigma]$, so that its expectation, $Ez_G = 0$, holds. Now its variance covariance matrix is:

$$\begin{aligned} Ez_G(z_G)' &= S_c X'\Omega^{-1} E[\epsilon/\sigma][\epsilon/\sigma]'\Omega^{-1}[XS_c'] \\ &= S_c X'\Omega^{-1}[\sigma^2\Omega/\sigma^2]\Omega^{-1}[XS_c'] \\ &= S_c X'\Omega^{-1} X S_c' = I_p. \end{aligned} \tag{28}$$

REMARK 4: Since the GPF has several desirable robustness properties useful in finite samples, it is a good candidate for statistical inference whenever one wants to avoid Wald-type statistics. The roots of $S(y, X, \beta) = 0$ are quasi maximum likelihood QML (or GLS) estimators of β. In (25) the scale factor $[(X'\Omega^{-1}X)^{-1/2}/\sigma] > 0$ cancels whenever the right side is zero. Therefore, the QML roots coincide with the p roots of the p pivot equations: $z_G = 0$. Next, we construct a 95% confidence interval by solving two equations $z_G = \pm 1.96$, using the familiar quantiles of N(0,1) variate.

If one is interested in multivariate joint confidence regions, Vinod (2000) suggests the p-dimensional ellipsoid defined over the p dimensional parameter space of β:

$$z_G' z_G \leq \chi^2(0.95, p), \tag{29}$$

where the right hand side has the upper quantile of a Chi-square variate with cumulative probability 0.95, and with p degrees of freedom available in usual tables.

A similar confidence region described in econometrics texts, Davidson and MacKinnon (2004, p. 193), uses the FPF of (27) above. Since practitioners cannot readily use ellipsoids, approximate confidence intervals need to be obtained by a grid search for the largest and smallest values of each component of β simultaneously satisfying (29). A good way to avoid grid search, in situations where it can be avoided (not always) is to use the Frisch-Waugh theorem as spelled out in the next subsection. Of course, there will be situations when one needs to explicitly take account of joint tests on two or more parameters in β and work with confidence regions.

5.1 Frisch-Waugh Theorem and scalar pivot for multiple regression problems.

Remark 3 at the end of Section 3 notes that after suitably pre-multiplying by the V_r matrix, we can simply set $\Omega = I$ for all practical purposes and again work with: $y = X\beta + u$, and $Euu' = I_T$. The Frisch-Waugh theorem, discussed in detail in Davidson and MacKinnon (2004) allows us rewrite the model after partition of $X = [X_1, X_2]$ as:

$$y = X_1\beta_1 + X_2\beta_2 + u, \tag{30}$$

where our X_1 has only one column and X_2 has $p \times 1$ columns. Our interest is temporarily focused on the coefficient β_1 of the first regressor. Define

$$M_2 = (I_T - X_2(X_2'X_2)^{-1}X_2'), \tag{31}$$

as a $T \times T$ projection matrix operator for getting the residuals from a regression. The Frisch-Waugh theorem states that the estimate of β_1 and residuals from (30) are numerically identical to the corresponding values of the following abridged regression:

$$M_2\, y = M_2\, X_1\beta_1 + u, \tag{32}$$

where M_2y and M_2X_1 are both $T \times 1$ vectors. We call (32) abridged, because the transformed regressor M_2X_1 has only one column achieving a 'radical' simplification reducing the β vector to a scalar β_1. The scalar GPF (23) can now be used to construct a confidence interval for β_1.

The choice of the first column in the partition $X = X_1 + X_2$ can be rotated to allow construction of confidence intervals for the remaining elements of β. Hence, there is little loss of generality in our sequential approach. Even if an economist's hypothesis involves two regression coefficients, certain simple rearrangements of the model are available to convert the inference problem as inference on β_1. For example, let the regression for a Cobb-Douglas production function be: $log(y) = \alpha_0 + \alpha_1 log(K) + \alpha_2 log(L)$, with output y, capital K and labor L. Assume that the researcher wants a confidence interval on the sum of two coefficients $(\alpha_1 + \alpha_2)$ as a measure of economies of scale. Let us rewrite the model as:

$$log(y) = \beta_1[log(K) + log(L)] + \beta_2\iota + \beta_3[log(K) - log(L)], \tag{33}$$

where ι denotes a column vector of ones for the intercept. Equating the right side of (33) with that of the original Cobb-Douglas model, we have following simple relations between the coefficients of the two versions: (i) $\alpha_0 = \beta_2$. (ii) The first coefficient of (33), $\beta_1 = (\alpha_1 + \alpha_2)$, indeed measures economies of scale. (iii) Since $\beta_3 = \beta_1 - \alpha_2$ and $\beta_3 = \alpha_1 - \beta_1$, it indirectly measures both α_1 and $-\alpha_2$, which means that α_1 and α_2 are not individually identified when one estimates (33).

Note that (32) amounts to regressing $y^o = M_2\, y$ on $x^o = M_2 X_1$, while forcing the line of regression through the origin. Hence the quasi score function $X'\epsilon$ becomes: $\sum_{t=1}^{T} x_t^o(u_t) = \sum_{t=1}^{T} x_t^o(y_t^o - x_t^o\beta_1)$. The score at time t becomes:

$$S_t = x_t^o(u_t) = (x_t^o y_t^o) - (x_t^o)^2\beta_1. \tag{34}$$

Next, we need the scale factor involving expectations: $E(S_t)^2 = E(x_t^o u_t)^2 = (x_t^o)^2 E(u_t)^2 = (x_t^o)^2\sigma^2$. Hence the scale factor is: $S_c = [\sum_{t=1}^{T}(x_t^o)^2\sigma^2]^{1/2}$ and the scalar GPF for the regression problem is given by:

$$z_G = \sum_{t=1}^{T}(\tilde{S}_t), \text{ where } (\tilde{S}_t) = S_t/S_c. \tag{35}$$

From our discussion near (28) recall that $z_G \Rightarrow N(0,1)$, in the scalar case $(p = 1)$. Following Remark 4 we solve two equations: $z_G = (\pm 1.96)$ from (35) for the single unknown slope parameter β_1 of (34) to estimate its confidence limits. Next, we write a loop to provide confidence limits for all regression coefficients by sequentially placing different columns of X into the first position. Thus the 'radical' simplification of the regression problem demanded by the scalar GPF of (23) eventually yields GPF confidence intervals for each coefficient in β.

