MISSPECIFICATION TESTING IN SYSTEMS OF EQUATIONS

by

John Campbell Robertson

Dissertation submitted to the Faculty of the
Virginia Polytechnic and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Agricultural Economics

APPROVED:

A. Spanos, Chairman

A. McGuirk

P. Driscoll

R. Foutz

J. Alwang

June, 1992

Blacksburg, Virginia

MISSPECIFICATION TESTING IN SYSTEMS OF EQUATIONS

by

John Campbell Robertson

Committee Chairman: Aris Spanos
Economics

(ABSTRACT)

This dissertation is a set of related papers on the application of the "principle of statistical adequacy" to single and multi-equation econometric models.

The first chapter lays out the intended scope of the dissertation and defines the principle of statistical adequacy.

The second chapter reviews the formulation of tests of statistical adequacy for multivariate models, and describes the implementation of these tests. The first approach that is discussed is to select particular functions of the variables involved that should be orthogonal under the null hypothesis of no misspecification, and the sample analog of these functions is used as a basis for constructing misspecification tests. As an extension of this idea, it is argued that viewing the model in explicit probabilistic terms provides a basis for developing a set of orthogonality conditions that can be tested in terms of all the probabilistic assumptions underlying the model. The formulation of misspecification tests via auxiliary regressions using general polynomial functions and the implementation of these tests via a menu-driven econometric modeling computer program is described.

The third chapter reports the results of an empirical application of the principle of statistical adequacy to the modeling of inflation/unemployment trade-offs for the U.S. Using a statistically adequate "reduced-form" as the basis, a number of competing theoretical models are

considered. The use of graphical techniques and formal misspecification tests in determining the adequacy of the statistical model are emphasized. It is found that none of the competing theoretical explanations of aggregate labor market behavior are acceptable in terms of the over-identifying restrictions imposed or their own statistical adequacy.

The final chapter is an example of how one might proceed when a specification failures the criteria of statistical adequacy. For U.S. interest rates, it is shown that linear-homoskedastic autoregessions do not adequately account for the leptokurtosis and non-linear temporal dependence in the data. Using the evidence provided by preliminary data analysis as a guide, the Student's t autoregressive model with dynamic heteroskedasticity is estimated for the log differences in three interest rate series. The estimation and misspecification testing results suggest that the STAR model adequately accounts for the probabilistic features of the data: bell-shape symmetry; leptokurtosis; first and second-order temporal dependence. In contrast, a number of other heteroskedastic specifications are estimated, and found to be statistically inadequate.

Acknowledgements

My deepest gratitude goes to Professor Aris Spanos for his guidance, friendship, and coffee, and for providing the econometric modeling framework that underlies the research reported in this dissertation. Thanks are also due to the other members of my committee: Drs McGuirk; Alwang; Driscoll; and Foutz. Their constructive criticisms, and numerous suggestions have improved the exposition substantially. Special thanks to Professor McGuirk for her continual support and encouragement over the last four years. The financial assistance of the Reserve Bank of New Zealand, the Ministry of Agricultural Fisheries and the Department of Agricultural Economics at Virginia Tech, is also gratefully acknowledged.

Many thanks to Bert Ward for encouraging my desire to learn, and to my fellow students, past and present, for being good friends and colleagues.

Finally, to my families in New Zealand and in the United States, your love and support has meant so much to me. I dedicate this dissertation to my loving wife Debra for putting up with this kiwi econometrician.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction and Scope

The purpose of this chapter is to briefly outline the objectives of the chapters that follow, placing them in the context of a general framework for econometric modeling.

The central theme of this dissertation is the "principle of statistical adequacy". That is, the ability of a model to adequately account for and summarize the probabilistic information contained in the data. Statistical adequacy is important because the parameters of a theoretical economic model can often be derived via reparameterization/restriction on the parameters of some type of statistical model such as a reduced-form (Hsiao (1983)), or an autoregressive distributed lag model (Hendry et.al., (1988)). The reliability of the results from structural parameter estimation, hypothesis testing, and forecasting/simulation then depends crucially on the adequacy of the statistical parameterization for the data in hand.

The starting point for the approach adopted here is the observation that a statistical model can be either defined directly by a set of assumptions on the conditional distribution of the dependent variable(s) or it can be derived through assumptions on the joint distribution of all the random variables involved in the model. This duality in model specification provides the basis for the Haavelmo reduction approach to statistical model specification, misspecification testing and respecification (Spanos (1990)).

The emphasis in the following chapters is on the practical aspects of this framework; particularly the formulation and implementation of informal graphical techniques and formal misspecification tests of statistical adequacy for multi-equation models.

The residuals from the statistical model are viewed as a mean-deviation realization from the conditional distribution defining the model and thus can be used as a basis for examining the statistical adequacy of the model. The assumptions defining the model can be tested by formulating alternatives that contain the null as a special case but are general enough to capture departures from the null in a wide range of circumstances. In this context, polynomial functions provide convenient forms that can be used in auxiliary regression-based tests of the orthogonality conditions implied by the model.

The formulation and implementation of misspecification tests for multi-equation models is described in Chapter 2. The emphasis on multi-equation regression tests is important for a number of reasons. First, multi-equation regression models are closely related to the simultaneous equation model, forming the basis for estimation, testing theoretical restrictions, and forecasting. Second, to account for the fact that the model is treated as a system, a test must accommodate the cross-equation effects of misspecification on the system as a whole. Third, while the importance of a multivariate perspective is often stressed in the literature (see Hendry et.al., (1988)), there is no econometric software that provides a comprehensive set of multivariate graphical methods and multivariate misspecification testing procedures. Consequently, this Chapter also emphasizes the practical aspects of implementing these techniques via SAM, an interactive menu-driven computer program. Finally, many of the multivariate tests can be formulated by simple auxiliary regressions, and hence are natural extensions of the "augmented conditional moment" and "omitted variables" approaches of Pagan (1984), and Spanos (1986).

The importance of the principle of statistical adequacy in application is demonstrated in Chapter 3. The implicit reduced-form for a number of simultaneous equation models of the labor market is viewed in terms of a general dynamic multivariate statistical model, whose form is chosen both by the data, in terms of the model's statistical adequacy, and as a convenient basis for considering the restrictions imposed by the various structural models.

It is found that the 1954-1991 quarterly data do not admit a linear-homoskedastic statistical model for wages, prices and unemployment. This result is due primarily to the presence of non-linear dependence and skewness in the unemployment series, and the effects of a number atypical events, such as the 1971-1974 wage-price controls. As an alternative, a multivariate dynamic linear regression model for wages and prices, incorporating a series of dummy variables as part of the constant, and using unemployment, productivity, and interest rates as additional conditioning variables are estimated. This model is found to be a statistically adequate representation based on a wide range of informal and formal diagnostic measures. Using this as a general reduced-form, several structural models are estimated, and the over-identifying restrictions imposed on the statistical model were tested. It is shown that all of the conventional models rejected these restrictions, and were statistically inadequate.

The final chapter presents an example of how one might proceed when a specification failures the criteria of statistical adequacy. For U.S. interest rates, it is shown that linear-homoskedastic autoregessions do not adequately account for the leptokurtosis and non-linear temporal dependence in the data. Using the evidence provided by preliminary data analysis as a guide, the Student's t autoregressive model with dynamic heteroskedasticity is estimated for the log differences in three interest rate series. In the STAR model, the conditional mean has a linear autoregressive form and the conditional variance is a quadratic recursive function of all the past history of the series. The estimation and misspecification testing results suggests that the STAR model adequately accounts for the probabilistic features of the data: bell-shape symmetry; leptokurtosis; first and second-order temporal dependence. In contrast, a number of other heteroskedastic specifications are estimated, and found to be statistically inadequate.

Chapter 2

Orthogonality Conditions and Multivariate Misspecification Tests

2.1 Introduction

Consider the regression equation,

$$y_t = \beta' \mathbf{x}_t + u_t, \quad t \in \mathbf{N}, \tag{2.1}$$

β is a kx1 parameter vector. Pagan and Vella (1988, p.31) summarize the assumptions on the error term of this model as:

(a) $E(\mathbf{w}_t u_t) = \mathbf{0}$; i.e., the px1 vector $\mathbf{w}_t$ is not incorrectly excluded from the regression;

(b) $E(\mathbf{w}_t(u_t^2 - \sigma^2)) = \mathbf{0}$; i.e., the errors are assumed to have constant variance, σ^2, that is unrelated to the $\mathbf{w}_t$'s;

(c) $E(u_t u_{t-j}) = 0$ ($j = 1,\ldots$); i.e., the errors have no serial correlation;

(d) $E(u_t^3) = 0$, $E(u_t^4 - 3\sigma^4) = 0$; i.e., the moments are those of a normally distributed random variable with mean zero and variance σ^2.[1]

Much of the recent literature on testing the adequacy of econometric models such as (2.1) focuses on the use of orthogonality restrictions of the type presented in (a)-(d). This chapter reviews a number of such misspecification testing strategies, describing the formulation and implementation of misspecification tests for multivariate regression models.

The moment testing approach (Pierce (1982), Newey (1985), and Tauchen (1985)) is summarized in Section 2.2. This approach utilizes sample analogues of a set of orthogonality conditions which are postulated to hold under the null. Moment tests are closely related to method of moments estimation (Hansen (1982)) summarized in Appendix 2A, and the auxiliary regression score-form of the Lagrange Multiplier test (Godfrey and Wickens (1981)). There is an extensive literature on these topics, as surveyed in White (1987), Pagan and Wickens (1989), and MacKinnon (1992). This section concludes by raising a number of issues concerning the generality of the moment testing approach in the absence of an explicit probabilistic framework.

Moment tests are appealing, in part because of its simplicity in application when the model is estimated by ML. However, the literature on moment testing provides little guidance as to what constitutes a complete set of orthogonality restrictions to test for a model. Moreover, the questions of how to formulate tests to have power in certain directions, or how to proceed if symptoms of misspecification are detected is not dealt with explicitly. In Section 2.3, I examine an approach to misspecification testing that attempts to address these issues. The Haavelmo reduction modeling framework of Spanos (1986, 1989, 1990, 1992) is used as a basis for developing a complete set of misspecification tests for a model, by explicitly relating misspecification testing to model specification and re-specification.

In this approach, the statistical model is viewed in terms of probabilistic assumptions on all the observable random variables involved. Hence, the relationship between the conditional representation and the unconditional distributions is explicit, and appropriate orthogonality conditions arise naturally.

The formulation of separate, joint, and simultaneous conditional moment misspecification tests are discussed for multivariate regression models are described in Section 2.4. Many of the tests can be formulated via auxiliary regressions, by using polynomial functions as a general method for approximating departures from the null of no misspecification. The implementation of these tests using the interactive econometric modeling computer package SAM is described in Section 2.5, and provides a precursor to the empirical application reported in Chapter 3.

2.2 Moment tests

In this section I describe the formulation of moment tests of model misspecification. Heuristically, a moment test of the specification of the single equation regression model (2.1) can be achieved by introducing a px1 vector of variables, w_t, and checking if they are correlated with some function of the error term, u_t. The moment restriction is then that $E[m_t(u_t,w_t;\theta)] = 0$ under the null, and the moment test is a test of this restriction using the sample analog,

$$\hat{\tau} = T^{-1}\sum_{t=1}^{T} m_t(\hat{u}_t, w_t; \hat{\theta}) \tag{2.2}$$

where $\hat{\theta}$ is an $O(T^{½})$ consistent estimator of the kx1 vector θ under the null.[2] For (2.1) the moment restriction take the form $m_t(u_t,w_t;\theta) = f(u_t;\theta)w_t$, where $f(u_t;\theta)$ are particular functions of the error term, (e.g., u_t, $(u_t^2 - \sigma^2)$, u_t^3, ...,) that have zero expectation under the null.[3]

When the statistical model is correctly specified, $T^{½}\hat{\tau}$ can be shown to be asymptotically normal, and an asymptotic χ^2 statistic can be formed as

$$T^2\hat{\tau}'V_{\hat{\tau}}^{-1}\hat{\tau} \sim \chi^2(p). \tag{2.3}$$

The form of the asymptotic covariance $V_{\hat{\tau}}$ depends on the way that θ is estimated. Expanding $T^{½}\hat{\tau}$ around the true parameter vector θ_0 via a first-order Taylor expansion we have,

$$T^{\frac{1}{2}}\hat{\tau} = T^{-\frac{1}{2}}\sum_{t=1}^{T} m_t(y_t,x_t,w_t;\theta_0) + T^{\frac{1}{2}}B(\hat{\theta} - \theta_0) + o_p(1), \tag{2.4}$$

where $B = T^{-1}[\Sigma_{t=1}^{T}\partial m_t(y_t,x_t,w_t;\theta)/\partial\theta]$.[4] Hansen (1982) shows that if $\hat{\theta}$ is an $O(T^{½})$ consistent estimator, and under certain regularity conditions,

$$T^{½}(\hat{\theta} - \theta_0) + T^{½}Ag(\theta_0) \xrightarrow{p} 0, \tag{2.5}$$

If $\hat{\theta}$ is a MM estimator, Hansen shows that $\mathbf{A} = (\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}^{-1}$, with $\mathbf{F} = E[\partial g(\theta)/\partial\theta]$ evaluated at θ_0, and $\mathbf{V} = E[g(\theta_0)g(\theta_0)']$ is the expected value of the cross-products of the kx1 vector of first-order conditions defining $\hat{\theta}$, $g(\theta) = \Sigma_{t=1}^{T} g_t(\theta)$ (see Appendix 2A).[5] In the case when θ is estimated by Maximum Likelihood (ML), Pagan and Vella (1989) show that $\mathbf{A}$ is minus the inverse of the asymptotic information matrix, $\mathbf{I}(\theta) = \lim_{T\to\infty} T^{-1}\mathbf{I}_T(\theta)$, where,

$$\mathbf{I}_T(\theta) = -E[H(\theta)] = E[s(\theta)s(\theta)'], \tag{2.6}$$

$s(\theta) = \Sigma_{t=1}^{T} s_t(\theta) = \partial \mathrm{Log}L_T(\theta)/\theta$, and $\mathbf{H}(\theta)$ is the Hessian matrix of second-order derivatives of the sample log-likelihood function, $\mathrm{Log}L_T$. Pagan and Vella go on to show that in this case, $T^{1/2}\hat{\tau}$ converges in distribution to a p-dimensional normally distributed random vector, with zero mean and asymptotic covariance $\mathbf{V}_{\hat{\tau}} = [\mathbf{I}_p\text{-}\mathbf{BA}]\mathbf{C}[\mathbf{I}_p\text{-}\mathbf{BA}]'$, where $\mathbf{I}_p$ is a pxp identity matrix, $\mathbf{C}$ is the asymptotic covariance of

$$\mathbf{G} = [T^{-1/2}\Sigma_{t=1}^{T}\mathbf{m}_t(y_t,x_t,w_t;\theta_0) \mid T^{-1/2}\Sigma_{t=1}^{T} s_t(\theta_0)]', \tag{2.7}$$

and $\mathbf{G}$ is assumed to be asymptotically normal.

An auxiliary regression approach to implementing the moment test under ML is developed by Newey (1985).[6] If $\mathbf{m}_t(u_t,w_t;\theta_0)$ and $s_t(\theta_0)$ are independent or martingale difference processes, then $\mathbf{C}$ can be consistently estimated by,

$$\hat{\mathbf{C}} = T^{-1}\begin{bmatrix} \hat{M}'\hat{M} & \hat{M}'\hat{S}_\theta \\ \hat{S}_\theta'\hat{M} & \hat{S}_\theta'\hat{S}_\theta \end{bmatrix}, \tag{2.8}$$

where $\hat{\mathbf{M}}$ is the Txp matrix of sample moment conditions with elements $\mathbf{m}_{ti}(\hat{\theta})$, i = 1,...,p, and $\Sigma_{t=1}^{T}\mathbf{m}_t(\hat{\theta})\mathbf{m}_t(\hat{\theta})' = \hat{\mathbf{M}}'\hat{\mathbf{M}}$. $\hat{\mathbf{S}}_\theta$ is the Txk matrix of scores for the estimated model evaluated at $\hat{\theta}$, with elements $s_{tj}(\hat{\theta})$, $j = 1,..,k$, and $T^{-1}\hat{\mathbf{S}}_\theta'\hat{\mathbf{S}}_\theta$ is a consistent estimator of $\mathbf{I}(\theta)$.[7] Hence, $[\mathbf{I}_p\text{-}\mathbf{BA}]$ can be consistently estimated by $[\mathbf{I}_p - \hat{\mathbf{M}}'\hat{\mathbf{S}}_\theta(\hat{\mathbf{S}}_\theta'\hat{\mathbf{S}}_\theta)^{-1}]$, and $\mathbf{V}_{\hat{\tau}}$ can be estimated by,

$$\hat{\mathbf{V}}_{\hat{\tau}} = T^{-1}[\hat{\mathbf{M}}'\hat{\mathbf{M}} - \hat{\mathbf{M}}'\hat{\mathbf{S}}_\theta(\hat{\mathbf{S}}_\theta'\hat{\mathbf{S}}_\theta)^{-1}\hat{\mathbf{S}}_\theta'\hat{\mathbf{M}}], \tag{2.9}$$

which is the residual covariance from a regression of $\hat{\mathbf{M}}$ on $\hat{\mathbf{S}}_\theta$. The moment test statistic is then,

$$T\hat{\tau}'\hat{\mathbf{V}}_{\hat{\tau}}^{-1}\hat{\tau} \equiv \iota'\hat{\mathbf{M}}[\hat{\mathbf{M}}'\hat{\mathbf{M}} - \hat{\mathbf{M}}'\hat{\mathbf{S}}_\theta(\hat{\mathbf{S}}_\theta'\hat{\mathbf{S}}_\theta)^{-1}\hat{\mathbf{S}}_\theta'\hat{\mathbf{M}}]^{-1}\hat{\mathbf{M}}'\iota, \tag{2.10}$$

where ι is a Tx1 vector of ones. Using results for the inverse of a partitioned matrix (see for example, Dhrymes (1978)), it can be deduced that $T\hat{\tau}'\hat{\mathbf{V}}_{\hat{\tau}}^{-1}\hat{\tau}$ is the explained sum of squares from the auxiliary regression (in sample form)

$$\iota = \hat{\mathbf{S}}_\theta\mathbf{b}_1 + \hat{\mathbf{M}}\mathbf{b}_2 + \nu. \tag{2.11}$$

Since the total sum of squares from such a regression is T, the statistic is equivalent to T minus the residual sum of squares, or T times the un-centered R^2.

This auxiliary regression approach is directly related to the implementation of the score form of the Lagrange Multiplier (LM) test (see Godfrey (1988)). To illustrate, let $\theta = (\beta,\alpha)$, where β is a kx1 vector of parameters, and α is a px1 vector of parameters that are zero under the null, and hence are taken to summarize potential misspecification. The sample Log likelihood function for the alternative is $\text{LogL}_T(\beta,\alpha)$ and the score vector is $s(\beta,\alpha)$.

The LM score formulation tests whether the part of s associated with α, when evaluated under the null $(\hat{\beta},\mathbf{0})$, is sufficiently close to zero. The asymptotic LM test statistic is

$$LM = T^{-1}s(\hat{\beta},\mathbf{0})\mathbf{I}_T(\hat{\beta},\mathbf{0})^{-1}s(\hat{\beta},\mathbf{0})' \sim \chi^2(p). \tag{2.12}$$

The k elements of the score that correspond to β are zero by construction when evaluated at $\hat{\beta}$. The p elements of the score corresponding to α should be close to zero when evaluated at $\alpha=\mathbf{0}$ if the restrictions are valid. Hence, the score test is a moment test where the moment restrictions are that $E[s_i(\beta,\mathbf{0})] = 0$, for $i = k+1,\ldots,k+p$.

An auxiliary regression form of the LM test arises by utilizing the average cross-products of the scores to estimate the information matrix under the null. Let $\hat{S}$ denote the $Tx(k+p)$ matrix of scores for $\theta = (\beta,\alpha)$, evaluated at $(\hat{\beta},\mathbf{0})$, so that the cross-product estimate of the information matrix under the null is $T^{-1}\hat{S}'\hat{S}$. Hence, the LM test statistic can be written as,

$$LM = \iota'\hat{S}(\hat{S}'\hat{S})^{-1}\hat{S}'\iota, \tag{2.13}$$

or, partitioning $\hat{S}$ as $[\hat{S}_\beta : \hat{S}_\alpha]$,

$$LM = \iota'\,\hat{S}_\alpha'[\hat{S}_\alpha'\hat{S}_\alpha - \hat{S}_\alpha'\hat{S}_\beta(\hat{S}_\beta'\hat{S}_\beta)^{-1}\hat{S}_\beta'\hat{S}_\alpha]^{-1}\hat{S}_\alpha'\iota, \tag{2.14}$$

which is the explained sum of squares from the auxiliary regression

$$\iota = \hat{S}_\beta \mathbf{b}_1 + \hat{S}_\alpha \mathbf{b}_2 + \nu, \tag{2.15}$$

and is T minus the residual sum of squares, or equivalently, TR^2. On substituting the Txp matrix of moment restrictions $\hat{\mathbf{M}}$ for $\hat{S}_\alpha$, we obtain the Newey-formulation (2.10).[8]

Tauchen (1985), and Pagan and Vella (1989) have proposed a further simplification of the moment test under ML by noting that under the null, $s(\hat{\theta}) = \iota'S_\theta = \mathbf{0}$. Hence, for the "reverse" auxiliary regression of $\mathbf{M}$ on ι and $\hat{S}_\theta$, since $\iota'\hat{S}_\theta = \mathbf{0}$, the coefficient on ι is $\hat{\tau}$,

$$\hat{\tau} = (\iota'\iota)^{-1}\iota'\hat{\mathbf{M}}, \tag{2.16}$$

and the covariance of $\hat{\tau}$ is $(\iota'\otimes\hat{\Sigma}^{-1})^{-1}$ where $\hat{\Sigma}$ is the estimated residual covariance matrix.