6 Bootstrap Using Godambe Pivots

The main appeal of any bootstrap is that it provides robust confidence limits, without assuming normality. Recall that many inference problems in economics suffer from the identification problem so that the usual Fisher pivot of a Wald test is problematic in light of Dufour's impossibility theorems. Now we avoid the problematic pivot and use the scalar GPF of (23) for bootstrap inference on a typical regression problem in econometrics.

Note that our bootstrap needs to create a large number (j=1,..,J=999, say) of resamples of scaled scores at time t:

$$z_F = \sum_{t=1}^{T} (\hat{S}_t)/\hat{S}_c, \quad (\hat{S}_t) = x_t^o \hat{u}_t^o, \quad \hat{S}_c = [\sum_{t=1}^{T} (x_t^o)^2 s^2]^{1/2}, \tag{36}$$

where the scale factor $\hat{S}_c$ does not change from one resample to the next. After computing the common scale factor once, we can focus on obtaining J resamples of scores at time t denoted by $(\hat{S}_t)_j$ for $(j = 1, \ldots J)$.

Clearly, since the right side of $(\hat{S}_t)$ in the middle part of (36) has two terms, we have three choices for rotation as follows. We can make J versions of $\hat{u}_t^o, (x_t^o)$, or both. Let us choose the first alternative for simplicity, resample $\hat{u}_t^o$, and denote the J versions of regression residuals by $(\hat{u}^o)_j$, leading to $(\hat{S}_t)_j = x_t^o(\hat{u}^o)_j$. After dividing by the common scale factor and adding over t we have our J scalar pivots, denoted upon inserting a second subscript j as: $z_{F,j}$ for $j = 1, 2, \ldots 999$.

Note that $z_{F,j}$ for j=1,2, .., 999 are sample realizations from an asymptotic N(0,1) population. We expect them to be different from N(0,1) in finite samples and the main appeal of the bootstrap here is that we are able to use actual realizations to get robust estimates of the confidence interval for z_F, instead of using the usual ± 1.96 for a 95% interval of parametric normal density. Denote the order statistics of $z_{F,j}$ after reordering them from the smallest to the largest, by $z_{(j)}$ for j=1,2, .., 999. Note that 95% of $z_{F,j}$ are inside the range $(z_{(25)}, z_{(975)})$, irrespective of the sign of $(z_{(25)}, z_{(975)})$, by construction.

Now two preliminary confidence limits on β_1 are given by directly solving the two equations $z_F = z_{(25)}$ and $z_F = z_{(975)}$ as:

$$\begin{aligned} b^{(1)} &= [\textstyle\sum_{t=1}^{T} x_t^o(y_t^o) - z_{(25)}\hat{S}_c]/\textstyle\sum_{t=1}^{T}(x_t^o)^2; \\ b^{(2)} &= [\textstyle\sum_{t=1}^{T} x_t^o(y_t^o) - z_{(975)}\hat{S}_c]/\textstyle\sum_{t=1}^{T}(x_t^o)^2. \end{aligned} \tag{37}$$

Finally, the 95% confidence intervals we seek are:

$$\beta_{1.LO} = \min(b^{(1)}, b^{(2)}), \quad \text{and} \quad \beta_{1.UP} = \max(b^{(1)}, b^{(2)}). \tag{38}$$

We reject the null hypothesis: $\beta_1 = c$ (a constant), unless the hypothesized constant c is inside the estimated interval $(\beta_{1.LO}, \beta_{1.UP})$ based on (38).

For illustration, we return to Irving Fisher's model (1), which contains inflationary expectations. Let us model them explicitly by fitting an AR(1) model to price data and estimate (2) giving $\hat{\mu}_1$ =0.9598, which is unbiased. The strict form of Fisher's null hypothesis: $\beta_1 = 1$ means the expected value of the coefficient of lagged price in our regression should be 0.9598. The OLS estimate at 2.3021 is considerably larger than Fisher's null (=0.9598), with the usual t statistic of 7.009. Since Fisher's null is rejected, lying outside the 95% confidence interval of (1.653, 2.951) based on the t statistic. Our robust GPF bootstrap confidence interval using (38) before any HAE correction is narrower: (2.170, 2.480) but still continues to reject.

Irving Fisher's example has significant autocorrelation due to near zero p-values of the Breusch-Godfrey statistic. Table 4 reports the results for all 20 models similar to Tables 1 and 3. Table 4 suggests heteroscedasticity from low p-values of the Cook-Weisberg statistic. However, since the p-values are all too close to zero in columns 4 for all comparable tables so far (Tables 1, 3 and 4), it would be interesting to compare the power of this statistic with the Breusch-Pagan and other tests in a Monte Carlo experiment, beyond the scope of this paper. Similar to Table 3, all (heuristic) p-values of column 3 of Table 4 exceed 0.05 suggesting no GARCH effects, unlike Table 1.

Figure 3 is a plot of heteroscedasticity correction revealing the non-constant squared residuals (solid line) and how the simplest straight line fit with time τ as the regressor (dashed line) fits them reasonably well (after they are sorted). Of course, the quartic with HC0 has the best fit according to column 5 of Table 4 and as depicted in Figure 4. Some details are omitted for brevity.

After HAE estimation (correcting for first order autocorrelation and heteroscedasticity, in that order) and then using (38) the relevant HC0 linear confidence interval is (1.673399 1.810466). A similar GPF interval for HC0 and quartic polynomial is (1.592199 1.799848). Since Fisher's null remains outside all these intervals, we must reject the strong form of Fisher's hypothesis. This is consistent with Sun and Phillips (2004). However, since the number zero is also outside all the estimated confidence intervals, our results support the weak form of Fisher's null similar to our conclusion in Section 2 based on Dhrymes' CCT statistic.

7 Summary and Final Remarks

This paper suggests some practical solutions to problems including: autocorrelated and or heteroscedastic errors, endogeneity, identification, spurious regression and incidental parameters. Section 1.1 shows how the identification of deep parameters becomes difficult in attempting to solve endogeneity and/ or spurious regression. Our illustration uses Vinod's (2004, 2006) new maximum entropy bootstrap, freely available as the 'meboot' package in the R software. It can help focus on economic specifications of equilibria while avoiding differencing and detrending. Remark 1 describes the unrealism of random walks due to their infinite memory, without discounting.

Feasible GLS needs estimates of the large $T \times T$ covariance matrix of errors, $\Omega(\phi)$, as a parsimonious function of ϕ. Remark 2 states a parsimony rule of thumb that number of coefficients in ϕ should be less than $(T/5) - p$. Hence, parsimony here becomes an upper limit on $dim(\phi)$. Since HAC estimates of standard errors, $SE(b)$, typically use all T diagonals of estimated Ω and several off diagonals, they suffer from the 'incidental parameters' problem. We propose heteroscedasticity cum heterogeneity and autocorrelation efficient (HAE) estimates of β by first fitting ARMA$(q, q-1)$ to residuals for autocorrelations. If squared residuals $(\hat{u})^2$ of a transformed model are nonconstant, we suggest making them monotonic by sorting, and finding their fitted values by regressing $\log(\hat{u})^2$ on a powers of $(1, 2, \ldots)$. A (t-dom $\leftrightarrow$ v-dom) map involving sorting recovers the original time subscript for $(\hat{u})^2$, fitted with very few additional parameters in ϕ and results in a new, practical and parsimonious correction for heteroscedasticity.

We discuss how mappings between the time domain and the new (ordered numerical values) v-domain are analogous to mappings in spectral analysis between time and frequency domains (through Fourier integrals). Assuming smoothness of heteroscedasticity in the v-domain, $f(\Omega_{(tt)})$, and further assumptions, a theorem proves consistency and asymptotic normality, efficiency and specification-robustness of our HAE estimator. The intuition behind the smoothness assumption is that conditional scaling of $E(y|X)$ is fundamentally related to the numerical magnitudes (of y and X) in the v-domain, and need not depend on their timing. A simulation experiment uses a published design where OLS is found hard to beat in small samples. We still use small samples ($T = 24, 46$) and report efficiency gains over OLS achieved by our new HAE estimators. The simulation supports the econometric practice of formally testing for heteroscedasticity before considering any corrections.

In the presence of even mild identification problems, inverting Wald-type tests based on Fisher's pivot into confidence intervals leads to wide intervals with poor coverage, Dufour (1997). Furthermore, a lack of invariance of Wald-type test statistics under nonlinear transformations of parameters has been noted in the literature. Proposition 1 in Vinod (1998) proved that Godambe pivot functions (GPFs) are a robust alternative to Fisher's pivot. Unfortunately, my 1998 implementation of GPF for regressions was somewhat impractical. This paper provides a practical implementation of a scalar version of the GPF for $p > 1$ dimensional β by rotating over regressor columns in X, while using the Frisch-Waugh theorem. Two examples demonstrate that proposals included here are novel, practical and worthy of further study and extension.

References

Carroll, R. J. and D. Ruppert. 1982. A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. Journal of the American Statistical Association 77, 878-882.

Cook, D. and S. Weisberg. 1983. Diagnostics for heteroscedasticity in regression. Biometrika 70, 1-10

Cragg, J. G., 1983. More efficient estimation in the presence of heteroscedasticity of unknown form. Econometrica 51(3), 751-763.

Cribari-Neto F., 2004. Asymptotic Inference Under Heteroskedasticity of Unknown Form. Computational Statistics & Data Analysis, 45, 215-233.

Davidson, R. and J. G. MacKinnon, 2004. Econometric Theory and Methods. Oxford Univ. Press, New York.

Dhrymes, P. J., 1998. Time Series, Unit Roots, and Cointegration. Academic Press, New York

Dixon, S. L and J. W. McKean, 1996. Rank-based analysis of the heteroscedastic linear model. Journal of the American Statistical Association 91, 699-712.

Dufour, Jean-Marie, 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65, 1365-1387.

Durbin, J., 1960. Estimation of parameters in time-series regression models. Journal of the Royal Statistical Society, Ser. B, 22, 139-153.

Engle, R.F. and C.W.J. Granger, 1987. Cointegration and error correction: Representation, estimation and testing. Econometrica 55, 251-276.

Godambe V. P., 1960. An optimum property of regular maximum likelihood estimation. Annals of Mathematical Statistics 31, 1208-1212.

Godambe, V. P., 1985. The foundations of finite sample estimation in stochastic processes. Biometrika 72, 419-428.

Godfrey, L.G., 2006. Tests for Regression Models with Heteroskedasticity of Unknown Form. Computational Statistics & Data Analysis, 50, 2715-2733.