Under the null hypothesis, $\hat{\tau} \xrightarrow{p} 0$, so Σ can be consistently estimated by,

$$\hat{\Sigma} = T^{-1}[\hat{M}'\hat{M} - T\hat{\tau}\hat{\tau}' - \hat{M}'\hat{S}_\theta(\hat{S}_\theta'\hat{S}_\theta)^{-1}\hat{S}_\theta'\hat{M}]$$

$$= T^{-1}[\hat{M}'\hat{M} - \hat{M}'\hat{S}_\theta(\hat{S}_\theta'\hat{S}_\theta)^{-1}\hat{S}_\theta'\hat{M}] + o_p(1) \approx \hat{V}_{\hat{\tau}}, \tag{2.17}$$

(Pagan and Vella (1989, p.34)). Thus, one can regress the px1 vector of sample moment conditions on a constant and the kx1 score vector in a system of equations, and test if the intercept terms are jointly zero using $T\hat{\tau}'\hat{\Sigma}^{-1}\hat{\tau}$. For individual restrictions ($p=1$), the test is simply a t-test on the intercept term. Note that because of the additional approximation used here, the Newey and Tauchen forms only coincide numerically in the scalar case (Tauchen (1985, p.436), MacKinnon (1992, p.133)).

Moment tests based on ML estimates are also closely related to the Information Matrix (IM) test (White (1982)) and the Hausman test (Hausman (1978)). The IM test is based on the idea that if the model is correctly specified, different consistent estimates of the information matrix should coincide asymptotically. The moment conditions that are tested are the differences between the elements of two different information matrix estimates (see also Hall (1987)). Chesher (1983) shows that the IM test can be computed via an auxiliary regression similar to that presented above. The Hausman test compares two estimators of θ, $\hat{\theta}$ and θ^*, that are both consistent and asymptotically normal under the null, and that converge to different limits when the model is misspecified. Rudd (1984), and Newey (1985) show that if $\Sigma_{t=1}^{T} m_t(\hat{\theta}) = 0$ defines $\hat{\theta}$, then there is an equivalence between $T^{1/2}(\hat{\theta} - \theta^*)$ and $T^{-1/2}\Sigma_{t=1}^{T} m_t(\theta^*)$ in the sense that testing $E[m_t] = 0$, with $\hat{\tau}$ based on θ^*, is asymptotically equivalent to the Hausman test.

In terms of finite sample properties, the actual size of the moment test statistic $T\hat{\tau}'\hat{V}_{\hat{\tau}}^{-1}\hat{\tau}$ is often much smaller than its nominal size. Hence, large values of the test statistic cannot safely be interpreted as evidence of misspecification (MacKinnon (1992)). Chesher and Spady (1988), Kennan and Neumann (1988), trace this result to the use of the cross-product estimate of the information matrix. Alternative approaches to estimating **C** are proposed by Newey and West (1987), and Andrews (1991). Also, higher-order terms could be important in the Taylor expansion of $\hat{\tau}$ in (2.4) (see Gregory and Veall (1985)).

The significance levels for each test could be computed by Monte Carlo or Boot-strapping methods (Breusch and Pagan (1979)). However, these are rather costly approaches if numerous tests are computed in assessing the statistical adequacy of a model and more than one model is estimated.

As a simple illustration of the implementation of moment tests, consider the linear regression model (2.1), where $\mathbf{w}_t$ is a px1 vector of "omitted variables". The sample analogues of the moment restrictions are as in (2.2). The Tauchen moment test for each of (a)-(d) is based on estimating a system of p auxiliary regressions,

$$m_i(\hat{u}_t, w_{it}\hat{\theta}) = \beta_{0i} + \beta_i' s_t(\hat{\theta}) + v_{it}, \quad i = 1,\ldots,p, \; t = 1,\ldots.,T, \tag{2.18}$$

by LS, where $s_t(\hat{\theta}) = \partial \log L_t(\theta)/\partial\theta$, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The t-statistics on β_{0i} are the moment test statistics, and an asymptotic chi-square χ^2 test can be obtained by testing that all $\beta_{0i} = 0$ using the system of equations. The Newey (1985) moment test for each restriction is based on the auxiliary regression,

$$1 = \boldsymbol{\beta}' s_t(\hat{\theta}) + \boldsymbol{\alpha}' \mathbf{m}_t(\hat{u}_t, w_t\hat{\theta}) + e_{it}, \tag{2.19}$$

and the t-statistics on each α_i, $i = 1,\ldots,p$ are numerically identical to those on β_{0i} (MacKinnon (1992, p.133)).

From this heuristic discussion we can see that the moment testing strategy is appealing in part because of its simplicity in application when the model is estimated by ML. However, the literature on moment testing provides little guidance as to what constitutes a complete set of testable orthogonality restrictions for a model. Moreover, the questions of how to formulate tests to have power in certain directions, or how to proceed if symptoms of misspecification are detected is not dealt with explicitly. In the next section an approach to misspecification testing that attempts to address these issues is examined. The Haavelmo reduction modeling framework of Spanos (1986, 1989, 1990, 1992) is used as a basis for developing a complete set of misspecification tests for a model by explicitly relating misspecification testing to model specification and re-specification. A number of separate and joint tests that arise as auxiliary regressions based on orthogonality conditions and appropriate choices of the "omitted variables" are described.

2.3 The Omitted Variables Approach to Misspecification Testing

This section describes the formulation of misspecification testing procedures based on the "omitted variables argument" of Spanos (1986). The statistical model is specified in terms of the decomposition of conditional moments and specific probabilistic assumptions. By postulating alternative parametric forms of the conditional moments, based on the same sample information, auxiliary regression-based tests for misspecification can be derived. The "omitted variables" are variables contained in the distribution of the sample, but ignored in specifying the statistical model. This approach has the advantage that tests that have power in certain directions can be obtained by relating the specification of the alternative to the underlying joint distribution, and hence can provide useful information as to possible respecifications of the model.

The rest of this chapter focuses on the interesting case of testing for misspecification in multivariate regression models. These models are important because they are directly related to the simultaneous equation model (see Chapter 3). However, misspecification testing for multivariate models is usually approached on an equation-by-equation basis (see Hendry, Neale and Srba (1988)). A single-equation approach to misspecification testing does not account for cross-equation effects that may be important in determining the statistical adequacy of the system. For example, the Durbin-Watson test for autocorrelation in the first equation of a two-equation system amounts to testing the significance of ρ, the correlation between the residuals u_{1t} and u_{1t-1}. A multivariate tests must also take account of $\rho(u_{1t},u_{2t-1})$, and $\rho(u_{1t-1},u_{2t})$.

Most multivariate statistical models can be viewed as extensions of the Multivariate Linear Regression (MLR) model,

$$\mathbf{y}_t = \mathbf{B}'\mathbf{x}_t + \mathbf{u}_t, \quad \mathbf{u}_t \mid \mathbf{X}_t \sim \mathrm{N}(\mathbf{0},\boldsymbol{\Omega}), \qquad t \in \mathbf{N}, \tag{2.20}$$

where $\mathbf{Z}_t = (\mathbf{y}_t',\mathbf{X}_t')'$, $\mathbf{y}_t$ is an mx1 random vector, $\mathbf{x}_t$ is a kx1 vector of observed realizations on the conditioning random variables $\mathbf{X}_t$ (note that $\mathbf{X}_t$ generally includes 1), $\mathbf{B}$ is the kxm matrix of unknown regression coefficients, and $\boldsymbol{\Omega}$ is the mxm conditional covariance matrix.

The MLR model is a statistical model based on the first two moments of the conditional normal distribution, and is defined by the following probabilistic assumptions:

$E(\mathbf{y}_t \mid \mathbf{X}_t=\mathbf{x}_t) = \mathbf{B}'\mathbf{x}_t$; (2.21a)

$D(\mathbf{y}_t \mid \mathbf{X}_t;\boldsymbol{\theta})$ is Normal, $\boldsymbol{\theta} \equiv (\mathbf{B},\boldsymbol{\Omega})$, $\mathbf{B} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$; (2.21b)

$Cov(\mathbf{y}_t \mid \mathbf{X}_t=\mathbf{x}_t) = \boldsymbol{\Omega}$ is homoskedastic (free of $\mathbf{x}_t$); (2.21c)

$\mathbf{Y} \equiv (\mathbf{y}_1,\ldots,\mathbf{y}_T)'$ is an independent sample sequentially drawn from $D(\mathbf{y}_t \mid \mathbf{X}_t;\boldsymbol{\theta})$; (2.21d)

(Spanos (1986, p.574)).[9] As will become clear, these assumptions are interrelated, and can not be chosen independently. Moreover, orthogonality conditions arise naturally via the statistical decomposition of a square integrable random vector for particular choices of the conditioning information set $\Im_t$, and the distributional assumptions on $\{\mathbf{Z}_t, t \in \mathbb{N}\}$.

Consider first the conditional mean decomposition,

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{u}_t, \tag{2.22}$$

where $\boldsymbol{\mu}_t = E(\mathbf{y}_t \mid \Im_t)$, and $\boldsymbol{\mu}_t \perp \mathbf{u}_t$ (orthogonal) by construction. One way to view the role of statistical modeling in this context is that of ensuring that, for the data in hand, we select $\Im_t$ so as to maximize $\boldsymbol{\mu}_t$ (minimize $\mathbf{u}_t$) in a square integrable sense.[10] For the MLR model it is assumed that the relevant information set is just the current realization of $\mathbf{X}_t$, denoted $\mathbf{X}_t=\mathbf{x}_t$, and the assumption of an independent sample implies that the past history of $\mathbf{y}_t$ is not relevant conditioning information. The conditional mean is assumed to be a linear function of the conditioning variables, and the form of the parameters follows directly from the assumption of linearity (see Spanos (1992a)).

For the conditional covariance we have the second-order decomposition,

$$\mathbf{u}_t\mathbf{u}_t' = E(\mathbf{u}_t\mathbf{u}_t' \mid \Im_t) + \mathbf{e}_t, \tag{2.23}$$

where $E(\mathbf{u}_t\mathbf{u}_t' \mid \Im_t) \perp \mathbf{e}_t$ by construction. This decomposition extends the modeling problem to one of simultaneously considering an appropriate choice of $\Im_t$ for both the first two moments.[11] For the MLR model, $E(\mathbf{u}_t\mathbf{u}_t' \mid \mathbf{X}_t=\mathbf{x}_t) = \mathbf{\Omega}$ is an assumption of homoskedasticity. In particular, the optimal predictor of $\mathbf{y}_t\mathbf{y}_t'$ (in conditional mean deviation form) is free of the conditioning variables. Under normality we need to consider only the first two moments because the higher odd conditional moments are zero, and the even conditional moments are proportional to the conditional variance. Independence implies that $E(\mathbf{u}_t\mathbf{u}_s' \mid \mathbf{X}_t=\mathbf{x}_t) = \mathbf{0}$, $t \neq s$.

While the MLR model is defined in terms of the conditional distribution $D(\mathbf{y}_t \mid \mathbf{X}_t;\boldsymbol{\theta})$, the properties of the model can not be separated from the properties of the joint distribution $D(\mathbf{Z}_t;\boldsymbol{\Psi})$ in two respects. First, we can readily deduce that:

(i) $E_z(u_{it}) = E_x\{E(u_{it} \mid \mathbf{X}_t=\mathbf{x}_t)\} = 0$, $i = 1,\ldots,m$;

(ii) $E_z(u_{it}u_{jt}) = E_x\{E(u_{it}u_{jt} \mid \mathbf{X}_t=\mathbf{x}_t)\} = \omega_{ij}$, $i,j = 1,\ldots,m$;

(iii) $E_z(u_{it}u_{jt}u_{kt}) = E_x\{E(u_{it}u_{jt}u_{kt} \mid \mathbf{X}_t=\mathbf{x}_t)\} = 0$, $i,j,k = 1,\ldots,m$;

(iv) $E_z(u_{it}u_{jt}u_{kt}u_{\ell t}) = E_x\{E(u_{it}u_{jt}u_{kt}u_{\ell t} \mid \mathbf{X}_t=\mathbf{x}_t)\} = \omega_{ij}\omega_{k\ell} + \omega_{ik}\omega_{j\ell} + \omega_{i\ell}\omega_{jk}$, $i,j,k,\ell = 1,..,m$;

(v) $E_z(\mu_{it}u_{it}) = E_x\{E(\mu_{it}u_{it} \mid \mathbf{X}_t=\mathbf{x}_t)\} = 0$, $i = 1,\ldots m$;

where $\omega_{ij} = Cov(u_{it}u_{jt} \mid \mathbf{X}_t=\mathbf{x}_t)$. Hence, an equivalent way to look at the problem of statistical model specification is as an attempt to ensure that $E_z(\mathbf{u}_t\boldsymbol{\mu}_t') = \mathbf{0}$. If $E_z(\mathbf{u}_t\boldsymbol{\mu}_t') \neq \mathbf{0}$ then there is information in $\mathbf{u}_t$ not utilized in the form of $\boldsymbol{\mu}_t$, violating the maximization principle. Second, while the model is specified in terms of the conditional distribution, it can also be obtained as a reduction from the underlying joint distribution of the random variables involved by imposing assumptions on the joint distribution. This joint-to-conditional approach is called the "Haavelmo reduction" (Spanos (1986)), and is particularly useful because it is often easier to deduce the plausibility of assumptions relating to $D(\mathbf{Z}_t;\psi)$ than to $D(y_t \mid \mathbf{X}_t;\theta)$.[12]

To illustrate the "Haavelmo reduction", note that the assumption that $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is Normal, Independent, and Identically Distributed (NIID) is sufficient (and essentially necessary) to derive the linear regression specification (Barra (1981), Spanos (1992a)). This connection can be presented in terms of the sequence of reductions

$$D(Z_1,\ldots,Z_T;\Psi) = \prod_{t=1}^{T} D(Z_t;\phi) = \prod_{t=1}^{T} D(y_t|X_t;\phi_1)D(X_t;\phi_2), \tag{2.24}$$

where the joint distribution is normal. The first equality imposes the independence and identical distribution assumptions, and the second is simply applying Bayes decomposition theorem. Moreover, since $\phi_2 = (\mu_x,\Sigma_{22})$, $\mathbf{X}_t$ is said to be "weakly exogenous" for $\theta = \phi_1$ (Engle, Hendry and Richard (1983)). In particular, the values that ϕ_2 takes do not constrain the possible values of θ and vice versa, so $D(\mathbf{X}_t;\phi_2)$ can be ignored for making inferences on θ.[13] (2.24) is an example of the Haavelmo reduction scheme of Spanos (1986, 1992), who shows that imposing alternative reduction assumptions relating to the:

(a) Distribution;

(b) Homogeneity; and

(c) Memory structure;

of the random variables underlying the data, gives rise to alternative statistical models. The Haavelmo reduction provides a method for examining a model's specification in terms of the conditional distribution definition (2.21a-d), and in terms of the distributional properties of the data via (2.24).

As described in Spanos (1992), the probabilistic information provided by preliminary data analysis can often provide a guide in choosing a conditional representation for the data, via the Haavelmo reduction. Misspecification testing and graphical information based on the estimated residuals are then used to assess the adequacy of the model specification, and also to suggest possible respecifications of the model.

Specifically, misspecification tests attempt to measure the effects of omitting potentially relevant information when choosing the form of the conditional moments of the model. The relevant information is demarcated by the variables contained in the underlying joint (Haavelmo) distribution $D(\mathbf{Z}_1,\ldots,\mathbf{Z}_T;\boldsymbol{\Psi})$. This ensures that the null and the alternative have a common basis for comparison, and that the orthogonality conditions under the null hypothesis are well-defined (Spanos (1986)).

To illustrate the basic rationale of the "omitted variables" approach, consider the construction of a test of conditional mean misspecification for (2.20). Misspecification implies that there is information left in the error term $\mathbf{u}_t$, that is not utilized in the form of $\boldsymbol{\mu}_t$; i.e., $E(\mathbf{u}_t\boldsymbol{\mu}_t') \neq \mathbf{0}$. The non-orthogonality can be tested via a parametric alternative which includes the null as a special case,

$$\mathbf{y}_t = \mathbf{B}_0'\mathbf{x}_t + \boldsymbol{\alpha}_0'\mathbf{w}_t + \mathbf{v}_t, \tag{2.25}$$

where $\mathbf{x}_t$ is a kx1 vector of regressors, and $\mathbf{w}_t$ is a px1 vector of well-defined functions of other variables contained in the joint distribution of the sample, $D(\mathbf{Z}_1,\ldots,\mathbf{Z}_T;\boldsymbol{\Psi})$. This includes, time trends, lagged regressands $\{\mathbf{y}_{t-j}\}_{j=1}^{p}$, contemporaneous and lagged regressors $\{\mathbf{x}_{t-j}\}_{j=0}^{p}$, and non-linear functions of these. The system (2.25) includes (2.20) as a special case under H_0: $\boldsymbol{\alpha}_0 = \mathbf{0}$.

Rearranging (2.25) we may write,

$$\mathbf{u}_t = (\mathbf{B}_0\text{-}\mathbf{B})'\mathbf{x}_t + \alpha_0'\mathbf{w}_t + \mathbf{e}_t \tag{2.26}$$

and the operational form of the alternative is,

$$\hat{\mathbf{u}}_t = (\mathbf{B}_0\text{-}\hat{\mathbf{B}})'\mathbf{x}_t + \alpha_0'\mathbf{w}_t + \mathbf{e}_t \tag{2.27}$$

where $\hat{\mathbf{u}}_t = \mathbf{y}_t - \hat{\mathbf{B}}'\mathbf{x}_t$, and $\hat{\mathbf{B}}$ is the LS estimate of $\mathbf{B}$ in (2.20). Note that the conditioning underlying (2.20) and (2.25) are different so that $\mathbf{B} \neq \mathbf{B}_0$ unless $\alpha_0 = \mathbf{0}$, or $\mathbf{w}_t \perp \mathbf{x}_t$.[14]

Although (2.20) is nested in (2.25), we need not think of α as a parameter vector in the "correct specification". The alternative formulation is chosen primarily by mathematical arguments of approximation (see for example, Rivlin (1969)). In this context, appropriately defined polynomials provide a convenient method of capturing departures from the null. Moreover, when $\alpha \neq \mathbf{0}$, we have an indication that there is a problem with the current specification, but it may not tell us precisely what the problem is. This uncertainty arises because all the assumptions of the statistical model are interrelated, and often a number of assumptions are invalid simultaneously. Hence, we must consider all the test results together, with each individual rejection being viewed as an indication of problems with the specification.

From this perspective, misspecification tests are viewed as pure significance tests; their role is that of determining whether there is evidence against the probabilistic assumptions defining the model, and rejection of the null does not imply endorsement of the alternative (Cox and Hinkley (1974, p.66)). The p-value (the smallest significance level at which the null hypothesis would be rejected) is computed under the assumption of no misspecification and, hence, gives a measure of how strongly the data contradict the hypothesis.

Because the restrictions implied by (2.25) are linear in the parameters the relevant test statistics can be selected from among the Wald (W), Likelihood Ratio (LR), and LM formulations described in Appendix 2B. These statistics can be easily justified on normal quasi-maximum likelihood grounds, where the implied quadratic forms provide a natural basis for measuring the distance between the null and alternative.

When there is evidence of misspecification, a respecification of the model can be suggested by tracing the "symptoms" back to the assumptions of the underlying joint distribution, and postulating an alternative set of reduction assumptions. For example, when there is evidence of autocorrelation, the multivariate dynamic linear regression model arises by replacing independence with some form of asymptotic independence, and identically distributed with stationarity in the Haavelmo reduction presented in (2.24). The relationships among the linearity, homoskedasticity and normality assumptions can also suggest a number of possible respecifications. When normality and homoskedasticity are rejected, a respecification along the lines of a non-normal Elliptical regression model may be appropriate (see Spanos (1991)). When there is also evidence of dynamic heteroskedasticity and/or autocorrelation, a non-normal Elliptical dynamic regression model might be considered (see Chapter 4 for an application). When the joint distribution is not a member of the Elliptic family, it is possible that the conditional mean will be non-linear in $\mathbf{x}_t$.