Greene, W. H., 2000, Econometric Analysis. 4-th. ed. New York: Prentice Hall

Heyde, C. C.,1997. Quasi-Likelihood and Its Applications. Springer Verlag, New York.

Kendall, M. G and A. Stuart, 1979. The Advanced Theory of Statistics. 4th ed. Vol. 2. MacMillan Publishing, New York.

Long, J. S. and L. H. Ervin, 2000. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. The American Statistician, 54, 217-224.

Maddala, G. S and In-Moo Kim, 1998. Unit roots, Cointegration and Structural Change. Cambridge Univ. Press, New York.

McCullough B. D. and H. D. Vinod, 2003.Verifying the Solution from a Nonlinear Solver: A Case Study. American Economic Review 93(3), 873-892.

McLeish, D. L., 1974. Dependent central limit theorems and invariance principles. Annals of Probability 2(4), 620-628.

Murray, M.P., 2006. Econometrics. A modern introduction. Pearson Addison Wesley, New York.

Newey. W. K., 1993. Efficient estimation of models with conditional moment restrictions. Ch. 16 in Maddala, G. S., C.R. Rao and H. D. Vinod (eds.)

Handbook of Statistics: Econometrics, Vol.11, New York: North Holland, Elsevier Science Publishers.

Noriega, A. E. and D. Ventosa-Santaulria, 2006. Spurious Regression Under Broken- Trend Stationarity, Journal of Time Series Analysis 27 (5), 671-684.

Phillips P. C. B., 2005. Econometric Analysis of Fisher's Equation. American Journal of Economics and Sociology, 64(1), 125-168.

Priestley, M.B. l98l. Spectral Analysis and Time Series, Vols. I and II, Academic Press, London.

R Development Core Team, 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07- 0, URL http://www.R-project.org.

Rao, P.S.R.S. 1977. Theory of the MINQUE- A Review. Sankhya 39, 201-210.

Robinson, P. M., 1987. Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. Econometrica 55(4), 875-891.

Spanos, A., 1999. Probability theory and statistical inference: econometric modeling with observational data. Cambridge University Press, New York.

Sun, Y. and P. C. B. Phillips, 2004. Understanding the Fisher Equation. Journal of Applied Econometrics 19, 869-886.

Vinod, H.D., 1985. Exact Maximum Likelihood Regression Estimation with ARMA (n,n- l) Errors. Economics Letters 17, 355-358.

Vinod, H.D. 1997. Using Godambe-Durbin Estimating Functions in Econometrics, in I. Basawa, V. P. Godambe and R. L. Taylor, (eds.) Selected proceedings of the symposium on estimating functions, IMS Lecture Notes-Monograph Series, Vol. 32, 215-237.

Vinod, H.D., 1976. Effects of ARMA Errors on the Significance Tests for Regression Coefficients, Journal of the American Statistical Association, 71, 929-933.

Vinod, H.D., 1998. Foundations of statistical inference based on numerical roots of robust pivot functions (Fellow's Corner). Journal of Econometrics, 86, 387-396.

Vinod, H. D., 2000. Foundations of Multivariate Inference Using Modern Computers, North Holland Elsevier Journal Linear Algebra and Its Applications. Vol. 321, 365-385.

Vinod, H. D., 2002. Econometric Applications of Generalized Estimating Equations for Panel Data and Extensions to Inference, in Ullah, A., T. K. Wan and A. Chaturvedi (Ed.s) Handbook of Applied Econometrics Marcel Dekker, New York, Ch. 26, 553-574.

Vinod, H. D., 2004. Ranking Mutual Funds Using Unconventional Utility Theory and Stochastic Dominance. Journal of Empirical Finance Vol. 11(3), 353-377.

Vinod, H. D., 2006. Maximum entropy ensembles for time series inference in economics. Journal of Asian Economics, 17(6), 955-978.

White H., 1980. A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. Econometrica, 48, 817-838.

Wedderburn, R. W. M., 1974. Quasi-likelihood functions, generalized linear models and the Gaussian method. Biometrika 61, 439-447.

Zeileis, Achim, 2007. Econometric Computing with HC and HAC Covariance Matrix Estimators. A vignette accompanying the package 'sandwich' of R

Figure 1: Consumption Function using Logs of consumption and income, HC0 & Linear Case.

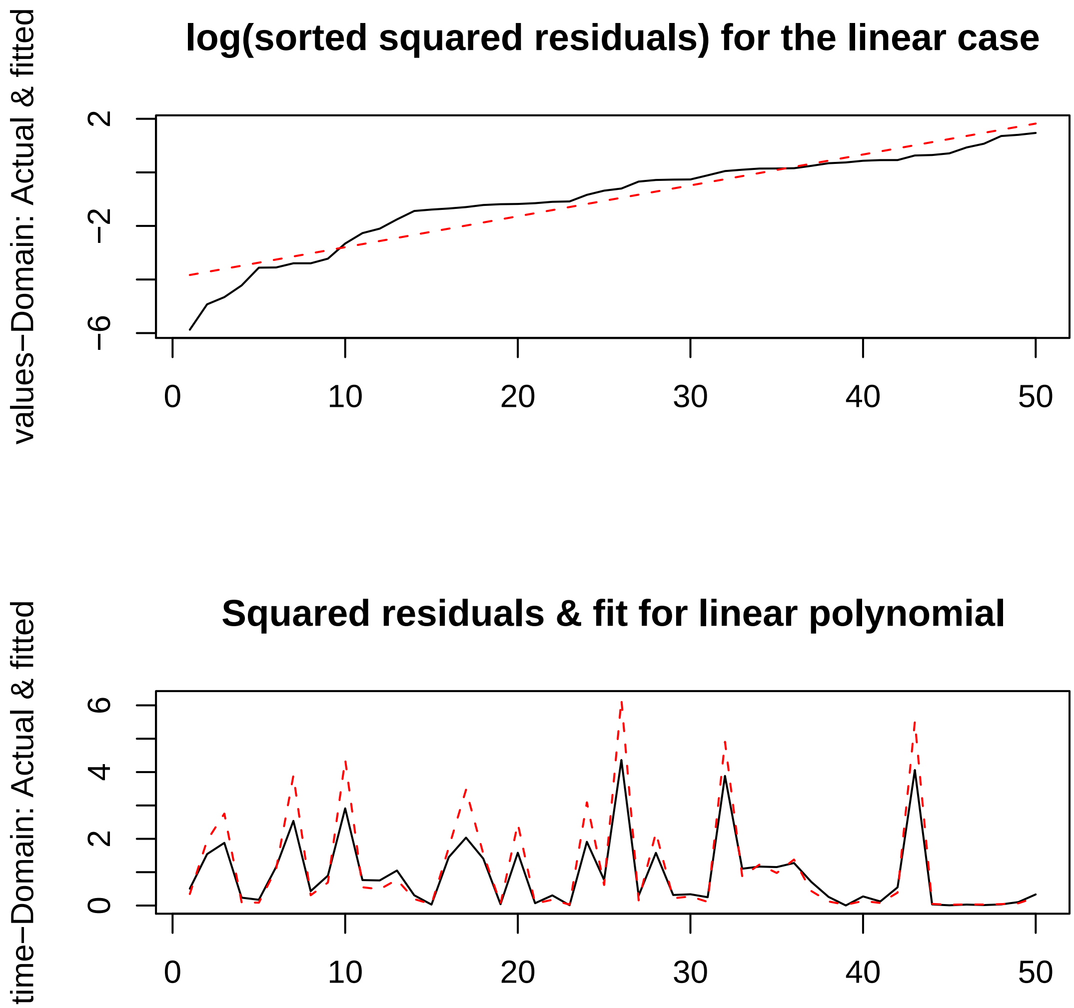

Figure 2: Consumption Function using logs of consumption and income, HC3 & Quartic polynomial.

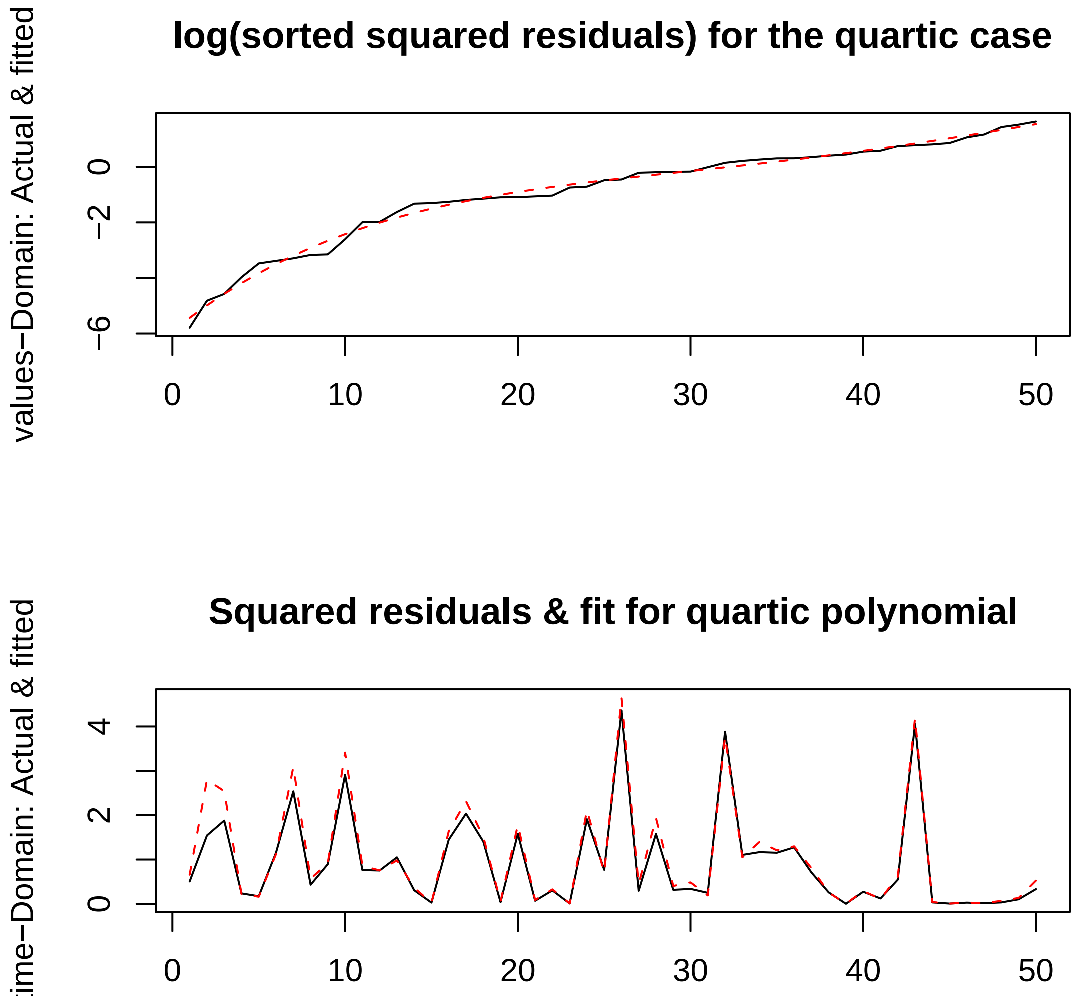

Figure 3: Heteroscedasticity correction for Irving Fisher's model (HC0, Linear)