This discussion suggests that the probabilistic structure of the data can be utilized in formulating misspecification tests that have power in certain directions. A number of such formulations are described in the next Section.[15]

2.4 Misspecification Tests for Multivariate Regression Models

The following discussion of auxiliary regression misspecification tests for the MLR and related statistical models is divided into four sub-sections. The first sub-section describes tests of the assumptions defining the conditional mean (linearity, parameter time-invariance, and independence), separately and jointly. The second sub-section considers separate and joint tests for the assumptions of the conditional covariance (time-invariance, and homoskedasticity). The third sub-section describes tests of normality (and implicitly the nature of the third and fourth moments). The final sub-section discusses a method for testing the assumptions of all the conditional moments simultaneously via a Seemingly Unrelated Regression (SUR) formulation.

To begin, it is useful to define a number of terms that will be used extensively below:

(a) $\boldsymbol{\phi}_t = (t-\frac{1}{2}(T+1),\ t^2-(T+1)t+(T+1)(T+2)/6,\ \ldots)'$, $t = 1,\ldots,T$, is a vector of orthogonal trend polynomials;

(b) $\boldsymbol{\psi}_t = (\psi_{1t},\ldots,\psi_{pt})'$, $\psi_{\ell t} = x_{it}x_{jt}$, $i \geq j$, $i,j = 1,\ldots,k$, $\ell = 1,2,\ldots,p$, $p = \frac{1}{2}k(k-1)$, are the second-order terms in the Kolmogorov-Gabor (KG) polynomial (see Ivakhnenko (1984))

$$h(\mathbf{x}_t) = a + \sum_{i=1}^{k} b_i x_{it} + \sum_{i=1}^{k}\sum_{i\geq j}^{k} c_{ij} x_{it} x_{jt} + \sum_{i=1}^{k}\sum_{j\geq i}^{k}\sum_{l\geq j}^{k} d_{ijl} x_{it} x_{jt} x_{lt} + \ldots \tag{2.28}$$

(note that the constant is excluded from $\mathbf{x}_t$ in this definition);

(c) A multivariate version of the RESET polynomial of Ramsey (1969) is,

$$r(\mathbf{x}_t) = \sum_{i=1}^{m} a_i \mu_{it} + \sum_{i=1}^{m}\sum_{i\geq j}^{m} b_{ij} \mu_{it} \mu_{jt} + \sum_{i=1}^{m}\sum_{j\geq i}^{m}\sum_{l\geq j}^{m} c_{ijl} \mu_{it} \mu_{jt} \mu_{lt} + \ldots \tag{2.29}$$

where $\mu_{it} = \boldsymbol{\beta}_i'\mathbf{x}_t$, $i = 1,\ldots,m$, and $\boldsymbol{\beta}_i$ is the kx1 conditional mean parameter vector from the i^{th} equation, $i = 1,\ldots,m$.[16]

2.4.1 Multivariate Conditional Mean Tests

This section describes the formulation of auxiliary regression tests of the specification of the conditional mean of the multivariate regression model. This includes tests for non-linearity, unmodeled trends, and autocorrelation, as well as a joint test of all these misspecifications.

Non-Linearity

Joint normality ensures that the conditional mean is a linear function of the conditioning variables. Hence, a test for linearity should be viewed as indirect evidence on the validity of normality.[17] Tests for non-linearity can be based on the null hypothesis H_0: $\boldsymbol{\alpha}_0 = \mathbf{0}$, in the auxiliary multivariate regression:

$$\hat{\mathbf{u}}_t = (\mathbf{B}_0 - \hat{\mathbf{B}})'\mathbf{x}_t + \boldsymbol{\alpha}_0'\mathbf{w}_t + \mathbf{e}_t \tag{2.30}$$

where $\mathbf{w}_t$ are the higher-order terms in the KG (2.28) or RESET (2.29) polynomials. These particular types of polynomials are shown to provide good approximations to many of the non-linear functions considered in the literature for single equation models (Ivakhnenko (1984), Godfrey (1988)).

In practice, second and third-order polynomials are usually sufficient, and the additional flexibility provided by the KG form suggests it may be preferable if degrees of freedom permit its use. For numerical reasons it may also be useful to standardize the variables by their sample variances, since the standardization does not affect the values of the estimated statistics. Multicollinearity problems may also necessitate that particular terms be excluded from the test. For example, only the squared terms in the KG polynomial might be used. Collinearity is particularly likely if the data are trending or the regression includes a large number of lagged variables.

One could employ principle component methods to deal with the collinearity problem. For example, let Ψ be the Txq ($q = \frac{1}{2}k(k+1)$) matrix of squared and cross-product terms from a second-order KG polynomial, then we may define the sample principle components $\Psi^* = \Psi\mathbf{P}_x$, where $\mathbf{P}_x$ is a qxq orthogonal matrix whose columns are the eigenvectors of $\Psi'\Psi$ (see Muirhead (1982)). The $r<q$ orthogonal linear combinations in Ψ^* that account for a large proportion of the variation, in the sense that $\lambda/\text{tr}(\Psi'\Psi) \simeq 1$, where λ is the vector of associated eigenvalues, provides a set of regressors that may be used in place of Ψ in the auxiliary regression. This approach has the advantage that the orthogonality of the principal components ensures that adding an extra term to the auxiliary regression does not effect the significance of the original r-components.

The power of linearity tests depends, *inter alia*, on the correlation between the polynomial approximations and the actual "omitted" non-linear terms. The power of these tests can be investigated through Monte Carlo experimentation by generating data from distributions that imply non-linear conditional means (e.g., multivariate Exponential). Traditionally, such studies have considered alternative functional forms while maintaining the assumption of normality (see Godfrey (1988)).

Trends

Time series data often exhibit evidence of trends (first-order non-stationarity). If a trend polynomial has not been included in $\mathbf{x}_t$ as part of the constant, or other variables in the regression do not have similar trends, it is quite likely that the estimated residuals will also have a trending mean. The constant term for a linear conditional mean model is defined as $\mathbf{c} \equiv \mu_y - \mathbf{B}'\mu_x$ (see Spanos (1992a)), where μ_y and μ_x, are the unconditional means. But, if the means of $\mathbf{y}_t$ and $\mathbf{x}_t$ are trending we have the time-dependent constant $\mathbf{c}(t) \equiv \mu_y(t) - \mathbf{B}'\mu_x(t)$. Using terms in ϕ_t as an approximation to $\mathbf{c}(t)$ gives rise to the alternative regression

$$\hat{\mathbf{u}}_t = (\mathbf{B}_0\text{-}\hat{\mathbf{B}})'\mathbf{x}_t + \alpha_0'\boldsymbol{\phi}_t + \mathbf{e}_t \tag{2.31}$$

Parameter instability due to "outliers" or "steps" in the mean of the data (particularly $\mathbf{y}_t$) can be examined by adding specific dummy variables to the vector of trend polynomials $\boldsymbol{\phi}_t$.[18]

Autocorrelation and Misspecified Dynamics

For economic time series data, the MLR model is generally inappropriate because of the temporal structure of such data, making testing for autocorrelation very important. Traditionally, a Vector AutoRegressive and Moving Average (VARMA)-type generating mechanism for the error vector is viewed as the leading cause of observed temporal correlation of the estimated residuals. For example, the first-order vector-autocorrelation mechanism is,

$$\mathbf{u}_t = \boldsymbol{\rho}_1'\mathbf{u}_{t-1} + \mathbf{e}_t, \tag{2.32}$$

where $\mathbf{e}_t \sim \text{NIID}(\mathbf{0},\boldsymbol{\Omega})$ is a vector innovation process, $\boldsymbol{\rho}$ is the mxm first-order cross-correlation matrix, and $\mathbf{u}_t$ is defined by (2.20). However, the assumption of error autocorrelation imposes very strong restrictions on the temporal structure of the data (Hendry and Mizon (1978), Spanos (1987)). In this context, a regression that contains (2.32) as a special case is the multivariate dynamic linear regression,

$$\mathbf{y}_t = \boldsymbol{B}_0'\boldsymbol{x}_t + \sum_{i=1}^{p}(\boldsymbol{A}_i'\boldsymbol{y}_{t-i} + \boldsymbol{B}_i'\boldsymbol{x}_{t-i}) + \boldsymbol{e}_t \tag{2.33}$$

where $\mathbf{e}_t$ is a vector innovation process, and whose the operational form for testing purposes is

$$\hat{\mathbf{u}}_t = (\boldsymbol{B}_0 - \hat{\boldsymbol{B}})'\boldsymbol{x}_t + \sum_{i=1}^{p}(\boldsymbol{A}_i'\boldsymbol{y}_{t-i} + \boldsymbol{B}_i'\boldsymbol{x}_{t-i}) + \boldsymbol{e}_t. \tag{2.34}$$

For the MLR model, the hypothesis of independence is H_0: $\mathbf{A}_i = \mathbf{B}_i = \mathbf{0}$, $i = 1,\ldots,p$. However, in general, this is only a test for linear independence, because non-correlation and independence only coincide under normality. We can see that (2.33) arises by introducing a lag polynomial for $\mathbf{y}_t$ and $\mathbf{X}_t$ into (2.20). This specification can be derived via an alternative reduction of the Haavelmo distribution by replacing the independence assumption with p^{th}-order Markovness in (2.24). Note that for $p=1$, (2.33) and (2.34) coincide when $\mathbf{B}\boldsymbol{\rho}_1 = -\mathbf{B}_1$.

A less general alternative formulation is,

$$\hat{\mathbf{u}}_t = (\boldsymbol{B}_0 - \hat{\boldsymbol{B}})'\boldsymbol{x}_t + \sum_{i=1}^{p} \boldsymbol{C}_i' \hat{\boldsymbol{u}}_{t-i} + \boldsymbol{e}_t \tag{2.35}$$

which uses the linear combinations $\hat{\mathbf{u}}_{t-i} = \mathbf{y}_{t-i} - \hat{\mathbf{B}}'\mathbf{x}_{t-i}$ as the "omitted variables" from (2.20). This form may be useful when degrees of freedom are at a premium.[19]

The auxiliary regression corresponding to (2.32) is,

$$\hat{\mathbf{u}}_t = \sum_{i=1}^{p} \boldsymbol{\rho}_i' \hat{\boldsymbol{u}}_{t-i} + \boldsymbol{e}_t \tag{2.36}$$

and is directly related to the multivariate Portmanteau test (Lütkepohl (1991)) and LM error autocorrelation test (Godfrey (1988)), based on the null hypothesis H_0: $\boldsymbol{\rho}_i = \mathbf{0}$, $i=1,\ldots,p$.

In practice, tests based on (2.33) are likely to be most informative because (2.33) represents a natural respecification of the model. Tests based on (2.36) are likely to have poor power properties because (2.36) already imposes strong restrictions on the dynamics. This result is confirmed by Monte Carlo experiments in the single equation case (see Godfrey (1988)). However, (2.36) is informative to the extent that $\boldsymbol{\rho}_1$ is an estimate of the first-order autocorrelation matrix.

Joint Conditional Mean Tests

Often, misspecifications such as conditional mean non-linearity, non-stationarity, and temporal dependence, can result in similar behavior in the residuals (see Spanos (1992)). Hence, there is merit in considering joint misspecification tests for the conditional mean to get a better idea of the source(s) of misspecifications. The joint test could be based on an auxiliary regression of the form,

$$\hat{\mathbf{u}}_t = (B_0 - \hat{B})' x_t + C' \phi_t + D' w_t + \sum_{i=1}^{p} A_i \hat{u}_{t-i} + v_t, \tag{2.37}$$

where $\mathbf{w}_t$ are higher-order terms of the KG or RESET polynomials. The joint tests for conditional mean misspecification can be based on the null hypothesis H_0: $\mathbf{C} = \mathbf{D} = \mathbf{A}_i = \mathbf{0}$, $i = 1,\ldots,p$. Alternatively, each type of conditional mean misspecification could be examined individually.

The actual regressors used in (2.37) is constrained by concerns of collinearity and the availability of degrees of freedom. The trend polynomials in $\boldsymbol{\phi}_t$ are orthogonal (to each other) by construction, and r principle components of the KG or RESET polynomials may be used instead of in $\mathbf{w}_t$.

Currently, there is little evidence as to the practical usefulness of (2.37) in distinguishing among misspecifications, and thus represents an important area for further research.

2.4.2 Multivariate Conditional Covariance Tests

Auxiliary regression-based tests of the conditional covariance of the multivariate regression model are examined in this section. Specifically, separate and joint tests for static and dynamic heteroskedasticity are described.

In general, the form of auxiliary regression tests for the conditional variance are similar to those underlying the conditional mean tests. The null representation takes the form,

$$\boldsymbol{\xi}_t \equiv \mathrm{vech}(\mathbf{u}_t\mathbf{u}_t') = \mathrm{vech}(\boldsymbol{\Omega}) + \mathbf{e}_t, \tag{2.38}$$

where vech($\cdot$) denotes the vectorized lower triangular portion of a square matrix. Hence, if $\mathbf{y}_t$ is mx1, then $\boldsymbol{\xi}_t$ is ½m(m+1)x1. In (2.38) $\mathbf{e}_t$ will certainly be skewed to the right, but under the null of homoskedasticity it will have zero mean and homoskedastic covariance. Moreover, all the restrictions being tested are linear, so the usual test statistics can be employed. Since the diagonal elements of $\mathbf{u}_t\mathbf{u}_t'$ are positive by construction, there may be some advantage to using a log transformation of $\boldsymbol{\xi}_t$ if both even and odd functions are used as regressors in the alternative regression.

Heteroskedasticity

Under joint normality, not only is the conditional mean linear, but the conditional variance is homoskedastic. When the form of heteroskedasticity is thought to be a function of contemporaneous conditioning variables one might consider the auxiliary regression,

$$\boldsymbol{\xi}_t = \mathbf{c}_0 + \mathbf{C}_0'\mathbf{w}_t + \mathbf{v}_t, \tag{2.39}$$

where $\mathbf{w}_t$ denotes a vector of terms from a p^{th}-order KG or RESET polynomials, or a principle components form of these, and the null hypothesis is H_0: $\mathbf{C}_0 = \mathbf{0}$. A multivariate version of the heteroskedasticity test of White (1980) can be obtained by selecting $\mathbf{w}_t$ to be the second-order terms of the KG polynomial. Spanos (1986, 1991) shows that this test is an exact test against the form of heteroskedasticity implied by the Multivariate Student's t distribution. Similarly, the conditional covariances of all the members of the symmetric elliptical family of multivariate distributions are characterized by particular functions of this quadratic form (the normal is the limiting case where the quadratic form is a constant - see Spanos (1991)). Hence, a test for heteroskedasticity that should have power against the whole elliptical family can be formulated in the context of the auxiliary regression,

$$\xi_t = \mathbf{d}_0 + \mathbf{D}_0'\lambda_t + \epsilon_t. \tag{2.40}$$

where $\lambda_t = (\psi_t^{1/2}, \psi_t, \psi_t^2, \ldots)$ are functions of the second-order KG terms.

The presence of dynamic heteroskedasticity implies that both the distributional and independence assumptions may be invalid. Even when the conditional mean is static the past history may be important in predicting volatility (Spanos (1991)). The most common form of dynamic heteroskedasticity considered in the literature is the AutoRegressive Conditional Heteroskedasticity (ARCH) formulation due to Engle (1982). The auxiliary regression underlying the multivariate version of this test is,

$$\hat{\xi}_t = g + \sum_{i=1}^{p} G_i' \hat{\xi}_{t-i} + e_t, \tag{2.41}$$

where the null hypothesis is H_0: $\mathbf{G}_i = \mathbf{0}$, $i = 1,\ldots,p$. This test can be shown to be directly related to a multivariate version of the univariate McLeod and Li (1987) non-linear Portmanteau test based on the autocorrelation function of the squared residuals.

A more general formulation is provided by the dynamic quadratic form

$$\hat{\xi}_t = h + H'\psi_t + \sum_{i=1}^{p} R_i' \psi_{t-i} + v_t. \tag{2.42}$$

Exclusion of the contemporaneous effects gives rise to a dynamic multivariate White-type test, and a test of H_0: $\mathbf{R}_i = \mathbf{0}$, $i = 1,..,p$ can be viewed as a test for dynamic heteroskedasticity with the static form of heteroskedasticity maintained under the null. This form is useful in testing for unmodeled dynamic heteroskedasticity in Student's t regression models, because these models have a quadratic-form of heteroskedasticity under the null (see Chapter 4). When there are a large number of terms or collinearity is a problem, one may wish to use the RESET polynomials or principal components of the KG polynomial.

Joint Conditional Variance Tests

It is important to attempt to distinguish between dependence of the conditional covariance on time and heteroskedasticity, because the source of such departures are different; heteroskedasticity can be traced to non-normality (Spanos (1991)), and time-dependence to non-stationarity (Spanos (1990a)). An auxiliary regression that may be useful in this context is,

$$\boldsymbol{\xi}_t = \mathbf{a} + \mathbf{C}'\boldsymbol{\phi}_t + \mathbf{D}'\boldsymbol{\psi}_t + \mathbf{v}_t, \tag{2.43}$$

where the null hypothesis is H_0: $\mathbf{C} = \mathbf{D} = \mathbf{0}$. The two components of H_0 can be tested either separately or jointly. (2.43) can be readily extended by including lagged $\boldsymbol{\xi}_t$ and $\boldsymbol{\psi}_t$ terms to capture dynamic effects. If degrees of freedom are at a premium, the principal components of the quadratic forms might be used instead. As with the joint conditional mean test, there is little empirical evidence as to the performance of the joint conditional variance formulation.

2.4.3 Multivariate Normality Tests

As has been argued in Sections 2.4.1 and 2.4.2, tests for conditional mean linearity and homoskedasticity are indirect tests for normality, since it is difficult to justify conditional normality on statistical grounds without these properties.

Numerous direct tests for univariate normality have been developed in the literature (see for example, MacDonald and White (1980), Mardia (1980)). In the multivariate case, skewness and kurtosis are measured by the quantities,

$$\alpha_{3,m} = E[(u_t'\Omega^{-1}u_t)^3]^{1/2}, \text{ and } \alpha_{4,m} = E[(u_t'\Omega^{-1}u_t)^2], \tag{2.44}$$

where $\alpha_{3,m} = 0$, and $\alpha_{4,m} = m(m+2)$ under multivariate normality (see Kendall et.al., (1987)).

The sample versions $\hat{\alpha}_{3,m}$ and $\hat{\alpha}_{4,m}$ are,

$$\hat{\alpha}_{3,m} = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} g_{ts}^3, \qquad \hat{\alpha}_{4,m} = \frac{1}{T}\sum_{t=1}^{T} g_{tt}^2 \tag{2.45}$$

where $g_{ts} = \hat{u}_t'\hat{\Omega}^{-1}\hat{u}_s$, $t,s = 1,2,\ldots,T$.

This sub-section describes a number of multivariate normality tests based on (2.44) and (2.45). Specifically, the tests of Mardia (1973,1974) and Mardia and Foster (1983); Bera and John (1983); and Small (1980). Modified versions of the Small tests and auxiliary regression-based tests of the third and fourth conditional moments are also described.

Mardia Tests

Mardia (1970) shows that under null hypothesis of multivariate normality,

$$MS_1 = (T/6)\alpha_{3,m}, \tag{2.46}$$

is asymptotically distributed chi-square, with $f = m(m+1)(m+2)/6$, degrees of freedom. Mardia (1974) deduces that asymptotically,

$$E(\hat{\alpha}_{3,m}) = \frac{m(m+2)}{(T+1)(T+3)}[(T+1)(m+1)-6] \tag{2.47}$$

and, thus, recommends the adjusted statistic

$$MS_1^* = \frac{T}{6}\hat{\alpha}_{3,m}\left[\frac{(m+1)(T+1)(T+3)}{T[(T+1)(m+1)-6]}\right] \sim \chi^2(f). \tag{2.48}$$

For kurtosis, Mardia (1970) shows that asymptotically

$$MK_1 = [\hat{\alpha}_{4,m} - m(m+2)]/[8m(m+2)/T]^{1/2} \sim N(0,1). \tag{2.49}$$

When $\hat{\alpha}_{4,m} < E(\hat{\alpha}_{4,m})$ (platykurtic data), Mardia (1974) recommends regarding $\hat{\alpha}_{4,m}$ as normally distributed with $E(\hat{\alpha}_{4,m}) = m(m+2)(T+m+1)/T$, and $Var(\hat{\alpha}_{4,m}) = 8m(m+2)/(T-1)$ if $T < 400$. If $\hat{\alpha}_{4,m} > E(\hat{\alpha}_{4,m})$ (leptokurtic data) Mardia recommends MK_1 as a test for excess kurtosis.