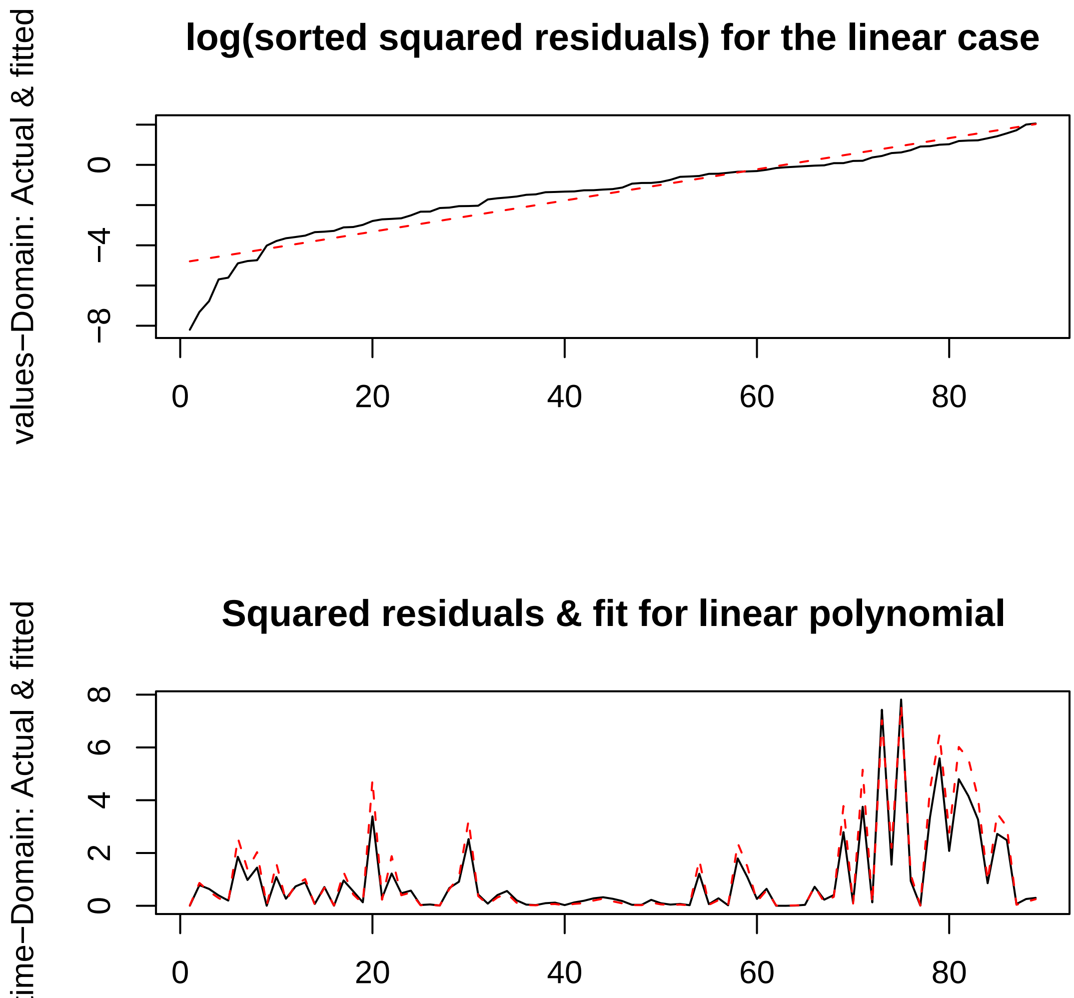

Figure 4: Heteroscedasticity correction for Irving Fisher's model (HC0, Quartic)