Mardia and Foster (1983) considers other normalizing transformations of $\hat{\alpha}_{3,m}$ and $\hat{\alpha}_{4,m}$. They show that asymptotically,

$$MS_2 = [6(2f)]^{-\frac{1}{2}}\left[6(\frac{4}{3}Tf^2\hat{\alpha}_{3,m})^{\frac{1}{3}} - 18f + 4\right] \sim N(0,1) \tag{2.50}$$

and, $$MK_2 = \left[\frac{2}{9\nu}\right]^{-\frac{1}{2}} \left[\left[1 - \frac{2}{9\nu}\right] - \left[\frac{1 - \frac{2}{\nu}}{1 + MK_1 \left[\frac{2}{\nu - 4}\right]^{\frac{1}{2}}} \right]^{\frac{1}{3}} \right] \sim N(0,1) \qquad (2.51)$$

where

$$\nu = 6+T^{½}\{8m(m+2)(m+8)^{-2}\}^{½}[\{½m(m+2)\}^{½}(m+8)^{-1}T^{½} + \{1+½m(m+2)m+8)^{-2}T\}^{½}].$$

To account for the covariance between $\hat{\alpha}_{3,m}$ and $\hat{\alpha}_{4,m}$. Mardia and Foster (1983) use a first-order Taylor expansion and taking expectations to show that, as an approximation

$$\begin{aligned} Cov(MS_2,MK_2) \equiv C = 3(f/2)^{½}\{72f\}^{-½}\{&-(40/9)(1 - (2/f))(1/(f-4)) + \\ &(T/3\sigma(\hat{\alpha}_{4,m}))(1-(2/f))^{¼}\{2/(f-4)\}^{½}Cov[b_{1m},b_{2m}]\}. \end{aligned} \qquad (2.52)$$

This can then be used to define the asymptotic joint normality test statistic

$$MSK_1 = [MS_2, MK_2] \begin{bmatrix} 1 & C \\ C & 1 \end{bmatrix}^{-1} [MS_2, MK_2]' \sim \chi^2(2). \qquad (2.53)$$

Bera-John Tests

The Jarque and Bera (1987) normality test (see Appendix 2C) is generalized to the multivariate case by Bera and John (1983), who consider the characterization of multivariate normality in the multivariate Pearson family (see Johnson and Kotz (1972)).

Bera and John define,

$$V_i = \Sigma_{t=1}^{T} v_{it}^3,\ V_{ii} = \Sigma_{t=1}^{T} v_{it}^4,\ V_{ij} = \Sigma_{t=1}^{T} v_{it}v_{jt}, \qquad (2.54)$$

where $\mathbf{V} = (\mathbf{v}_1,\ldots,\mathbf{v}_m)' = \mathbf{Ru}$, such that $\mathbf{R'R} = \mathbf{\Omega}^{-1}$.

Under the null, they show that

$$E[V_i] = 0,\ Var[V_i] = 6/T,\ E[V_{ii}] = 3,\ Var[V_{ii}] = 24/T,\ E[V_{ij}] = 1,\ Var[V_{ij}] = 4/T. \quad (2.55)$$

Using these results they formulate the asymptotic chi-square skewness and kurtosis test statistics,

$$C_1 = T\Sigma_{i=1}^{m} V_i^2/6 \sim \chi^2(m), \quad (2.56)$$

$$C_2 = T[\Sigma_{i=1}^{m}(V_{ii}-3)/24 + \Sigma_{i<j}^{m}(V_{ij}-1)/4] \sim \chi^2(m(m+1)/2), \quad (2.57)$$

and the joint normality test statistics,

$$C_3 = T[\Sigma_{i=1}^{m} V_i^2/6 + \Sigma_{i=1}^{m}(V_{ii}-3)/24] \sim \chi^2(2m), \quad (2.58)$$

and

$$C_4 + C_2 \sim \chi^2(m(m+3)/2). \quad (2.59)$$

Bera and John find that both the Mardia and multivariate Jarque-Bera statistics often have very low power, and the size of all the C_i statistics are well below their nominal values for samples sizes less than 500. This result suggests neither of these families of tests may be particularly useful in application, and could be due to the nature of the approximations to asymptotic normality used in constructing the tests.

Small Tests

Small (1980) suggests a multivariate normality test based on the simultaneous testing of the marginal skewness and kurtosis coefficients, by combining them into an appropriate quadratic form. Let $\mathbf{z}_1$ be the mx1 vector of marginal skewness coefficients with covariance $\mathbf{V}_1$; let $\mathbf{z}_2$ and $\mathbf{V}_2$ be similarly defined for the marginal kurtosis coefficients. Kendall et.al., (1987) have shown that the distributions of $\mathbf{z}_1$ and $\mathbf{z}_2$ will approach normality only slowly and the quadratic forms $\mathbf{z}_i'\mathbf{V}^{-1}\mathbf{z}_i$ $(i=1,2)$ are not well approximated by a χ^2 distribution.

Applying the Johnson (1949) S_U normalizing transformations component-wise to the marginal coefficients yields,

$$\mathbf{y}_1 = \delta_1 \sinh^{-1}(\mathbf{z}_1/\lambda_1), \text{ and } \mathbf{y}_2 = \gamma_2\mathbf{1} + \delta_2\sinh^{-1}\{(\mathbf{z}_1-\xi\mathbf{1})/\lambda_2\} \tag{2.60}$$

The marginal distribution of the components of $\mathbf{z}_1$ and $\mathbf{z}_2$ have been derived, from which the appropriate values of the transformation parameters can be obtained using the corresponding moments of the elements of $\mathbf{z}_1$ and $\mathbf{z}_2$ (Kendall et.al., (1987)). Since the S_U transformation makes the components of $\mathbf{y}_1$ and $\mathbf{y}_2$ are approximately standard normal, the main diagonal of their covariance matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ will be unity. Asymptotically, Small (1980) shows that the correlation between the i^{th} and j^{th} elements of $\mathbf{z}_1$ is ρ_{ij}^3 and of $\mathbf{z}_2$ is ρ_{ij}^4, where ρ_{ij} is the correlation between the corresponding components of $\mathbf{u}_t$. The skewness and kurtosis statistics can then be formed as

$$MS_2 = \mathbf{y}_1'\mathbf{C}_1^{-1}\mathbf{y}_1, \text{ and } MK_2 = \mathbf{y}_2'\mathbf{C}_2^{-1}\mathbf{y}_2. \tag{2.61}$$

Since MS_3 and MK_3 are approximately independent, an asymptotic chi-square test for normality can be based on

$$MSK_2 = MS_2 + MK_2 \sim \chi^2(2m). \tag{2.62}$$

A modification of these tests is obtained by using the D'Agostino (1970) and the Anscombe and Glynn (1983) transformations rather than the S_U approximation (see Appendix 2C). This formulation has the additional advantage that it can be readily implemented without using interpolated values from the tables associated with the S_U translation.

Clitic and Kurtic Tests

One of the key assumptions underlying normality tests based on the skewness and kurtosis coefficients is the assumption that $\alpha_{3,m}$ and $\alpha_{4,m}$ are constants. However, any test based on this assumption will be invalid if there exists heteroclicitity and heterokurticity of the third and fourth conditional moments, respectively. This suggests that testing the validity of these assumptions in conjunction with any parametric normality tests is important.

One may consider individual tests along the lines discussed previously, or via joint formulations such as,

$$\hat{\kappa}_t = \mathbf{a}_1 + \mathbf{C}_1'\phi_t + \mathbf{D}_1'\psi_t + \mathbf{v}_{1t}, \tag{2.63}$$

and

$$\hat{\eta}_t = \mathbf{a}_2 + \mathbf{C}_2'\phi_t + \mathbf{D}_2'\psi_t + \mathbf{v}_{2t}, \tag{2.64}$$

where $\hat{\kappa}_t = \mathrm{vec}(\hat{u}_t^2\hat{u}_t')$ as an m^2x1 vector, and $\hat{\eta}_t = \mathrm{vech}(\hat{u}_t^2\hat{u}_t^{2\prime})$ is a $\frac{1}{2}m(m+1)$x1 vector. These two multivariate regressions could be estimated jointly, and can be readily extended by including lagged terms to account for possible non-linear dynamics.

2.5.4 Simultaneous Test Formulations

In addition to the separate and joint tests of conditional moments described above, one can also consider misspecification tests for the various conditional moments simultaneously. Simultaneous testing has been explored in the single equation context by Bera and Jarque (1982), and Pagan and Hall (1983). The obvious auxiliary regression formulation is the SUR model of Zellner (1962) because it allows for exclusion restrictions within particular sub-systems of the model, and contains the usual uniform linear restrictions as a special case (see Appendix 2B).

A general alternative specification would take the form,

$$\hat{\pi}_t = C_0'x_t + C_1'\phi_t + C_2'w_t + \sum_{i=1}^{p} C_{3i}'v_{t-i} + \sum_{j=1}^{n} C_{4j}'\hat{u}_{t-j} + \zeta_t \tag{2.65}$$

where $\hat{\pi}_t = (\hat{u}_t', \xi_t', \hat{\kappa}_t', \hat{\eta}_t')'$, w_t are higher order KG, RESET terms, or modified version thereof, and v_{t-i} could be w_{t-i} or ξ_{t-i}. In the case where x_t is excluded from the specification of the higher-order moment sub-systems, the SUR formulation would be used.

Tests based on (2.65) extends the usual simultaneous testing approach in that it is based on a respecification argument that allows for effects that are not always apparent when concentrating exclusively on the conditional distribution.

Except when sample size is sufficiently large, the practical usefulness of (2.65) is somewhat restricted. Thus, it is often the results from a number of separate misspecification tests that must be used in judging the statistical adequacy of a model, and more importantly, how to respecify the model in order to take account of any misspecification(s). MacKinnon (1992) has argued that the multiple comparison of tests results requires a great deal of care, since it is quite possible that several tests formulated against different alternatives might be rejected, even though there is only a single source of misspecification.

As described in Section 2.3, the Haavelmo reduction framework plays an important role in tracing the symptoms of misspecification to the underlying probabilistic structure of sample. Utilizing knowledge of various distributional relationships (e.g., between homoskedasticity and normality), the information provided by data plots and misspecification tests provides a basis for considering alternative reduction assumptions and model specifications. The implementation of such misspecification tests and related diagnostic measures is the focus of the next section.

2.5 Implementing Misspecification Tests: SAM.

In this section I describe how a number of the misspecification tests presented in Section 2.4 can be implemented using the menu-driven interactive computer program SAM (Statistically Adequate Model). This program is written jointly by Dr's Aris Spanos, Anya McGuirk, and myself.[20]

While there are a number of diagnostic-based interactive econometrics packages (e.g., PC-GIVE (Hendry (1991)), MICROFIT (Pesaran and Pesaran (1990)), these programs currently provide no multivariate data analysis tools and multi-equation misspecification testing options. SAM attempts to address this issue by providing a coherent and practical tool for conducting data analysis for: multivariate time-series data; the estimation of small dynamic (multivariate) linear regression-type statistical models; single-equation and system misspecification testing; linear SEM estimation; testing; and forecasting.[21]

SAM consists of a set of procedures (sub-routines) called from a main program written in the GAUSS matrix language (Aptech Systems (1992)). Currently, GAUSS version 3.0 and at least a 386 co-processor are required to run SAM. GAUSS is chosen primarily because it is an efficient matrix language, although future versions of SAM may be translated into a format that can be accessed independently of the GAUSS compiler/interpreter. The structure of the program is summarized in Figure 2.1. The solid lines represent relationships among menus called for each equation in the statistical model, while the dashed lines connect the components of SAM used for multi-equation statistical models.

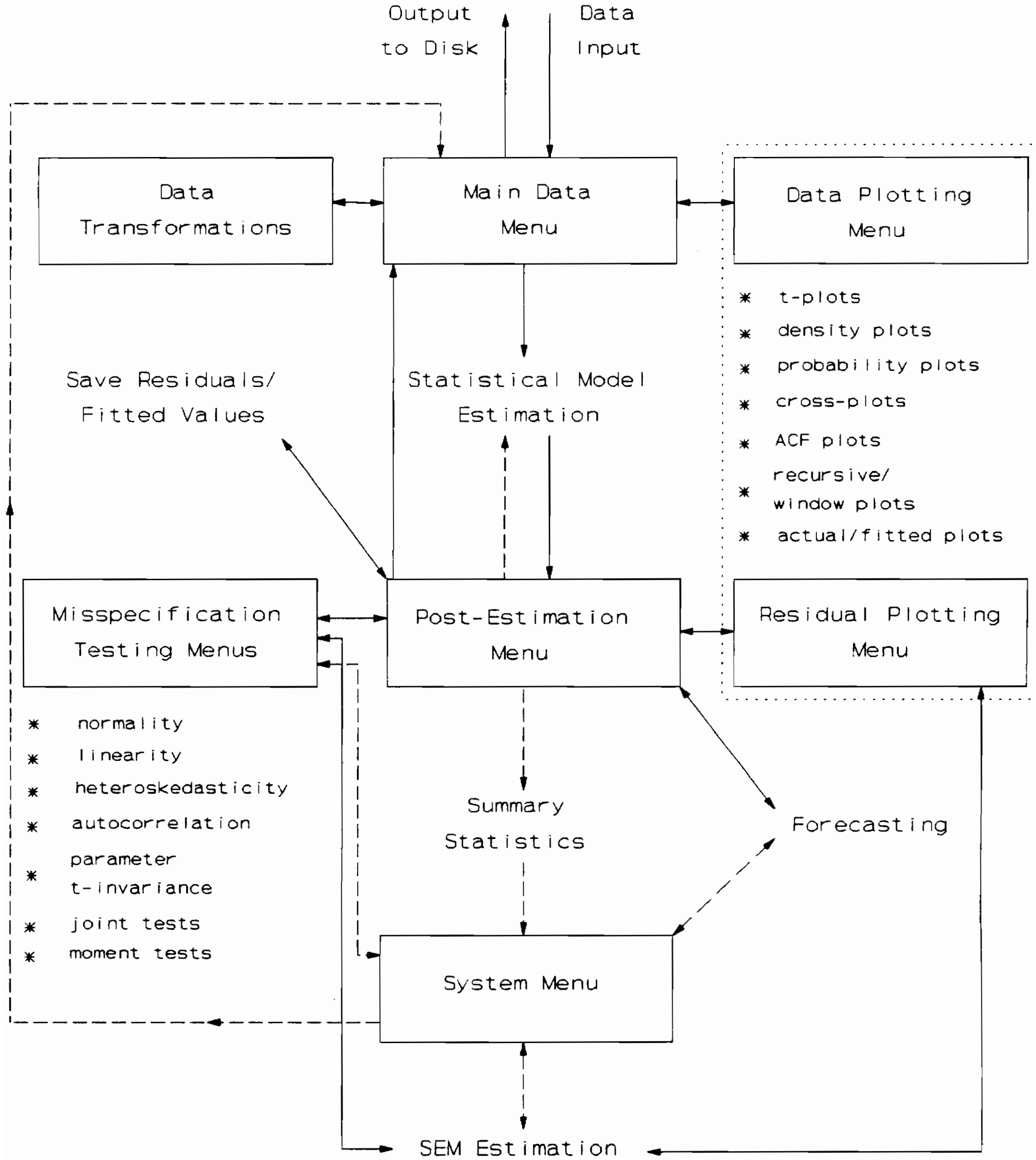


Figure 2.1 - SAM Program Structure Overview

The first operational menu in SAM is the main data menu. This menu provides: data transformation options; reports summary statistics; options for saving data and directing output to disk; and a data plotting menu. The available data transformations include: logarithm/exponent of a variable; differencing a variable and creating lags; scaling a variable; adding/subtracting or multiplying/dividing two variables; listing, renaming and deleting variables; creating a constant, orthogonal trends, and dummy variables.

The summary statistics reported are: the mean; standard deviation; minimum/maximum; and skewness and kurtosis, for each variable in the data set. Correlations among variables are also optionally reported.

The data plotting menu contains several methods of graphical exploratory data analysis (see Spanos (1992) for an extensive discussion of plot reading in this context):

(a) Plots of a variable against the index set, with up to three series plotted on one graph. All the plotting procedures contain options for modifying the axis range, scaling, axis labeling, title, output and format conversion. The t-plots can be useful in determining the homogeneity, memory, and distributional features of the data.

(b) Two and three-way cross plots (scatter diagrams), which can give some idea of the nature of dependence among variables and the distributional shape.

(c) Univariate autocorrelation and partial autocorrelation functions, and cross-correlation functions for two or three variables. Together these can be used to indicate the nature of linear temporal dependencies in the data over time and among the variables.

(d) Plots of recursive and τ-period window (rolling) estimates of the variance of a series and covariance between two series over time, measured around either fixed or trending means. These plots are useful for indicating t-heterogeneity in the second moments of the data.

(e) Probability plots and an empirical marginal cumulative distribution function (cdf) for a series. The use of probability plots is a popular tool for analyzing the distribution of random samples (D'Agostino and Stephens (1986)). By transforming the scale of the empirical cdf, the normal probability plot is a straight line, which makes visual inspection for non-normality rather straightforward. For platykurtic data the probability plot has an S shape, and an inverted S-shape for leptokurtic data.

(f) Non-parametric kernel density estimates for the marginal distribution of a series (see Silverman (1986) for an extensive discussion). A kernel estimator is a sum of "bumps" placed at the ordered observations. Options are available to select the kernel function and the bandwidth (smoothing parameter) which determine the shape of the "bumps" and their width, respectively. Reference distributional plots can also be super-imposed on the density estimate for comparative purposes. Non-parametric bi-variate kernel density estimates are available using a standardized bivariate normal or Epanechinikov kernel with the same correlation as the data. Options allow plotting of the estimated density surface, contour plots and marginal density projections. Reference bi-variate densities are available for comparison purposes.

After specifying the dimension of the regression model, and selecting the relevant regressors (including orthogonal trends, dummies and lagged variables) from the data set, the statistical model is estimated by (multivariate) LS (maximum likelihood under normality). The regression results for each equation in the statistical system are reported in standard regression format. The GAUSS LS routine is modified to include White's (1980) heteroskedastic consistent standard errors. The availability of the residual sum of squares in the output permits the user to compute F-tests for linear restrictions on an equation that are not reported automatically.

Residual diagnostics are available for each regression equation in the system via the post-estimation menu. The residual plotting menu essentially contains all the options from the data plotting menu, plus plots of the actual and fitted series, and linearity/heteroskedasticity cross-plots. These latter plots are composed of a scatter diagram of the residuals and squared residuals (or logs of these) against various functions of a regressor, and may be useful in identifying the sources non-linearity and heteroskedasticity in a model. An option for fitting the LS line is also available.

The single equation misspecification testing menu provides the F-type formulations of the multivariate tests discussed in the previous section. For each test, the value of the statistic, its distribution, and the p-value are reported. Currently, the following tests are available:[22]

(a) Normality - D'Agostino/Pearson and Jarque-Bera tests (see Appendix (2C));

(b) Linearity - RESET and KG tests;

(c) Heteroskedasticity - RESET, KG, and White tests;

(d) Dynamic Heteroskedasticity - ARCH, dynamic White, and McLeod-Li (1978) tests;

(e) Autocorrelation - modified LM, Ljung-Box (1978) portmanteau tests;

(f) Parameter stability - Recursive and window LS estimation, and Chow-type tests;

(g) Joint conditional mean, variance, skewness and kurtosis tests;

(h) Newey (1985) and Tauchen (1985) moment tests.

The parameter stability sub-menu allows plotting of recursive and window LS estimates of all the parameters of the equation (including σ^2), with and without standard error bars, and plots the CUSUM and CUSUMSQ statistics (Brown, Durbin and Evans (1975)). F-type tests for parameter t-dependence of the parameters of the conditional mean, and the conditional variance are computed and are reported with user-defined break-points (see Spanos (1986, pp.474-478)). A number of influential observation diagnostics are available, included t-plots of the leverage points, DFITs, and Studentized residuals (see Myers (1990)). The moment testing option reports Newey (1985) tests for the linear regression model as described in Section 2.1.

The post-estimation menu also contains an option for examining the ex-post predictive performance of the equation based on the last p observations. Standard forecast performance statistics are reported, and the actual and predicted series can be plotted over time. Finally, the post-estimation menu allows the residuals and fitted values from the equation to be saved in the data set for subsequent analysis.

For multivariate statistical models the system menu is available. This menu provides options for structural model estimation based on two and three- stage LS (instrumental variables); system ex-post forecast analysis; misspecification testing; and reporting multivariate model summary statistics. The system "goodness-of-fit" statistics reported are based on the matrix quantity $\mathbf{G} = (\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{U}}'\hat{\mathbf{U}})(\mathbf{Y}'\mathbf{Y})^{-1}$, where $\hat{\mathbf{U}}$ is the Txm residual matrix, and $\mathbf{Y}$ is in mean deviation form. $\mathbf{G}$ varies between an identity matrix when $\hat{\mathbf{U}} = \mathbf{0}$, and zero when $\mathbf{Y} = \hat{\mathbf{U}}$. Hooper (1959) suggests the quantity tr$\mathbf{G}$/m (trace correlation) or $|\mathbf{G}|$ (vector alienation correlation). These can be easily computed using the eigenvalues of $\mathbf{G}$. The system log-likelihood value, various system information criteria, $\hat{\Omega}$ and $|\hat{\Omega}|$, are also reported so that tests of linear restrictions on the system can also be computed by the user.

The menu of multivariate misspecification tests currently provides the following tests:

(a) Multivariate Normality - Mardia and modified Small tests;

(b) Linearity - multivariate RESET, KG, and modified KG tests;

(c) Heteroskedasticity - multivariate RESET, KG, White, and modified KG tests;

(d) Dynamic Heteroskedasticity - multivariate ARCH and dynamic White tests;

(e) Autocorrelation - multivariate modified LM, and portmanteau tests;

(f) Joint multivariate conditional mean, variance, skewness and kurtosis tests.

Figure 2 is a typical example of how the output is presented in SAM for multivariate tests. The first is the normality test output screen, which reports the Mardia (MS_1, MK_1, MSK_1) and the modified version of the Small (1980) normality test statistics (see Section 2.4.3), p-values ($\cdot$) for the tests, and the marginal skewness/kurtosis coefficients. The second screen is for a first-order test of autocorrelation using the modified LM formulation (2.35). As with all the auxiliary regression misspecification tests the Wald, adjusted LR, Rao F-approximation, and the LM statistics, along with the sample size (T), the number of equations (m), the number of restrictions per equation (p), the number of regressors in per unrestricted equation (k), and the Bartlett LR adjustment ($T^* = T\text{-}k\text{-}.5(m\text{-}p+1)$), are reported (see Appendix 2B for details).

Structural models can be analyzed via the SEM estimation menu. Single equation IV, and system two- and three-stage LS estimators are available. For each equation in the SEM the identification status, and tests of the over-identifying restrictions are reported (Anderson and Rubin (1949) and Basmann (1960)). The restricted structural equation residuals can be examined by the residual plotting and misspecification testing menus.

Multivariate Normality Tests			
Mardia (1983) Normality test:	5.676585	(0.058526)	[Chi^2(2)]
Mardia (1974) Skewness test:	10.472118	(0.033183)	[Chi^2(4)]
Mardia (1970) Kurtosis test:	6.970955	(0.008284)	[Chi^2(1)]
Modified-Small Normality test:	17.409903	(0.001609)	[Chi^2(4)]
Modified-Small Skewness test:	8.313493	(0.015658)	[Chi^2(2)]
Modified-Small Kurtosis test:	9.096410	(0.010586)	[Chi^2(2)]
Marginal Normality statistics:			
	α_3 = 0.588789	0.097805	
	α_4 = 4.760007	2.902309	

Multivariate Autocorrelation Tests (l = 1)			
T = 149, T* = 131.5, m = 2, p = 2, k = 17			
Wald test (trace form):	4.903372	(0.297357)	[Chi^2(4)]
LR test (Bartlett adj):	4.800144	(0.308425)	[Chi^2(4)]
LR test (Rao F-approx):	1.206449	(0.308432)	[F(T-k*,4)]
LM test (trace form):	4.735415	(0.315538)	[Chi^2(4)]

det(GHAT):	0.000861	det(GTIL):	0.000636
tr(GHAT):	0.037147	tr(GTIL):	0.035874
max(GHAT):	0.036302	max(GTIL):	0.035031

Figure 2.2 - Misspecification Test Output Screens

Appendix 2A

The Method of Moments

The Method of Moments (MM) estimation strategy has been proposed in the literature as a unified framework for estimation and inference in econometrics in a wide range of applications (see Pagan and Wickens (1989) for a recent survey). Heuristically, it is taken that the econometric model of interest implies a set of orthogonality conditions (as suggested by economic theory, say), where there are at least as many orthogonality conditions as there are parameters to be estimated. An MM estimator of the parameter vector is computed by finding the elements of the parameter space that set the sample estimates of the orthogonality conditions as close to zero as possible. In this sense, MM is based on a moment analogy principle where functions of sample statistics are chosen so as to have the same property (at least asymptotically) as the functions of parameters to be estimated (Manski (1988)). Applications of MM include models of market efficiency which utilize the implied informational orthogonality conditions (Hansen and Hodrick (1980)), and rational expectations models (Hansen and Singleton (1982)).

The statistical theory surrounding MM estimators is asymptotic, and consistency and asymptotic normality results have been obtained for random (IID) processes, and for stationary and ergodic (SE) processes (Hansen (1982)). Hansen (1982, p.1031) argues, while MM estimators are not usually fully efficient, in the absence of prior information about the underlying distribution, this loss in efficiency is a reasonable price to pay for a general estimation framework. Relative efficiency arguments are available within the class of estimators considered, where efficiency is taken to mean that the asymptotic variance is smallest given the information used in defining the orthogonality conditions (see Chamberlain (1987)).

To define the MM estimator a number of regularity and continuity conditions are made for the moment functions, and the parameter space. The key operational assumptions are:

(a) $\mathbf{Z}_t \equiv (y_t, \mathbf{X}_t')'$ (mx1) is a random vector with finite second moments, representing a realization of an IID or SE stochastic process $\{\mathbf{Z}_t, t \in \mathrm{N}\}$ with density $D(\mathbf{Z}_t;\boldsymbol{\theta})$. $\boldsymbol{\theta}_0$ is the (px1) vector of unknown parameters to be estimated, $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$.

(b) $\mathbf{g}_t(\mathbf{Z}_t,\boldsymbol{\theta})$: $\mathbf{R}^m \times \boldsymbol{\Theta} \rightarrow \mathbf{R}^q$, $q \geq p$ is a vector of Borel functions of $\mathbf{Z}_t$ continuous for each $\mathbf{Z} \in \mathbf{R}^m$, and $\mathbf{g}_T(\boldsymbol{\theta}) = T^{-1}\Sigma_{t=1}^T \mathbf{g}_t(\mathbf{Z}_t;\boldsymbol{\theta})$ is the vector of sample analogues.

(c) $E[\mathbf{g}_t(\mathbf{Z}_t,\boldsymbol{\theta}_0)] = \int \mathbf{g}_t(\mathbf{Z}_t,\boldsymbol{\theta}_0)D(\mathbf{Z}_t;\boldsymbol{\theta})d\mathbf{Z}_t = \mathbf{0}$.

Assumption (a) imposes general restrictions on the heterogeneity and memory structure of the stochastic process of interest. Assume that $\{\mathbf{Z}_t, t \in \mathrm{N}\}$ is an IID process for convenience in what follows. Assumption (b) defines a set of continuous and measurable functions of $\mathbf{Z}_t$ and $\boldsymbol{\theta}$. $E(\mathbf{Z}_t\mathbf{Z}_t') < \infty$ implies that $\mathbf{V} = E[\mathbf{g}_t(\mathbf{Z}_t;\boldsymbol{\theta})\mathbf{g}_t(\mathbf{Z}_t;\boldsymbol{\theta})'] < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ if $\boldsymbol{\Theta}$ is well-defined. Assumption (c) defines the orthogonality moment restriction so that when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ the mean of the chosen functions are simultaneously zero. Note, the expectation operator is with respect to the joint distribution of $\mathbf{Z}_t$.

From (c) the MM estimator is obtained by solving the sample analog $\mathbf{g}_T(\boldsymbol{\theta}) = \mathbf{0}$. Hansen (1982) suggests choosing $\hat{\boldsymbol{\theta}}$ so as to minimize,

$$\mathbf{g}_T(\hat{\boldsymbol{\theta}})'\mathbf{W}\mathbf{g}_T(\hat{\boldsymbol{\theta}}) \tag{A2.1}$$

where $\mathbf{W}$ is some appropriately chosen positive definite weighting matrix defining the metric by which $\mathbf{g}_T(\boldsymbol{\theta})$ is made close to zero.

Assuming that $\{\mathbf{Z}_t, t \in \mathrm{N}\}$ is stationary and ergodic implies that $\mathbf{g}_t(\mathbf{Z}_t;\boldsymbol{\theta}_0)$ is also stationary and ergodic. Thus the (uniform) Law of Large Numbers (LLN) and central limit theorem for stationary and ergodic processes can be applied (see Spanos (1986, p.180)). For an IID process, the consistency and asymptotic normality results are straightforward. For asymptotic normality define,

$$\mathbf{g}_{\theta t}(\mathbf{Z}_t;\theta) = [\mathbf{g}_{\theta t}(\mathbf{Z}_t;\theta)_{ij}] = \partial g_{it}(\mathbf{Z}_t;\theta)/\partial\theta_j, \text{ i,j} = 1,\ldots q, \tag{A2.2}$$

and assume the (qxp) matrix $\mathbf{F} = E[\mathbf{g}_{\theta t}(\mathbf{Z}_t;\theta)] < \infty$, and is continuous at θ_0 and of rank p. Then by Theorem 3.1 of Hansen (1982, p.1042) (see also Manski (1988, p.118)),

$$T^{1/2}(\hat{\theta} - \theta_0) \Rightarrow N(0,(\mathbf{F}'\mathbf{W}\mathbf{F})^{-1}(\mathbf{F}'\mathbf{W}\mathbf{V}\mathbf{W}\mathbf{F})(\mathbf{F}'\mathbf{W}\mathbf{F})^{-1}), \tag{A2.3}$$

where $\Rightarrow$ denotes convergence in distribution. If p = q, the MM estimator is said to be "just identified", and **W** can be any positive definite matrix. If q > p, θ is "over-identified" and Hansen (1982) (following Sargan (1958)) shows that the best choice of weighting matrix (in the sense of minimizing the asymptotic covariance of $\hat{\theta}$) is $\mathbf{V}^{-1}$.

In general, **V** is not known, so a two stage Generalized MM (GMM) procedure is used. First obtain a consistent estimator of θ_0 and **V** (pxp) using only p of the moment restrictions. Then, obtain the estimator of θ_0,

$$\hat{\theta} = \min_{\theta\in\Theta}\left[T^{-1}\sum_{t=1}^{T} g_t(Z_t;\theta)'\right]\hat{V}^{-1}\left[T^{-1}\sum_{t=1}^{T} g_t(Z_t;\theta)\right] \tag{A2.4}$$

and $T^{1/2}(\hat{\theta} - \theta_0) \Rightarrow N(0,(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1})$. Note, that for the estimators,

$$\hat{\mathbf{F}} = T^{-1}\sum_{t=1}^{T} g_{\theta t}(Z_t;\hat{\theta}), \qquad \hat{V} = T^{-1}\sum_{t=1}^{T} g_t(Z_t;\hat{\theta})g_t(Z_t;\hat{\theta}) \tag{A2.5}$$

under mild regularity conditions $\hat{\mathbf{F}}^{-1\prime}\hat{\mathbf{V}}\hat{\mathbf{F}}^{-1} \xrightarrow{p} (\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}$ (Hansen (1982, Lemma 3.3)).

To see the relationship of MM to ML estimation, recall that ML of θ solves a px1 system of equations (scores) $s(\theta)$. Since p = q, MM and ML will coincide when $g_t(Z_t;\theta) = T^{-1}s_t(\theta)$, and

$\mathbf{W}$ is any positive definite matrix. In general, $\hat{\theta}_{ml}$ has the property that it is consistent, asymptotically normal and (fully) efficient in that $T^{1/2}(\hat{\theta}_{ml} - \theta_0) \Rightarrow N(0,\mathbf{I}(\theta)^{-1})$, where $\mathbf{I}(\theta)$ is the asymptotic information matrix, $\lim_{T\to\infty} T^{-1}\mathbf{I}_T(\theta)$, and $\mathbf{I}_T(\theta)$ is sample Information matrix. This property arises because ML utilizes all the available information in defining the orthogonality conditions.

To illustrate the use of MM, consider the modeling of the spot/forward rate relationship in the foreign exchange market as described in Hansen and Hodrick (1980). Let $y_{t+k} = s_{t+k} - f_{t,k}$ be the k-period ahead forward premium; i.e., the difference between the t+k spot price of some asset and the k-period-ahead forward rate at time t. According to economic theory, $f_{t,k}$ represents the "rational expectation" of the future spot price, incorporating all relevant information available at time t, $\mathfrak{I}_t$. In particular, decompose the future spot price as,

$$s_{t+k} = E(s_{t+k} \mid \mathfrak{I}_t) + u_t \qquad \text{(A2.6)}$$

where $E(s_{t+k} \mid \mathfrak{I}_t) = f_{t,k}$, $E(u_{t+k} \mid \mathfrak{I}_t) = 0$, $E[f_{t,k}u_{t+k}] = 0$. The efficiency argument would set y_{t+k} orthogonal to all time t information, $\mathfrak{I}_t = \sigma(\mathbf{X}_t)$, where $\mathbf{X}_t$ is a q-dimensional vector of conditioning variables such as the past history of the premium. The proposed specification embodying the rational expectations argument is

$$y_{t+k} = \boldsymbol{\delta}_0'\mathbf{x}_t + \epsilon_{t+k}. \qquad \text{(A2.7)}$$

We have $E[\mathbf{x}_t'\epsilon_{t+k}] = E[g_t(y_{t+k},\mathbf{x}_t,\boldsymbol{\delta}_0)] = 0$ as the set of orthogonality conditions for $\boldsymbol{\delta} = \boldsymbol{\delta}_0$, and if $k > 1$, ϵ_{t+k} will be autocorrelated. The estimates of $\boldsymbol{\delta}_0$ are obtained by minimizing $\mathbf{g}_T(\boldsymbol{\delta})'\mathbf{W}\mathbf{g}_T(\boldsymbol{\delta})$, where $\mathbf{W}$ is a qxq symmetric, positive definite weighting matrix. Because the orthogonality conditions are formed linearly with $\mathbf{X}_t$, it is argued that autocorrelation does not effect the consistency of an estimator such as OLS. However, the asymptotic covariance of the estimator depends on **W**. To obtain the "optimal" covariance matrix $(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}$, **F** and $\mathbf{V}^{-1}$ can be estimated by

$$\hat{\mathbf{F}} = T^{-1}\sum_{t=1}^{T} \frac{\partial g_t(y_{t+k}, x_t; \hat{\delta})}{\partial \delta}, \qquad \hat{V} = \left[R_T(0) + w(j,k-1)\sum_{j=1}^{k-1}\left(R_T(j) + R_T(j)'\right)\right] \tag{A2.8}$$

where $$\mathbf{R}_T(j) = T^{-1}\sum_{t=j+1}^{T} g_t(y_{t+k}, x_t; \hat{\delta})g_t(y_{t+k-j}, x_{t-j}; \hat{\delta})', \quad w(j,k-1) = 1 - \left[\frac{j}{k}\right] \tag{A2.9}$$

and $\hat{\delta}$ is the OLS estimator of δ_0. $\hat{\mathbf{W}} = \hat{\mathbf{V}}^{-1}$ is proposed by Newey and West (1987) as a consistent covariance estimator that ensures positive semi-definiteness (see also Andrews (1991)). The w(j,k-1) term is a spectral weight used to smooth the sample autocovariance functions (see Anderson (1971)).

From this discussion we can note that the theoretical rational expectations argument is that the conditional mean representation of y_{t+k} coincides with $f_{t,k}$ apart from a white noise forecasting error. However, from a statistical perspective, the appropriate conditional mean representation is that function of the conditioning variables yielding an error process that contains no information useful for predictive purposes. Hence, it is crucial that $\mathfrak{I}_t$ be specified explicitly, and the choice of $\mathfrak{I}_t$ is demarcated by the joint distribution of the sample. Moreover, the reliance on consistent estimators in the MM approach can be misleading. While error autocorrelation does not affect the consistency of LS in this example (Spanos (1986, p.502)), residual autocorrelation due to misspecified dynamics does lead to inconsistency (Spanos (1986, p.500)). The possibility of lack of correspondence between the statistical and theoretical expectations provides a source of specification tests for the model.

Appendix 2B

Testing Linear Restrictions in Multivariate Regression Models

To motivate the multivariate tests based on auxiliary regressions described in Section 2.4, consider the estimation of the parameters of the MLR model (2.20). The log-likelihood function for this model is,

$$\mathrm{LogL}(\theta) \propto -\frac{T}{2}\ln(|\Omega|) - \frac{1}{2}\sum_{t=1}^{T} u_t'\Omega^{-1}u_t = -\frac{1}{2}\left(T\ln(|\Omega|) + tr(\Omega^{-1}U'U)\right) \tag{B2.1}$$

where $\mathbf{U} = \mathbf{Y} - \mathbf{XB}$, $\mathbf{Y}$: Txm, $\mathbf{X}$: Txk, $\mathbf{B}$:kxm. Solving the resulting first-order conditions yields the ML estimators,

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \text{ and } \hat{\Omega} = \mathrm{T}^{-1}\hat{\mathbf{U}}'\hat{\mathbf{U}}, \tag{B2.2}$$

where $\hat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}$, $\mathrm{T} \geq \mathrm{m+k}$, $\mathbf{S} \equiv \hat{\mathbf{U}}'\hat{\mathbf{U}} = \mathbf{Y}'\mathbf{M}_x\mathbf{Y}$ is the unrestricted residual moment matrix, and $\mathbf{M}_x = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is an idempotent orthogonal projection matrix.

For misspecification testing the most useful type of restriction is the uniform type,

$$\mathrm{H}_0: \mathbf{DB} - \mathbf{C} = \mathbf{0}, \tag{B2.3}$$

ρ is # given with Γ, Δ and D

where $\mathbf{D} = (\mathbf{0},\mathbf{I}_p)$: pxk, $\mathbf{B} = (\mathbf{B}_1',\mathbf{B}_2')'$, $\mathbf{B}_1$: $\mathrm{k_1xm}$, $\mathbf{C} = \mathbf{0}$: pxm, $\mathrm{k} = \mathrm{k_1+p}$, $\mathrm{rank}(\mathbf{D}) = \mathrm{p} \leq \mathrm{k}$.[23]

The restricted log-likelihood function is

$$\mathrm{LogL}(\theta_0) \propto -\frac{1}{2}\left(T\ln(|\Omega|) + tr(\Omega^{-1}U'U) - tr(\Lambda'(DB - C)\right) \tag{B2.4}$$

where Λ is the pxm matrix of Lagrange multipliers. Solving the first-order conditions (Spanos (1986, p.582)) yields the restricted ML estimator

$$\tilde{\mathbf{B}} = \hat{\mathbf{B}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}'[\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}']^{-1}(\mathbf{D}\hat{\mathbf{B}} - \mathbf{C}), \text{ and } \tilde{\mathbf{\Omega}} = \mathrm{T}^{-1}\tilde{\mathbf{U}}'\tilde{\mathbf{U}}. \tag{B2.5}$$

The restricted residuals may be written as,

$$\tilde{\mathbf{U}} = \hat{\mathbf{U}} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}[\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}']^{-1}(\mathbf{D}\hat{\mathbf{B}} - \mathbf{C}), \tag{B2.6}$$

so that

$$\tilde{\mathbf{U}}'\tilde{\mathbf{U}} = \mathbf{S} + (\mathbf{D}\hat{\mathbf{B}} - \mathbf{C})[\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}']^{-1}(\mathbf{D}\hat{\mathbf{B}} - \mathbf{C}). \tag{B2.7}$$

Hence, the restricted covariance estimator can be written as

$$\tilde{\mathbf{\Omega}} = \hat{\mathbf{\Omega}} + \mathrm{T}^{-1}(\mathbf{D}\hat{\mathbf{B}} - \mathbf{C})[\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}']^{-1}(\mathbf{D}\hat{\mathbf{B}} - \mathbf{C}). \tag{B2.8}$$

For testing H_0 it is natural to consider,

$$\mathbf{H} = \tilde{\mathbf{U}}'\tilde{\mathbf{U}} - \hat{\mathbf{U}}'\hat{\mathbf{U}} = (\mathbf{D}\hat{\mathbf{B}} - \mathbf{C})[\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}']^{-1}(\mathbf{D}\hat{\mathbf{B}} - \mathbf{C}) = \mathbf{Y}'\mathbf{M}_d\mathbf{Y} \tag{B2.9}$$

where $\mathbf{M}_d = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}'[\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}']^{-1}\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{M}_d\mathbf{M}_x = 0$. Standardizing $\mathbf{H}$ via $\mathbf{H}\mathbf{\Omega}^{-1}$, and choosing the unrestricted covariance estimate as the standardization gives rise to the estimated quadratic form

$$\mathbf{G} = \mathbf{H}\mathbf{S}^{-1} = (\tilde{\mathbf{U}}'\tilde{\mathbf{U}} - \hat{\mathbf{U}}'\hat{\mathbf{U}})(\hat{\mathbf{U}}'\hat{\mathbf{U}})^{-1}. \tag{B2.10}$$

To reduce $\mathbf{G}$ to a scalar, we may use the functionals,

$$\text{(i) } \mathrm{tr}(\mathbf{G}) = \Sigma_{i=1}^{m}\lambda_i,\ i = 1,\ldots,\ell,\ \ell=\min(m,p), \tag{B2.11}$$

$$\text{(ii) } |\,\mathbf{G}\,| = \Pi_{i=1}^{m}\lambda_i = \Sigma_{i=1}^{m}\ln\lambda_i, \tag{B2.12}$$

where λ_i are the non-zero eigenvalues of $\mathbf{G}$, and tr denotes the trace. The asymptotic chi-square test statistics (under H_0) are all related via (B2.11) and (B2.12) and are described below.