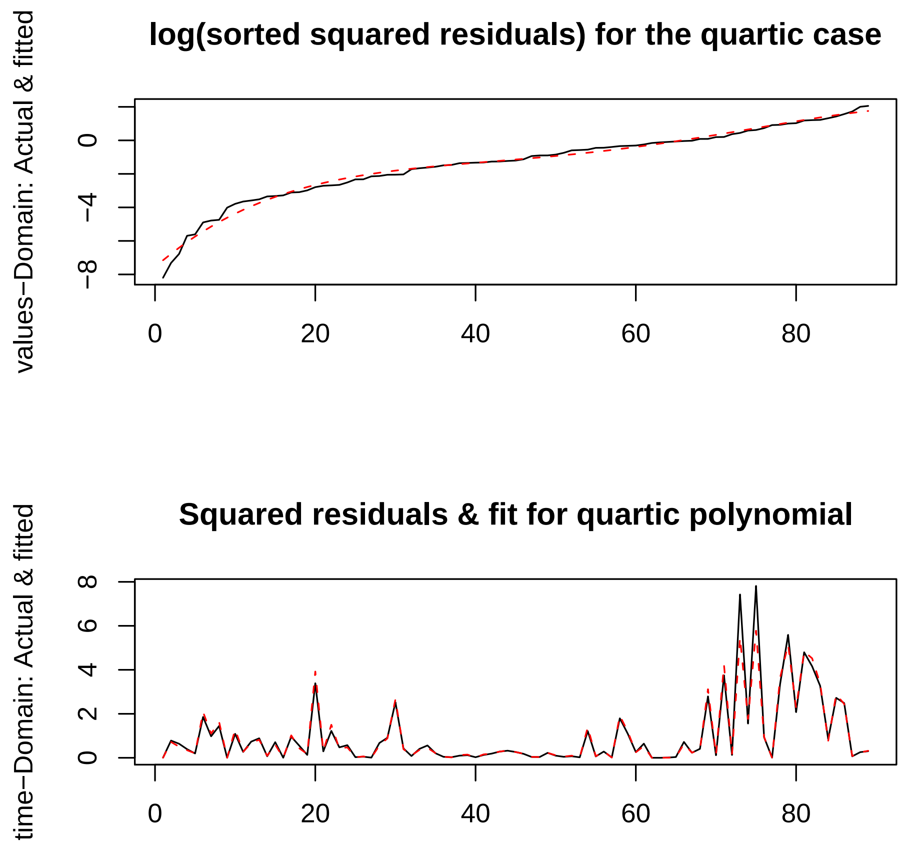

Figure 5: SC1 to SC5 scedasticities built from Student's t (df=5) with lines marked 1 to 5, Results of 999 experiments

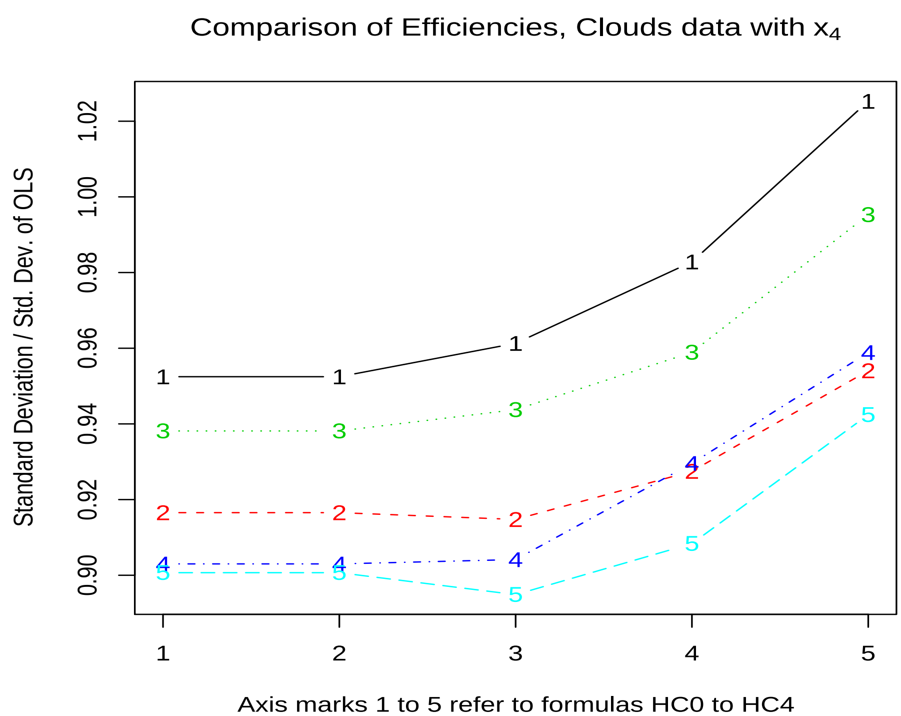

Figure 6: SC1 to SC5 scedasticities built from Student's t (df=5) with lines marked 1 to 5, Results of 999 experiments

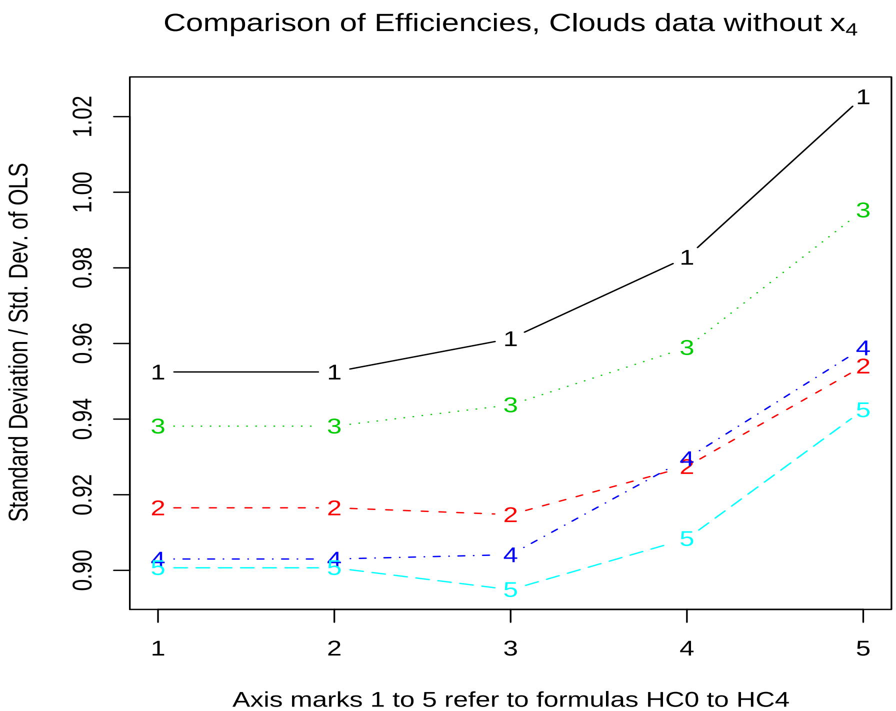

Table 1: Consumption Function (logs) Heteroscedasticity Tests and Efficient Estimation