Multivariate Test Criteria

The trace of **G** gives rise to the asymptotic chi-square *Wald* statistic,

$$W = (T\text{-}k)^{-1}\tau_1 \sim \chi^2(mp) \tag{B2.13}$$

where $\tau_1 = tr(\mathbf{G}) = \Sigma_{i=1}^{m}\lambda_i$ is the Hotelling-Lawley trace criterion (Lawley (1938) Hotelling (1931)), and the characteristic equation of **G** is

$$|\,\mathbf{HS}^{-1} - \lambda\mathbf{I}\,| = |\,\mathbf{H} - \lambda\mathbf{S}\,| = 0. \tag{B2.14}$$

As shown by Anderson (1984), the exact distribution of τ_1 is that of a product of independent Beta random variables, and the critical values of this distribution are given in Kres (1983).

The *likelihood ratio* criterion (Wilks (1932)) is,

$$\tau_2 = [L(\tilde{\theta})/L(\hat{\theta})]^{2/T} = |(\hat{U}'\hat{U})(\tilde{U}'\tilde{U})^{-1}| = |\mathbf{S}|\,|\mathbf{S}+\mathbf{H}|^{-1} = \Sigma_{i=1}^{m}\ln\mu_i, \tag{B2.15}$$

where $\mu_i = 1/(1+\lambda_i)$ are the roots of the characteristic equation

$$|\,\mathbf{S}(\mathbf{S}+\mathbf{H})^{-1} - \mu\mathbf{I}\,| = |\,\mathbf{S} - \mu(\mathbf{S}+\mathbf{H})\,| = 0. \tag{B2.16}$$

The Bartlett (1947) modified LR statistic is

$$LR = -T^{*}\ln(\tau_2) = -T^{*}\ln(\,|(\hat{U}'\hat{U})(\tilde{U}'\tilde{U})^{-1}|\,) \sim \chi^2(mp), \tag{B2.17}$$

which is asymptotically distributed chi-square under H_0, and $T^{*} = T\text{-}k\text{-}.5(m\text{-}p+1)$, $p \geq m$. The asymptotic expansion of LR is given by Box (1949), where it is shown that the first term of this expansion satisfies $C(\alpha)\chi^2(\alpha) = LR$, where $C(\alpha)$ is the α-level error approximation and is tabulated in Anderson (1984). Anderson shows that $C(\alpha)$ is close to unity for $m^2+p^2 \leq T^{*}/3$, and that T^{*} is a better finite-sample adjustment than (T-k).

The asymptotic chi-square *Lagrange Multiplier* statistic is,

$$LM = (T\text{-}k)\tau_3 \sim \chi^2(mp), \tag{B2.18}$$

where $\tau_3 = \mathrm{tr}[(\mathbf{S} + \mathbf{H})^{-1}\mathbf{H}] = \mathrm{tr}[(\hat{\mathbf{U}}'\hat{\mathbf{U}} - \tilde{\mathbf{U}}'\tilde{\mathbf{U}})(\tilde{\mathbf{U}}'\tilde{\mathbf{U}})^{-1}]$. This statistic was proposed by Pillai (1955), and is referred to as Pillai's trace criterion. τ_3 can be written in terms of the roots of the equation,

$$|\ (\mathbf{S}+\mathbf{H})^{-1}\mathbf{H} - \theta\mathbf{I}\ | = |\ \mathbf{H} - \theta(\mathbf{S}+\mathbf{H})\ | = 0, \tag{B2.19}$$

where $\theta_i = 1 - \mu_i$. The exact distribution of τ_3 in special cases is derived by Nanda (1950). The asymptotic expansion of LM in terms of χ^2 distribution is given in Muirhead (1982), which is used to derive the $\chi^2(mp)$ limiting distribution.

The relationship among these test criteria can be summarized in terms of eigenvalues as,

$$\lambda_i = \theta_i/(1\text{-}\theta_i) = (1\text{-}\mu_k)/\mu_k,\ \ \theta_i = \lambda_i/(1+\lambda_i) = 1 - \mu_k,\ \ \mu_k = 1/(1+\lambda_i) = 1\text{-}\theta_i, \tag{B2.20}$$

(Kres (1983)). The relationship in (B2.21) is discussed in detail in Berndt and Savin (1977) who show that $\tau_1 \geq \ln\tau_2 \geq \tau_3$, because $x \geq \ln(1+x) \geq x/(1+x) \geq 0$, which corresponds to the well-known inequality $W \geq LR \geq LM$.[24] This inequality raises the possibility of conflict between the various criteria. Note also that as each non-zero root approaches infinity the LM test is bounded by (T-k)m, so the unadjusted tests will reject the null hypothesis for some critical value. Hence some caution is suggested in interpreting these statistics in the extreme case where (T-k)m is close to the critical value.

There are a number of exact results for these statistics in special cases. Anderson (1984), shows that when $m = 1$, $(T\text{-}k)[(1\text{-}\tau_2)/\tau_2]/p \sim F(p,T\text{-}k)$, and when $p = 1$ we obtain the Hotelling T^2 statistic (Lawley (1938), Hotelling (1931)), $[(1\text{-}\tau_2)/\tau_2](T\text{-}k\text{-}m+1)/m \sim F(m,T\text{-}k\text{-}m+1)$. The

T^2 statistic arises, say, when testing that the intercept terms are all zero. When $m = 2$, we have $[(1-\tau_2^{1/2})/\tau_2^{1/2}](T-k-1)/p \sim F(2p,2(T-k-1))$, and for $p = 2$, $[(1-\tau_2^{1/2})/\tau_2^{1/2}] \sim F(2m,2(T-k-m+1))$. Rao (1951) suggests a general approximation as $[(1-\tau_2^{1/s})/\tau_2^{1/s}](T^*s-r)/pm \sim F(pm,T^*s-r)$, where $s = [(p^2 \cdot m^2-4)/(p^2+m^2-5)]^{1/2}$, $r = m \cdot p/2 - 1$. This approximation is more accurate than the χ^2 versions for small T (Anderson (1984)).[25]

Except when the maximal root of **G** is substantially larger than the others, the trace and determinant tests tend to have good power (Schatzoff (1966)). This is because they use all of the roots of **G**, i.e., all distinct linear combinations to "explain" the normalized variation. The power of the tests for linear hypotheses is examined in Giri (1977), Hart and Money (1976), Olson (1974), and Schatzoff (1966) using Monte Carlo experimentation. In general, the LM test is preferred if the roots are approximately equal under the alternative, and W is preferred if they are substantially different. The LR test is intermediate in either case.

Appendix 2C

Univariate Normality tests

This Appendix describes a number of univariate normality tests that are directly related to the multivariate tests described in Section 2.4.3.

The most commonly employed measures of distributional shape in the univariate case are the standardized third and fourth central moments:

$$\text{(i) } \textit{skewness}: \alpha_3 = \mu_3/\sigma^3, \quad \text{(ii) } \textit{kurtosis}: \alpha_4 = \mu_4/\sigma^4, \tag{C2.1}$$

where $\mu_r = E[u_t^r]$. This popularity follows from the well known result that the Pearson family of univariate distributions is characterized by their first four moments (see Kendall et.al., (1987)). In particular, normality is defined by $\alpha_3 = 0$ and $\alpha_4 = 3$. The skewness and kurtosis coefficients provide one measure of asymmetry, and the relative peakedness of the distribution respectively, and are both invariant to location and scale changes. Note however, Ord (1968) shows that there exist asymmetrical distributions for which $\alpha_3=0$, and Balanda and MacGillivray (1988) present a number of highly non-normal (but symmetric) distributions for which $\alpha_4 = 3$. In the statistics literature, Pearson (1965) examines approximations to the distribution of sample estimates of the standardized moments, whose limiting behavior is obtained by Pearson (1930). It is well known that the distribution of $\hat{\alpha}_3$ for NIID samples is symmetric, leptokurtic and converges to normality reasonably quickly (for T > 100), and that of $\hat{\alpha}_4$ is skewed and converges to normality very slowly (even for T > 1000). Moreover, following Fisher (1930), Shenton and Bowman (1975) have shown that $\hat{\alpha}_3$ and $\hat{\alpha}_4$ have non-zero covariance. Hence, the directional asymptotic test statistics,

$$S_1 = [(T/6)\hat{\alpha}_3^2]^{1/2} \sim N(0,1) \text{ and } K_1 = (T/24)^{1/2}(\hat{\alpha}_4\text{-}3) \sim N(0,1), \qquad (C2.2)$$

or the joint test statistic

$$SK_1 = S_1^2 + K_1^2 \sim \chi^2(2). \qquad (C2.3)$$

These tests are very popular in the econometrics literature where they have been formulated as LM tests within the Pearson family (Jarque and Bera (1987), Bera and Jarque (1982)), as LM tests against the Stable family of distributions (Bera and McKenzie (1983)), or via a Gram-Chalier expansion (Kiefer and Salmon (1983)).

As an alternative approach, D'Agostino (1970), D'Agostino and Pearson (1973,1974), and Pearson, D'Agostino and Bowman (1977) obtain a good normalizing transformation for $\hat{\alpha}_3$ based on Johnson's (1949) S_U normalizing transformation, and Anscombe and Glynn (1983) utilize the Wilson and Hilferty (1937) normalizing transformation for a χ^2 random variable to obtain good approximations for $\hat{\alpha}_4$ for $T > 30$. The D'Agostino-Pearson tests for normality as described in D'Agostino and Stephens (1986), and D'Agostino, Belanger and D'Agostino (1990), are based on these two normalizing transformations. The approximate skewness test statistic take the form,

$$S_2 = \delta \ln[Y/\gamma + (1 + (Y/\gamma)^2)^{1/2}] \sim N(0,1), \qquad (C2.4)$$

where $Y = \hat{\alpha}_3[(T+1)(T+3)/(6(T\text{-}2))]^{1/2}$, $\delta = \ln(W)^{-1/2}$, $\gamma = (2/(W^2\text{-}1))^{1/2}$, $W^2 = 2\alpha_4(\hat{\alpha}_3)\text{-}1)^{1/2}\text{-}1$, and

$\alpha_4(\hat{\alpha}_3) = \dfrac{3(T^2+27)T-70)(T+1)(T+3)}{(T-2)(T+5)(T+7)(T+9)}$. The kurtosis test statistic is

$$K_2 = \left[\left[\frac{1-\frac{2}{\nu}}{1+X\sqrt{\frac{2}{\nu-4}}}\right]^{\frac{1}{3}} - \left[1-\frac{2}{9\nu}\right]\right]\left[\frac{2}{9\nu}\right]^{-\frac{1}{2}} \sim N(0,1) \qquad (C2.5)$$

where $X = (T/24)^{1/2}(\hat{\alpha}_4 - 3)$, $\quad \nu = 6 + \frac{8}{\alpha_3(\hat{\alpha}_4)}\left[\frac{2}{\alpha_3(\hat{\alpha}_4)} + \left[1 + \frac{4}{\alpha_3^2(\hat{\alpha}_4)}\right]^{\frac{1}{2}}\right]$, and (C2.6)

$$\alpha_3(\hat{\alpha}_4) = \frac{6(T^2 - 5T + 2)}{(T+7)(T+9)}\sqrt{\frac{6(T+3)(T+5)}{T(T-2)(T-3)}}\ .$$

Combining S_2 and K_2 we obtain the joint statistic,

$$SK_2 = S_2^2 + K_2^2 \sim \chi^2(2), \tag{C2.7}$$

where S_2 and K_2 are uncorrelated and nearly independent.

The adjustments leading to S_2 and K_2 can also be applied to the multivariate normality test of Small (1980), as described in Section 2.4.3.

To examine the finite sample performance of these various tests I computed the test statistics for 5000 replications of random samples of standard normal pseudo-random numbers. The results are reported below in Table 1 for several sample sizes. We can see that there is a very close correspondence between the actual and nominal size for the D'Agostino-Pearson tests (S_2,K_2,SK_2), while the actual size of the Jarque-Bera tests (S_1,K_1,SK) tend to be considerably smaller than the nominal size, suggesting that the test will tend to reject the null hypothesis of normality too often in all but very large samples.[26] The sample mean, variance, skewness, kurtosis and correlations ($\rho(S_i,K_i)$, i = 1,2) for the 5000 replications of the tests are reported in Table 2. These results show the relative accuracy of the normalizing transformation for the D'Agostino-Pearson tests, compared with that of the Jarque-Bera tests. In particular, the D'Agostino-Pearson test statistics have approximately zero mean, unit variance, zero skewness, zero excess kurtosis, and are almost uncorrelated for all the sample sizes considered.

Table 2.1 - Actual Size of Univariate Normality Tests (5000 replications)

T	S_1	S_2	K_1	K_2	SK_1	SK_2
30 $\alpha=0.10$	0.076	0.108	0.028	0.106	0.050	0.103
50	0.082	0.102	0.039	0.103	0.054	0.095
100	0.088	0.099	0.056	0.105	0.069	0.105
200	0.094	0.098	0.067	0.098	0.071	0.098
500	0.098	0.100	0.091	0.103	0.088	0.100
30 $\alpha=0.05$	0.034	0.053	0.019	0.053	0.033	0.059
50	0.039	0.048	0.024	0.054	0.038	0.056
100	0.048	0.052	0.030	0.052	0.041	0.053
200	0.046	0.048	0.032	0.047	0.041	0.048
500	0.049	0.050	0.044	0.051	0.048	0.052

Table 2.2 - Normality Test Summary Statistics

	S_1	S_2	K_1	K_2
Mean, T = 30	0.022	0.024	-0.222	-0.011
Variance	0.834	1.016	0.615	1.016
Skewness	0.039	0.018	1.644	-0.012
Kurtosis	3.438	2.943	8.173	2.962
$\rho(S_i,K_i)$	0.032	0.015	0.032	0.015
Mean, T = 50	-0.009	-0.009	-0.175	-0.008
Variance	0.867	0.977	0.729	1.024
Skewness	-0.028	-0.018	1.422	-0.111
Kurtosis	3.477	3.056	6.652	3.110
$\rho(S_i,K_i)$	0.033	0.019	0.033	0.019
Mean, T = 100	0.009	0.009	-0.135	-0.021
Variance	0.937	0.996	0.858	1.050
Skewness	0.051	0.042	1.163	-0.155
Kurtosis	3.217	2.965	5.691	3.163
$\rho(S_i,K_i)$	0.018	0.009	0.018	0.009
Mean, T = 200	-0.016	-0.016	-0.099	-0.008
Variance	0.951	0.981	0.883	0.960
Skewness	0.034	0.029	1.017	-0.029
Kurtosis	3.148	2.997	5.330	3.165
$\rho(S_i,K_i)$	0.011	0.007	0.011	0.007
Mean, T = 500	0.010	0.010	-0.098	0.001
Variance	0.981	0.992	1.010	1.046
Skewness	-0.015	-0.015	0.648	-0.071
Kurtosis	3.081	3.016	4.202	3.181
$\rho(S_i,K_i)$	0.002	0.001	0.002	0.001

Notes

1. See also Pagan and Hall (1983), Pagan (1984), Godfrey (1988) for similar statements. Interestingly, the assumption of zero skewness and excess kurtosis used by Pagan and Vella (1988) to characterize normality is a little misleading. There exist asymmetrical distributions for which $E(u_t^3) = 0$ (Ord (1968)), and non-normal distributions for which $E(u_t^4 - 3\sigma^4) = 0$ (Balanda and MacGillivray (1988)).

2. The assumption of $O(T^{1/2})$ consistency is convenient for establishing standard $T^{1/2}$ asymptotic normality results, but will be invalid when the regressors are trending or elements of $\mathbf{w}_t$ include higher order powers. Anderson (1971) has provided more general conditions for asymptotic normality.

3. The asymptotic theory for the moment tests is due to Pierce (1982), Newey (1985, 1985a), and Tauchen (1985). Applications of moment tests include testing for heteroskedasticity (Pagan and Pak (1991), Nelson (1991)), and testing for misspecification in limited dependent and censored sample models (Pagan and Vella (1989), Cameron and Trivedi (1990)).

4. This method of approximation is referred to as the δ-method and relies on the mean of higher-order derivatives converging to a constant and the assumption that $\hat{\boldsymbol{\theta}}$ is $O_p(T^{1/2})$ consistent. In that case, $\mathbf{B}$ converges in probability to the possibly non-zero constant vector $E[\partial \mathbf{m}_t(u_t,w_t;\boldsymbol{\theta})/\partial\boldsymbol{\theta}]$ (see Pagan and Vella (1989)).

5. As shown in Appendix 2A, inferences based on consistent estimators do not generally lead to more precise or reliable evidence than those based on ML estimators. This is because there are no general finite sample results for MM estimators and the question of full efficiency can not be addressed.

6. It is interesting to note that the use of the auxiliary regression form requires an adequate specification of the distribution to ensure $-E[\mathbf{H}(\boldsymbol{\theta})] = E[\mathbf{s}(\boldsymbol{\theta})\mathbf{s}(\boldsymbol{\theta})']$, $E[\partial\mathbf{m}(\boldsymbol{\theta})/\boldsymbol{\theta}] = -E[\mathbf{m}(\boldsymbol{\theta})\mathbf{s}(\boldsymbol{\theta})']$ (the generalized information inequality of Beran (1977)). This tends to go against the general desire in this literature for a minimal assumption procedure.

7. For dependent processes, modifications along the lines of Newey and West (1987), Andrews (1991) have been suggested to account for a non-zero covariance between each vector over time (see Appendix 2B).

8. Note, although the columns of $\mathbf{S}_\beta$ are orthogonal to $\boldsymbol{\iota}$ by definition under the null, exclusion of them will lead to a statistic that is too small except when the information matrix is block diagonal between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

9. $\boldsymbol{\Sigma}_{11} = \mathrm{Cov}(\mathbf{y}_t)$ (mxm), $\boldsymbol{\Sigma}_{12} = \mathrm{Cov}(\mathbf{y}_t,\mathbf{X}_t)$ (mxk), and $\boldsymbol{\Sigma}_{22} = \mathrm{Cov}(\mathbf{X}_t)$ (kxk).

10. Possible choices of $\mathfrak{F}_t$ include (a) the trivial σ-field $\{S,\varnothing\}$; (b) the degenerate σ-field $\{\mathbf{X}_t=\mathbf{x}_t\}$; (c) the σ-field generated by the random vector $\mathbf{X}_t$, $\sigma(\mathbf{X}_t)$; (d) the σ-field generated by the past history of y, $\sigma(y_{t-i},\ i = 1,\ldots)$; (e) the σ-field $\{\sigma(y_{t-i}),\ \mathbf{X}_{t-j}=\mathbf{x}_{t-j}\}$, $i = 1,\ldots.,\ j = 0,1,\ldots$; and (e) $\{\sigma(y_{t-i}),\ \mathbf{X}_{t-j}=\mathbf{x}_{t-j}\}$, $i = 1,\ldots.,\ j = 1,\ldots$. The choice of (a) leads to the unconditional mean representation; (b) the static regression representation; (c) the stochastic regression representation; (d) the AutoRegressive (AR) representation; (e) Dynamic regression representation; and (f) the Vector AutoRegressive representation of y_t.

11. The importance of considering this problem simultaneously has been demonstrated in Spanos (1991) for the Student's t linear regression model, where non-correlation and independence do not coincide. This leads to a conditional mean that is static and a conditional variance that is dynamic, and the appropriate conditioning information set is $\{\sigma(y_{t-i}),\mathbf{X}_{t-j}=\mathbf{x}_{t-j}, i = 1,\ldots,\ j = 0,1,\ldots\}$.

12. In a sense, the only observable information on the latter is provided by the residuals, which are an estimated realization from the conditional distribution in mean deviation form. Moreover, different theories imply different conditional models, and the interrelationship among these is not always apparent. Placing competing theoretical models on a common statistical basis is discussed in Chapter 3.

13. Note, this restriction does not hold in general for non-normal distributions. For example, for the Student's t linear regression model (Spanos (1991)), the weak exogeneity assumption does not hold because Σ_{22} is a statistical parameter of interest in the conditional distribution.