Transform of $\hat{u}^2$	Order of τ^h polynomial	p-value GARCH	p-value Cook Weisberg	Adjusted R^2	Coef Y_{t-1} after V_h	t- stat $\hat{\beta}_3$
1	2	3	4	5	6	7
HC0 & HC1*	linear	0.017035	1.78E-07	0.898769	0.113114	2.740047
HC2	linear	0.017667	1.80E-07	0.899142	0.119206	2.881599
HC3	linear	0.015957	2.17E-07	0.899629	0.126247	3.021648
HC4	linear	0.016931	3.26E-07	0.900355	0.127205	3.045798
HC0 & HC1	quadratic	0.017035	7.97E-06	0.962918	0.126987	3.717827
HC2	quadratic	0.017667	7.92E-06	0.963051	0.136624	3.991833
HC3	quadratic	0.015957	8.24E-06	0.962909	0.142172	4.119112
HC4	quadratic	0.016931	8.48E-06	0.962857	0.142589	4.134217
HC0 & HC1	cubic	0.017035	2.15E-06	0.989095	0.133066	4.279686
HC2	cubic	0.017667	2.17E-06	0.989188	0.145817	4.6842
HC3	cubic	0.015957	2.54E-06	0.989195	0.15069	4.821263
HC4	cubic	0.016931	3.89E-06	0.989822	0.151141	4.844722
HC0 & HC1	quartic	0.017035	7.30E-06	0.989817	0.133713	4.31633
HC2	quartic	0.017667	7.48E-06	0.990041	0.146747	4.731231
HC3*	quartic	0.015957	8.77E-06	0.990152	0.151421	4.863972
HC4	quartic	0.016931	1.23E-05	0.990951	0.1519	4.890656

Table 2: Heteroscedasticity and autocorrelation efficient (HAE) feasible GLS Estimates of Coefficients of Consumption Function. HC3 weights and quartic polynomial.

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.13513	0.18261	0.740	0.463
C_{t-1}	0.85103	0.03216	26.466	< 2e-16 ***
Y_{t-1}	0.15142	0.03113	4.864	1.33e-05 ***
Confidence intervals	are given	below		
limits →	2.5 %	97.5 %		
(Intercept)	-0.23223443	0.5024996		
C_{t-1}	0.78634028	0.9157164		
Y_{t-1}	0.08879294	0.2140481		
GPF interval				
C_{t-1}	0.8382396	0.8662133		
Y_{t-1}	0.1365905	0.1640065		

Notes: Residual standard error: 1.002 on 47 degrees of freedom, F(2,47): 5.098e+06, p-value: < 2.2e-16.

Table 3: Consumption Function (levels of consumption and income) Heteroscedasticity Tests and Efficient Estimation

Transform of $\hat{u}^2$	Order of τ^h polynomial	p-value GARCH	p-value Cook Weisberg	Adjusted R^2	Coef Y_{t-1} after V_h	t-stat $\hat{\beta}_3$
1	2	3	4	5	6	7
HC0+HC1	linear	0.681493	2.00E-12	0.750405	-0.01265	-0.24757
HC2	linear	0.720334	2.00E-12	0.749499	-0.00685	-0.13298
HC3	linear	0.727052	2.00E-12	0.748637	-0.00261	-0.05013
HC4	linear	0.834826	2.00E-12	0.747702	0.000515	0.009834
HC0+HC1	quadratic	0.681493	1.45E-06	0.816421	0.004893	0.103459
HC2	quadratic	0.720334	1.59E-06	0.81611	0.012737	0.267426
HC3	quadratic	0.727052	1.77E-06	0.815715	0.02009	0.415827
HC4	quadratic	0.834826	2.46E-06	0.816499	0.021627	0.446801
HC0+HC1	cubic	0.681493	7.00E-12	0.928736	0.041506	0.886886
HC2	cubic	0.720334	7.00E-12	0.928604	0.048026	1.023175
HC3	cubic	0.727052	7.00E-12	0.928071	0.058709	1.233592
HC4	cubic	0.834826	7.00E-12	0.927259	0.061022	1.281598
HC0+HC1	quartic	0.681493	6.20E-11	0.944189	0.055225	1.139573
HC2	quartic	0.720334	6.50E-11	0.944342	0.060819	1.250255
HC3	quartic	0.727052	7.10E-11	0.944218	0.07102	1.442987
HC4	quartic	0.834826	1.01E-10	0.944388	0.076907	1.55493

Table 4: Heteroscedasticity Tests and Efficient Estimation of Irving Fisher's Interest Model

Transform of $\hat{u}^2$	Order of τ^h polynomial	p-value GARCH	p-value Cook Weisberg	Adjusted R^2	Coef Y_{t-1} after V_h	t-stat $\hat{\beta}_3$
1	2	3	4	5	6	7
HC0 & HC1	linear	0.091043	0	0.904123	1.71222	17.19788
HC2	linear	0.089014	0	0.905313	1.711027	17.18616
HC3	linear	0.080592	0	0.906315	1.705157	17.115
HC4	linear	0.082642	0	0.906847	1.670015	16.23623
HC0 & HC1	quadratic	0.091043	0	0.938854	1.720717	22.15223
HC2	quadratic	0.089014	0	0.939214	1.720425	22.09011
HC3	quadratic	0.080592	0	0.939367	1.717629	22.00533
HC4	quadratic	0.082642	0	0.938702	1.666006	20.36984
HC0 & HC1	cubic	0.091043	0	0.980066	1.696319	26.98915
HC2	cubic	0.089014	0	0.979928	1.696755	26.89194
HC3	cubic	0.080592	0	0.979593	1.696202	26.79802
HC4	cubic	0.082642	0	0.978245	1.617693	24.35959
HC0 & HC1	quartic	0.091043	0	0.987116	1.666412	28.2777
HC2	quartic	0.089014	0	0.986846	1.667331	28.1693
HC3	quartic	0.080592	0	0.986363	1.667648	28.06705
HC4	quartic	0.082642	0	0.984885	1.582449	25.71994