14. Pagan and Hall (1983), Pagan (1984) and MacKinnon (1992) provide related discussions. However, in these studies no distinction between $\mathbf{B}$ and $\mathbf{B}_0$ is made, and the presence of $\mathbf{x}_t$ in the auxiliary regression is motivated by concerns of obtaining the "correct" covariance for α_0. In the present case (2.25) is an alternative conditioning, and the term $(\mathbf{B}_0\text{-}\hat{\mathbf{B}})'\mathbf{x}_t$ arises by construction.

15. Wooldridge (1990, 1991) has proposed using "robust" auxiliary regression tests. For example, White's (1980) heteroskedastic consistent covariance estimator is often suggested in formulating test statistics for the conditional mean (Pagan and Hall (1983)). However, robustness is only a meaningful concept when defined relative to the probabilistic features of the data. For example, observations that may be outliers for the normal distribution, may be quite "normal" for the Student's t distribution. Moreover, "robust" adjustments are not a substitute for parameterizing the source of the departure when an appropriate form exists. For example, Spanos (1992) compares the Normal with the Student's t dynamic model, and shows that using heteroskedastic consistent standard errors led to quite different inferences than those of the statistically adequate parametric alternative.

16. See Davis (1975) and Rivlin (1969) for a discussion of LS approximation theory in square integrable Hilbert space in the context of general functional approximation theory.

17. There may exist cases where a non-linear conditional mean can arise from an alternative joint distribution without violating conditional normality. However, this author is unaware of any such cases.

18. Other graphical tools are available for assessing parameter t-invariance. See Spanos (1986) for a discussion and Section 2.5 for a brief description of the implementation of these techniques in SAM.

19. In the single equation context, this regression gives rise to the so-called modified LM test (Durbin (1970)).

20. SAM was originated in the summer of 1990 by Anya McGuirk and Aris Spanos. I then wrote the main sub-routines and established the program design for SAM over the following year. SAM is continually undergoing revision by all three authors, and new procedures are written as required depending on the particular empirical problem in hand. Thanks are due to Christoph Hinkleman for his initial assistance, and Mico Loretan for use of his fast Fourier transform density estimation algorithm. SAM has been used by the authors in a number empirical studies; in developing the exercises in Spanos (1992b); and by graduate students from the Departments of Economics and Agricultural Economics at Virginia Tech.

21. Related stand-alone programs have been written for estimation and testing of the dynamic Student's t models of Spanos (1992b), for cointegration analysis, and for random number generation (RANDOM), and for estimating linear models subject to error autocorrelation.

22. See Spanos (1986) for an extensive discussion of the single equation auxiliary regression tests.

23. This is a special case of the general class of uniform mixed linear constraints of the form $\mathbf{DBP} = \mathbf{C}$ (see Berndt and Savin (1977)). In the present case $\mathbf{P}$ is an mxm identity matrix and $\mathbf{C} = \mathbf{0}$, implying that the same set of zero restrictions is being applied to each equation: $\mathbf{DB}_i - \mathbf{C}_i = \mathbf{0}$, $i = 1,\dots,m$ (see Spanos (1986, p.580)). A

more general formulation is $\mathbf{R}\mathbf{B}^* = \mathbf{r}$, where $\mathbf{B}^* = \text{vec}(\mathbf{B})$. This allows the imposition of separate within equation, and cross-equation restrictions by reaching each coefficient in **B** directly. This is particularly useful when using the SUR formulation (see Spanos (1986, p.584)).

24. In the case where $\min(m,p) = 1$, there is only one non-zero root, so the W, LR, and LM tests are related via $W = (T\text{-}k)\lambda$, $LR = T^{*}\ln(1+\lambda)$, and $LM = (T\text{-}k)\lambda/(1+\lambda)$.

25. Alternative formulations of the multivariate LM and Wald tests arise by considering the determinants rather than the trace.

26. The tendency for the finite sample p-value to be much smaller than its asymptotic size has been noted by Jarque and Bera (1980) in Monte Carlo experiments.

Chapter 3

Modeling Wages and Unemployment in a Small Dynamic System

3.1 Introduction

This chapter describes the modeling of inflation/unemployment tradeoffs in the U.S. An adequate statistical model (unrestricted reduced-form) for the variables suggested by theory is estimated. This model is then used as a basis for estimating a number of competing structural models and testing the over-identifying restrictions implied by economic theory.

The attitude taken in this paper is that the empirical validity of any structural inferences depends crucially on the appropriateness of the underlying statistical model chosen for the data. Thus, careful statistical model specification and misspecification analysis are emphasized, and a number of the procedures described in Chapter 2 can be utilized.

In the next section I discuss the modeling approach in more detail, and in Section 3.3, I briefly summarize several alternative structural models in order to motivate the econometric analysis that follows. In Section 3.4, the probabilistic features of the data are described. These provide a guide in specifying a common statistical model for the data which can be related to the structural models of interest via reparameterization/restriction.[1] A dynamic model for wages and prices, using unemployment, labor productivity, and interest rates as additional conditioning series is estimated, and multivariate misspecification testing and graphical analysis are used in checking for statistical adequacy. In Section 3.5, I estimate a number of structural models by instrumental variable methods, using all the regressors in the statistical model as instruments. The identifying restrictions of the estimated structural models, and their statistical adequacy are tested. The conclusions are presented in Section 3.6.

3.2 Statistical Models and Identification

The structural model implied by economic theory can often be represented in terms of the classical Simultaneous Equations Model (SEM),

$$\mathbf{\Gamma}'\mathbf{y}_t + \mathbf{D}'\mathbf{X}_t = \epsilon_t, \qquad \epsilon_t \,|\, \mathbf{X}_t \sim \text{NIID}(\mathbf{0},\mathbf{\Sigma}), \quad t\in\mathbf{N}, \tag{3.1}$$

where $\mathbf{y}_t$ is an mx1 vector of endogenous variables, $\mathbf{X}_t$ is a kx1 vector of exogenous and otherwise predetermined variables, and ϵ_t is the mx1 structural error vector. Assuming that $\mathbf{\Gamma}'$ is non-singular, the reduced-form implied by (3.1) is,

$$\mathbf{y}_t = \mathbf{\Pi}'\mathbf{X}_t + \mathbf{u}_t, \qquad \mathbf{u}_t \,|\, \mathbf{X}_t \sim \text{NIID}(\mathbf{0},\mathbf{\Lambda}), \quad t\in\mathbf{N}, \tag{3.2}$$

where the parameters of (3.1) and (3.2) are related via $\mathbf{\Pi\Gamma} + \mathbf{D} = \mathbf{0}$, and $\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda\Gamma}$ (see Hsiao (1983)).

For present purposes, the formulation of a reduced-form for (3.1) is approached in more general terms. The reduced-form is viewed as a *statistical model* that is postulated in terms of the data as a convenient summarization of the sample information that can be related to the structural model via reparameterization/restriction. In other words, theory suggests the relevant variables and the general form of the statistical model, and its final form is chosen so as to adequately account for the probabilistic features of the data. There are several advantages to viewing structural modeling from this perspective:

(a) The statistical adequacy of the reduced-form is an important pre-requisite for estimation of the structural parameters and hypothesis testing. The statistical model provides a convenient framework for misspecification analysis, as described in Chapter 2.

(b) Since the statistical model does not have any direct theoretical meaning, respecification of the statistical model does not involve changing the structural model.

(c) Since many structural models impose restrictions on the dynamics of the data there is a rich source of identifying restrictions in terms of a dynamic statistical model.

(d) The statistical model provides a link between the structural form and the data, thereby relating the traditional SEM approach with multivariate time series analysis, and the literature on model encompassing (see Spanos (1990)).

In Chapter 2 the Multivariate Linear Regression (MLR) model was discussed as a basic statistical model, specified in terms of the first two moments of the conditional normal distribution $D(\mathbf{y}_t \mid \mathbf{X}_t; \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0 = (\mathbf{B}, \boldsymbol{\Omega}_0)$ are the statistical parameters of interest. For economic time series, this model is generally insufficient to capture the non-stationarity and dynamics that are an inherent feature of such data. To account for the dynamics one might consider the dynamic MLR (MDLR) model,

$$\mathbf{y}_t = \boldsymbol{B}_0' \boldsymbol{x}_t + \sum_{i=1}^{p} [\boldsymbol{C}_i' \boldsymbol{y}_{t-i} + \boldsymbol{B}_i' \boldsymbol{x}_{t-i}] + \boldsymbol{v}_t, \qquad \boldsymbol{v}_t \mid \boldsymbol{X}_t, \boldsymbol{Z}_{t-1}^{o} \sim N(\mathbf{0}, \boldsymbol{\Omega}_1), \tag{3.3}$$

with statistical parameters $\boldsymbol{\theta}_1 \equiv (\mathbf{B}, \mathbf{C}, \boldsymbol{\Omega}_1)$, or the Vector AutoRegressive (VAR) model

$$\mathbf{Z}_t = \sum_{i=1}^{p} \boldsymbol{A}_i' \boldsymbol{Z}_{t-i} + \boldsymbol{V}_t, \qquad \boldsymbol{V}_t \mid \boldsymbol{Z}_{t-1}^{o} \sim N(\mathbf{0}, \boldsymbol{\Omega}_2), \tag{3.4}$$

with parameters $\boldsymbol{\theta}_2 \equiv (\mathbf{A}, \boldsymbol{\Omega}_2)$, where $\mathbf{Z}_t \equiv (\mathbf{y}_t', \mathbf{X}_t')$, and $\mathbf{Z}_{t-1}^{o} = (\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \ldots, \mathbf{Z}_1)$.

These models arise as natural extensions of the MLR model via alternative reductions of the underlying Haavelmo distribution under the assumptions of joint normality, stationarity, and asymptotic independence (see Spanos (1986)). The basic difference between the VAR and MDLR model is that the VAR has np + ½n(n+1)) parameters (excluding constant terms, n = m+k), while the MDLR model has (m+p)k + mp + ½m(m+1) parameters, and for forecasting purposes we need to impose Granger non-causality from $\mathbf{y}_t$ to $\mathbf{X}_t$ on the latter (see Engle et.al., (1983)).[2] Another possible statistical model in this context is the partial VAR

$$\mathbf{y}_t = \sum_{i=1}^{p} [D_i' y_{t-i} + G_i' x_{t-i}] + e_t, \qquad e_t | Z_{t-1}^{o} \sim N(0, \Omega_3), \tag{3.5}$$

where $\theta_3 \equiv (\mathbf{D},\mathbf{G},\Omega_3)$. This model may be useful when the informational demands of the VAR and MDLR models on the available sample information are excessive.[3]

These multivariate dynamic regression-type models are generally adequate for modeling data that can be viewed as realizations of a jointly normal, weakly dependent process, for which the non-stationarity can be accounted for by modifying the "constant terms" to include time trends, dummy variables etc., or by imposing parametric differencing restrictions (see Spanos (1992a)). The logarithmic transformation may also be used to stabilize the variability of a series without necessarily sacrificing the interpretation of the structural models, since many theoretical models relate to growth rates, relative prices etc.

As will be demonstrated in Chapter 4, failure to adequately account for the probabilistic features of the data can have important implications on the reliability of inferences from the model. Hence, a careful preliminary data analysis is imperative before specifying and estimating the statistical model(s). The graphical tools in SAM can be utilized to help judge the appropriateness of the modeling assumptions by examining the relationship between the statistical model specification assumptions and the underlying Haavelmo distribution reduction.

Once the statistical model is estimated, residual plots and misspecification testing can be used to check the adequacy of the estimated form. If the statistical model is misspecified there is little point in imposing further restrictions on it, since tests thereof will be against an invalid baseline. In that case, the statistical model must be respecified to account for the misspecifications. Possible respecifications can arise by tracing the symptoms of misspecification to their source in terms of the underlying probabilistic structure of the data and considering an alternative Haavelmo reduction (see Spanos (1992) for a discussion). Interestingly, the question of "data-mining" does not arise because the statistical model is chosen to be a summary of the probabilistic features of the data, as demarcated by the Haavelmo (joint) distribution, and these features are invariant to the particular model specification used.

To proceed from the well-specified statistical model for the data to the structural form involves changing the statistical parameterization into a theoretically meaningful one by imposing restrictions. Identification of the structural parameters can be viewed as a mapping H(.) from the statistical parameter space to the structural parameter space A via $H(\cdot)$: $\theta \rightarrow A$. The structural parameters are said to be identified under $H(\cdot)$ if a unique solution for $\alpha \in A$ exists. If G denotes the mapping from θ_2 to θ_1, then the parameters of A are also identified via a composite mapping $G(H(\cdot))$: $\theta_2 \rightarrow A$. A rational expectations model can also be related via a composite mapping $C(D(\cdot))$: $\theta \rightarrow A$, where $C(\cdot)$: $\theta \rightarrow \phi$, and $D(\cdot)$: $\phi \rightarrow A$, and ϕ is the parameter space of the mechanism defining the expectations (see Appendix 3A).[4]

Estimation of the parameters of the structural model poses the question: "given the statistical parameters what are the implications for the structural parameters subject to the identifying restrictions?". The class of Instrumental Variable (IV) estimators such as two and three-stage least squares are applicable in this case (see Bowden and Turkington (1984)), where the appropriate instruments are defined by all the conditioning variables (exogenous and predetermined) from the underlying statistical model.[5]

Testing the a priori structural over-identifying restrictions in terms of a well-specified statistical model provides a means for assessing the empirical validity of the structural implications of competing models. When the identifying restrictions are of the linear homogenous (exclusion) form, the identification status of each structural equation is determined by the number of excluded exogenous variables ($k-k_1$), relative to the total number of included endogenous variables in the equation (m_1-1). If $k-k_1 > m_1-1$, the structural equation is over-identified; if $k-k_1 = m_1-1$ the equation is just-identified (i.e., no restrictions are imposed on the statistical parameters); and if $k-k_1 < m_1-1$, the equation is not identified.[6]

When a structural model is estimated by IV, a Lagrange Multiplier (LM) test for over-identifying restrictions arises by comparing the parameterization of the restricted and unrestricted (i.e., including all the exogenous variables) structural equations in terms of the restrictions imposed on the corresponding equation in the statistical model.[7] In other words, by reversing the role of the statistical and structural parameters from estimation, a test of over-identifying restrictions poses the question: "given the restricted structural parameters, what are the implications for the statistical parameterization?". The LM test was initially proposed by Anderson and Rubin (1949) (see Phillips (1983)). Denoting $\mathbf{X}$ as the Txk matrix of "exogenous" variables in the statistical model; $\mathbf{X}_1$ as the Txk_1 ($k_1 < k$) matrix of "exogenous" variables included in the restricted structural equation; $\mathbf{Y}_1$ as the Txm_1 matrix of included endogenous variables in the structural equation for $\mathbf{y}_1$; the test criteria for one equation in the structural model is,

$$\tau_1 = \frac{(y_1 - Y_1\hat{\gamma}_{IV})'(M_{X1} - M_X)(y_1 - Y_1\hat{\gamma}_{IV})}{(y_1 - Y_1\hat{\gamma}_{IV})'M_X(y_1 - Y_1\hat{\gamma}_{IV})} \tag{3.6}$$

where $\mathbf{M_X} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{M}_{X1} = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$, are orthogonal projection matrices, and IV denotes an instrumental variables estimator.[8] Anderson and Rubin show that $T\tau_1$ was asymptotically distributed as $\chi^2(q)$, where $q = k-k_1-m_1+1$ is the number of over-identifying restrictions. Basmann (1960) shows that $(T-k)\tau_1/q$ was approximately distributed $F(q,T-k)$ and was a better approximation in finite samples.

These test statistics can be modified for testing other restrictions such $\gamma_1 = \bar{\gamma}_1$ by replacing $\hat{\gamma}_{IV}$ with $\bar{\gamma}_1$. The resulting F-type test is approximately distributed as $F(k_2,T-k)$, where k_2 is the number of restrictions on γ_1. This can also be used for testing parameters on the included exogenous variables as long as all of the coefficients of γ_1 are also specified under the null (see Spanos (1986, p.650)).[9]

In the event that more than one structural model is found to be adequate, in terms of both the over-identifying restrictions and its own statistical adequacy, a number of criteria have been suggested for discriminating among the models (see Hendry and Richard (1982)). On the other hand, if none of the structural models survive to this stage it may be necessary to modify the structural models so as to attempt to account for the misspecifications. This "patch-work" approach may be necessary when the theoretical model does not coincide with the estimable form of the model for the data in hand. This suggests that the data can still have an important role in determining some of the features of the structural model, such as the dynamics (Spanos (1986), Hendry et.al., (1988)).

From this discussion we can see that the statistical and structural models are complements. The statistical model summarizes the probabilistic features of the data (the distributional, memory and heterogeneity), while the theory suggests the underlying structural model and the variables of interest. The structural model(s) can then be tested against the statistical model by means of specification tests.

In the next section, a number of theoretical models of the aggregate labor market are described. Several of these models will then form the focus of the econometric analysis that follows.

3.3 Theoretical Models of Wages, Prices and Unemployment

The original work on the wage, price, and unemployment nexus is attributed to Phillips (1958) (see Desai (1984)).[10] Phillips postulated that the rate of growth in nominal wages depends on:

(a) the level of excess demand of labor, as measured by the unemployment rate;

(b) the rate of change in unemployment, which reflects the business cycle effects on unemployment (demand for labor) and wages; and

(c) the cost-of-living adjustments in negotiating wage contracts based on changes in prices.

To represent the behavior of prices, or the implicit supply curve of firms, it is typically assumed that prices are determined as a constant markup of the value of output over the wage bill, with wages deflated by the growth in labor productivity, and by the difference between actual and capacity output (using the unemployment rate as a proxy).[11] Raw material and import prices are also included in some formulations of the price equation (see Sargan (1964)).

A stylized version of this basic wage-price mechanism takes the form,

$$\begin{aligned} \Delta w_t &= \alpha \Delta p_t + g(u_t, \Delta u_t), \qquad \frac{\partial g}{\partial u_t} \le 0, \quad \frac{\partial g}{\partial \Delta u_t} \le 0, \\ \Delta p_t &= \Delta w_t - \Delta q_t + f(u_t), \qquad \frac{\partial f}{\partial u_t} \le 0, \end{aligned} \tag{3.7}$$

where p_t, w_t, and q_t are the logarithms of the price level, nominal wage, average labor productivity respectively, and u_t is the unemployment rate.[12] Phillips postulated an inverse relationship between the rate of unemployment and wage inflation, and thus on substitution, the general level of price inflation (the Phillips curve).[13]

Lipsey (1960) interpreted the wage equation as a nominal wage adjustment mechanism, where wages changed proportionally to the gap between the supply and demand for labor relative to the total labor supply.[14] Lipsey argued that it takes time to change jobs, so the wage equation would represent the speed at which nominal wage rates adjust to a given disequilibrium (see also Lipsey and Parkin (1970)).

Phelps (1967), Friedman (1966, 1968), among others, have argued that it is unanticipated inflation rather than observed inflation that is important in determining wages because of the uncertainty facing workers in forming contracts for the following year, and in searching for employment. Friedman (1966, pp.58-60) summarized the criticism of the traditional Phillips curve model as follows:

> "*The basic fallacy is to suppose that ... by inflating more over any long period of time, you can have on average a lower level of unemployment .... By speeding up the rate of monetary expansion and aggregate demand, you can increase output and employment temporarily ... but only until people adjust their anticipations.*"

In this context, assuming that $\alpha = 1$, $\Delta q_t = a$, and replacing p_t with $\hat{p}_t$, the "rational expectation" of inflation given t-1 information (see Appendix 3A), we can combine this expectations-augmented wage equation with the price equation in (3.7) to yield

$$\Delta p_t = \Delta \hat{p}_t - a - f(u_t) + g(u_t, \Delta u_t). \tag{3.8}$$

This characterization of price inflation is now an integral component of most standard IS-LM models (see Sargent (1987), Blanchard and Fischer (1989)). Equation (3.8) implies that movements along the Phillips curve only correspond to unexpected inflation, while the whole curve will shift up by the full amount of anticipated inflation ($\alpha = 1$).

The implied steady state for this model is,

$$f(u) + g(u,0) - a = 0, \tag{3.9}$$

the solution of which gives the so-called constant "natural" rate of unemployment.[15] For example, Alogoskoufis and Smith (1991) examine the expectations-augmented wage equation,

$$\Delta w_t = \alpha_0 + \alpha_1 \Delta \hat{p}_t - \alpha_3 \Delta u_t - \alpha_4 u_{t-1}, \tag{3.10}$$

where $\Delta \hat{p}_t \equiv E(\Delta p_t \mid \Im_{t-1}) = \pi(1-\rho) + \rho \Delta p_{t-1}$, and $\Im_{t-1}$ is the information set. Substituting into (3.10) we obtain

$$\Delta w_t = \alpha_0 + \alpha_1 \pi(1-\rho) + \alpha_1 \Delta p_{t-1} - \alpha_3 \Delta u_t + \alpha_4 u_{t-1} + e_t. \tag{3.11}$$

Sargan (1964) proposed a modification of the basic Phillips wage equation to include real wages effects. The rationale for this was the idea that producers acted as price setters, deciding on the price level within the current period in view of labor costs. On the other hand, workers only adjust wages slowly through the bargaining process so as to attain their desired level of real wage. In this form, the wage equation is interpreted as a dynamic adjustment equation where nominal wages adjust proportionally to the gap between actual and target (equilibrium) real wages in the previous period. A simple form of the implied wage equation is,

$$\Delta w_t = \alpha_0 + \alpha_1 \Delta p_t - \alpha_2 u_{t-1} - \alpha_3 (w_{t-1} - p_{t-1}) + \alpha_4 t, \tag{3.12}$$

where the growth path of target real wages is given by the equilibrium solution to the model by writing the composite term

$$-\alpha_3 \left[\ln \left[\frac{w_{t-1}}{p_{t-1}} \right] - \ln \omega_t \right], \quad \omega_t = \left[\frac{w_0}{p_0} \right] \exp \left[\frac{\gamma + \alpha_4 t}{\alpha_3} \right]. \tag{3.13}$$

Sargan did not treat unemployment as completely exogenous in this model, using instruments for both prices and unemployment. Given an associated price mechanism he deduced the steady state rate of unemployment as a function of real wages, and trend.

This "error-correction" formulation is attributed to Phillips (1954) (see Alogoskoufis and Smith (1991a)), and has provided the basis for a number of subsequent empirical studies. For example, Hall (1989, 1990), Nymoen (1989) propose wage equations similar to (3.12) except that a measure of productivity is used rather than a trend, and expected inflation is substituted for actual inflation. Hatton (1988) includes real unemployment benefits in the model, and Alogoskoufis and Smith (1991a) consider a number of related modifications to the dynamic structure of the model.

As noted by Ashenfelter and Card (1982), the inability of empirical models of wage inflation developed during the 1960's to predict the simultaneous high inflation and high unemployment of the 1970's has raised the level of interest in reexamining the theoretical foundations of these models.

One strand of theoretical work has attempted to combine the rational expectations literature with that on micro-economics of the wage setting process. For example, Fischer (1977) and Taylor (1979, 1980) have proposed models of output, prices and wages in which wages are determined via a sequence of over-lapping contracts. That is, not all workers/employers are able to immediately negotiate new wage contracts when there are changes in the price level. For this reason, the Taylor model assumes that current wages are determined, inter alia, by the previous periods wage, and the expectation of next periods wage rate. A specification embodying this idea is

$$\Delta w_t = \alpha_0 + \alpha_1 \Delta \hat{w}_{t+1} + \sum_{j=1}^{k} \beta_j \Delta w_{t-j} + \sum_{i=0}^{h} \gamma_i u_{t-i} . \qquad (3.14)$$

A quite different theoretical approach is provided by the neoclassical inter-temporal consumption/leisure substitution model of Lucas and Rapping (1968). The equilibrium conditions for derived labor demand and supply functions are used to obtain an equation for the unemployment rate as a function of expected real wages. This short-run "inverted" Phillips curve contrasts with the previous excess labor demand interpretations because the causality implicitly runs from real wages to unemployment. For example, when current real wages or real interest rates are unexpectedly high, households will supply more labor this period, and unemployment will fall. However, as with the expectations-augmented Phillips curve model, the forces initiating the unanticipated changes in real wages were not explicitly specified.

Lucas and Rapping used an adaptive mechanism to generate expectations for their empirical model. Subsequently, Lucas (1973), Altonji (1982), Ashenfelter and Card (1982), Zeldes (1989), among others, have proposed alternative expectation formation schemes.

The Lucas-Rapping intertemporal substitution model is closely related to the real business cycle growth model, such as in Kydland and Prescott (1980). This models also yields a positive relationship between real wages (average productivity) and employment (hours worked), with the source of uncertainty being attributed to shocks in technology. Christiano and Eichenbaum (1992) have modified this model to allow private and public spending shocks to have different effects on employment.

This selective review of the literature has highlighted the importance of adjustment mechanisms of nominal or real wages, and expected inflation, in explaining the dynamics of the aggregate labor market. In Section 4 alternative forms of several of these models are estimated, and the structural implications investigated in the context of a well-defined general statistical model (reduced-form) for the data. To this end, in the next section I describe the data used, and the specification, estimation and evaluation of the statistical model.

3.4 Statistical Model Specification, Estimation and Testing

Ashenfelter and Card (1982), Blanchard (1986), among others, have argued that one problem that has plagued traditional empirical models of the labor market is that not enough "facts" are used in discriminating among alternative theoretical explanations. They use linear autoregressive and VAR models to describe the temporal properties of the data, and argue that a useful theory of the aggregate labor market should be consistent with these characterizations. In this section I also examine the "facts". However, the approach taken here is somewhat more rigorous in the sense that careful attention is paid to all the probabilistic features of the data; distributional, memory and heterogeneity, before specifying a parametric statistical representation of wages, prices and unemployment.

In view of the discussion of the theoretical models in Section 3.3, and the available sample size, I will focus attention on a number of the key variables. In particular, I consider quarterly data for the period 1954:1-1991:3 (T=152) on:

(a) w_t - the logarithm of the hourly wage rate in manufacturing;

(b) p_t - the logarithm of the GNP price deflator;

(c) u_t - the unemployment rate defined as l_t - e_t, where l_t and e_t are the logarithms of the civilian labor force and employment respectively;

(d) q_t - labor productivity defined as y_t - e_t, where y_t is the logarithm of real GNP; and

(e) r_t - the 3-month Treasury Bill interest rate.

Labor productivity is included because some theoretical models postulate that equilibrium real wages grow proportionally with productivity over time (see Sargan (1964), Alogoskoufis and Smith (1991a)). Interest rates are a measure the opportunity costs associated with consumption/investment decisions, and tends to reflect the general level of economic activity. Other studies have used variables such as the money supply, tax rates, government spending, and import prices (see for example, Blanchard (1986)) .

The data are obtained from the Citibank Economic Data Tapes, and all but r_t are seasonally adjusted.[16] In this notation, a variable Δx_t can be interpreted as an approximate growth rate, and x_t - z_t as a relative term.

The key variables, w_t, p_t, q_t, and u_t, and relevant transformations of these, are plotted over time (t-plots) in Figure 1. As noted in Section 3.2, the information in these data plots relates to the associated joint and marginal distributions, and may provide a preliminary guide in choosing a statistical model for the data, and in isolating potential problems with a proposed specification.

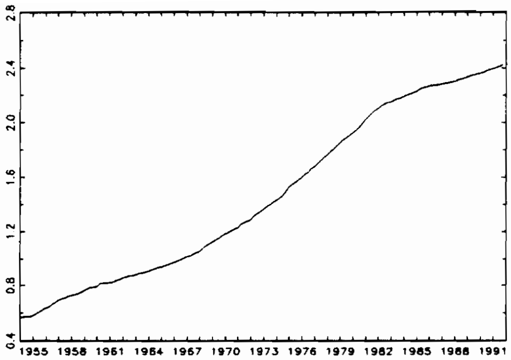
From Figure 3.1(a), we can note that there is a dominant upward trend in w_t over time, that was approximately linear from 1954-1960, exponential from 1961 to 1980, and much flatter subsequently. The effect of this mean non-stationary is also evident from the plot of Δw_t in Figure 3.1(c) (solid line). This plot shows that while wage inflation tended to rise through the 1960's and 70's, the 1980's represents a period of much lower and stable growth in nominal wages. Also evident from Figure 3.1(c) is the positive first-order temporal dependence (autocorrelation), and a pattern of clustering of large and small changes (in absolute value) over time, which suggests the presence of non-linear temporal dependence (see Chapter 4). When the data is dominated by non-stationarity and temporal dependence in this manner, judging the distributional assumption is difficult. However, as will be shown below, even when the first-order dependence is accounted for, the assumption of normality seems to be inappropriate for these series.

Figure 3.1(b)-(c) (dashed line) are t-plots of p_t and Δp_t, respectively. The trend behavior of p_t is very similar to that of w_t, except that during the 1980's, the rate of growth in prices did not declined as markedly as wages. This implies that real wages, $(w_t - p_t)$, have declined over this period. This behavior is evident from the plot of real wages; the solid line in Figure 3.1(d).

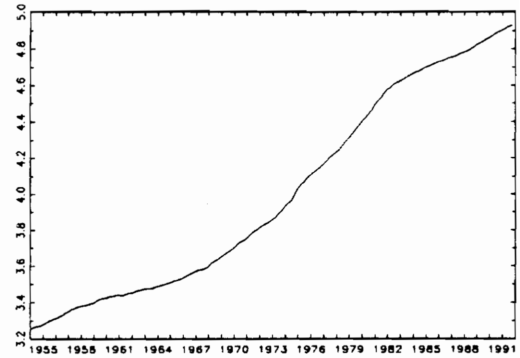
Labor productivity (the dashed line in Figure 3.1(d)), has also not grown at a constant rate over time.[17] Moreover, while the decline in the rate of productivity growth during the 1970's corresponds to the relative slow down in real wage growth, productivity has grown since the end of the 1982 recession, despite the fall in real wages. This implies that the labor share of output $([w - p] + e - y)$ has declined over this period.

The unemployment rate u_t, and the rate of acceleration in unemployment Δu_t, are plotted in Figures 3.1(e)-(f). Unlike the w_t and p_t, u_t displays only a gradual upward trend over time and is dominated by long cycles. These cycles tend to correspond closely to the expansions and recessions associated with the business cycle (see Layard, et.al., 1991). However, as can be seen from the t-plot of Δu_t, while the series has a stationary mean, there is strong second-order dependence effects; periods of relatively high volatility clustered in the late 1950's and late 1970's. Moreover, the data are skewed to the left, with increases in the unemployment rate tending to be larger in absolute value than decreases.

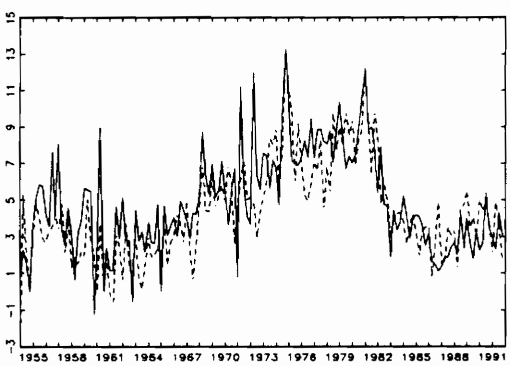
These preliminary results suggest that a linear, homoskedastic system may be an inappropriate model for the wage, price, and unemployment data.[18] For leptokurtic, and non-linearly dependent data, the Student's t VAR with dynamic heteroskedasticity is a possible specification (Spanos (1991)). However, the software for this model is not currently available. When the data are skewed, the question of an appropriate statistical model is less clear. Spanos (1992) has surveyed a number of potential statistical models in this context.[19]



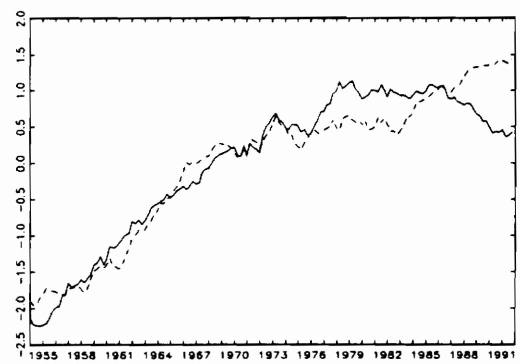
(a) w_t



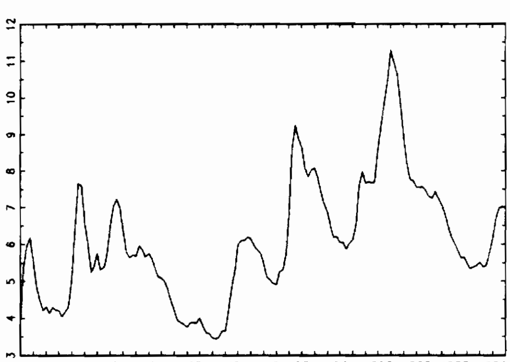
(b) p_t



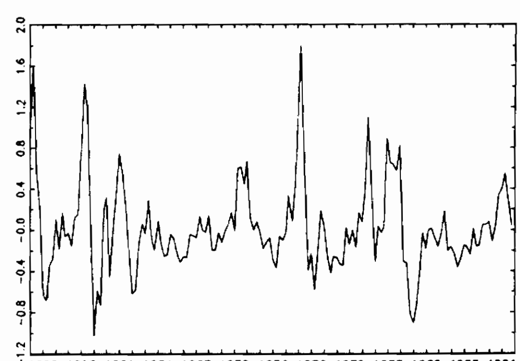
(c) Δw_t (——) Δp_t (- - -)



(d) $w_t - p_t$ (——) q_t (- - -)



(e) u_t



(f) Δu_t

Figure 3.1 - Wage, Price and Unemployment Data

To demonstrate the empirical problems, I proceeded to estimate a three-equation partial-VAR model for $\mathbf{y}_t = (w_t,p_t,u_t)'$, with $\mathbf{Z}_{t-j} = (w_{t-j},p_{t-j},u_{t-j},q_{t-j},r_{t-j})'$, $j=1,..4$, as conditioning variables, and including a unit vector and a linear time trend as part of the constant. The parameters of this model are not of interest per se, but the p-values for a number of system misspecification tests are reported in Table 3.1.[20] These tests provide strong evidence of first and second-order dependence, non-linearity and non-normality in the system.[21]

The residuals from the three-equation system are plotted over time in Figure 3.2(a)-(c). We can see that there is strong second-order dependence in the wage and unemployment equations, with the latter being systematically skewed. In contrast, the price equation residuals are generally well-behaved, except for the effect of one observation in 1974. Adding more lags to the model increased the misspecification problems by introducing spurious correlation into the residuals (over-differencing). Using the log of u_t or first differences or including a number of dummy variables did not remove the skewness and non-linear dependence problems in the unemployment equation.

Based on these results, I decided that the unemployment data are too skewed and non-linearly dependent to be adequately modeled by conventional statistical methods.[22] As an alternative approach, I considered the possibility of treating unemployment as exogenous to the system. This omission is clearly unappealing from a theoretical point of view, but was necessary given the absence of technology to handle the distributional properties of the unemployment series. In this regard, a MDLR model for $\mathbf{y}_{1t} = (w_t,p_t)'$ was estimated using $\mathbf{X}_t = (u_t,q_t,r_t)$, and $\mathbf{Z}_{t-j}$, as conditioning variables, $j = 1,..4$, and including dummy variables for 1959:4, 1960:2, 1971:3, 1972:2, and 1974:2-4 to account for the large spikes in the data.[23] The misspecification test results for the model are reported in Table 3.2.[24]

Table 3.1 - Partial VAR Misspecification Test Results

P-values for System Misspecification Tests

	Wald	Likelihood Ratio	F-Approximation	Lagrange Multiplier
MPCKG	.004*	.008*	.008*	.014*
MLRESET	.020*	.023*	.023*	.042*
MAC(1)	.683	.698	.698	.707
MAC(4)	.005*	.011*	.011*	.045*
MHRESET	.027*	.041*	.042*	.065
MPCWHITE	.017*	.021*	.022*	.032*
MARCH(1)	.176	.237	.238	.295
MARCH(4)	.000*	.006*	.007*	.343

P(8)	.833	MDS	.000*	MDK	.000*

Table 3.2 - MDLR Misspecification Test Results

P-values for System Misspecification Tests

	Wald	Likelihood Ratio	F-Approximation	Lagrange Multiplier
MPCKG	.074	.086	.086	.099
MLRESET	.169	.182	.182	.196
MAC(1)	.649	.654	.654	.656
MAC(4)	.342	.370	.370	.446
MHRESET	.420	.441	.441	.457
MPCWHITE	.316	.326	.326	.348
MARCH(1)	.167	.184	.184	.197
MARCH(4)	.263	.308	.309	.452

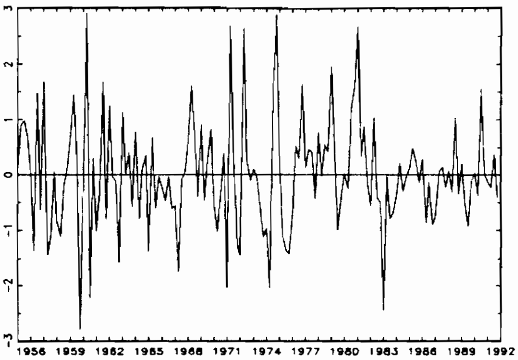
P(8)	.720	MDS	.242	MDK	.098

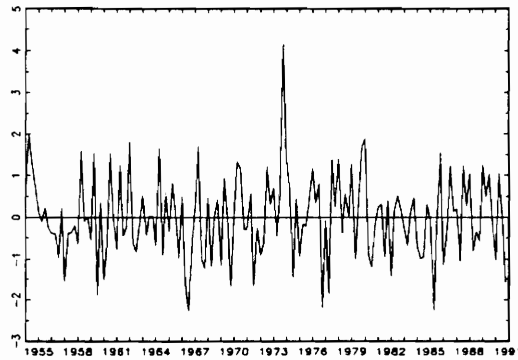
The multivariate tests reported in Table 3.1-3.2 are described in detail in Chapter 2. MPCKG denotes a four-term multivariate principal components version of the second-order Kolmogorov-Gabor polynomial linearity test; MLRESET is a second-order RESET polynomial test for linearity; MAC is the modified LM multivariate autocorrelation test; MHRESET is a second-order RESET polynomial test for heteroskedasticity; MPCWHITE is a four-term principal components version of a multivariate White (1980) heteroskedasticity test; and MARCH is a multivariate version of the Engle (1982) ARCH test for dynamic heteroskedasticity.[25] P(8) denotes an eighth-order multivariate chi-square portmanteau auto-correlation described in Lutkepohl (1991); and MDS and MDK are the chi-square Small (1980) multivariate skewness and kurtosis tests, using the adjustments suggested in D'Agostino and Stephens (1986).

The t-plots of the residuals from this model are presented in Figure 3.2(d)-(e). We can see that the outliers are not as pronounced as for the partial VAR residuals, and there is no evidence of unmodeled temporal dependence, or trends. This is confirmed by the misspecification tests reported in Table 3.2, although there is some marginal evidence of functional form misspecification of the conditional mean, and excess kurtosis.

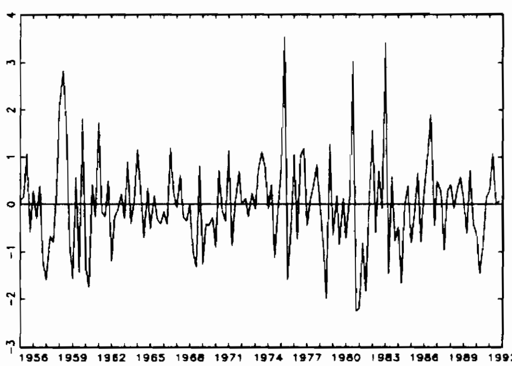
As a further indication of the appropriateness of the lag structure used, Figure 3.3(a) is a plot of the 12-period cross-correlation function for the residuals. The asymptotic 0.05 significance level for these correlations is $2/T^{1/2} = 0.17$, suggesting that the are no serious autocorrelation problems for this model.

Another issue concerns the stability of the parameter estimates over time. In this regard, Figure 3.3(b) is a t-plot of a recursive estimate of the correlation between the residuals using the first 70 observations to initialize the estimates. This plot suggests that, while the (contemporaneous) covariance structure is not constant, it has fluctuated in a relatively tight bound (0.41-0.46) around the full sample LS value of 0.44.

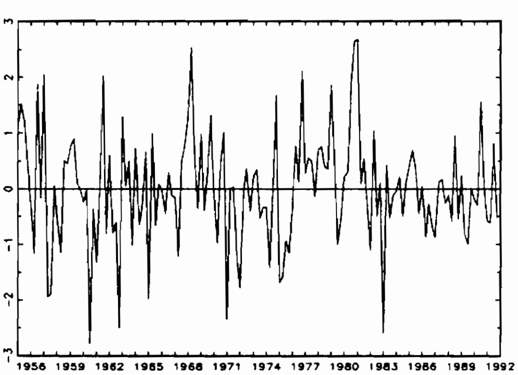


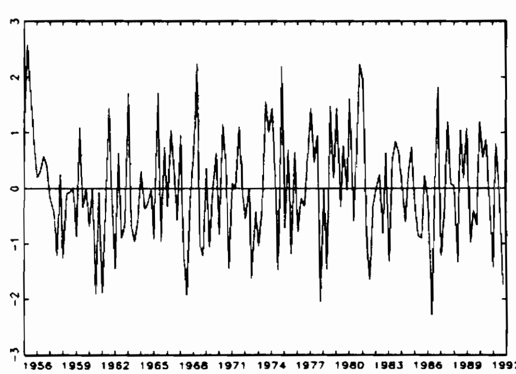


(a) VAR $\hat{u}_{wt}$ (b) VAR $\hat{u}_{pt}$



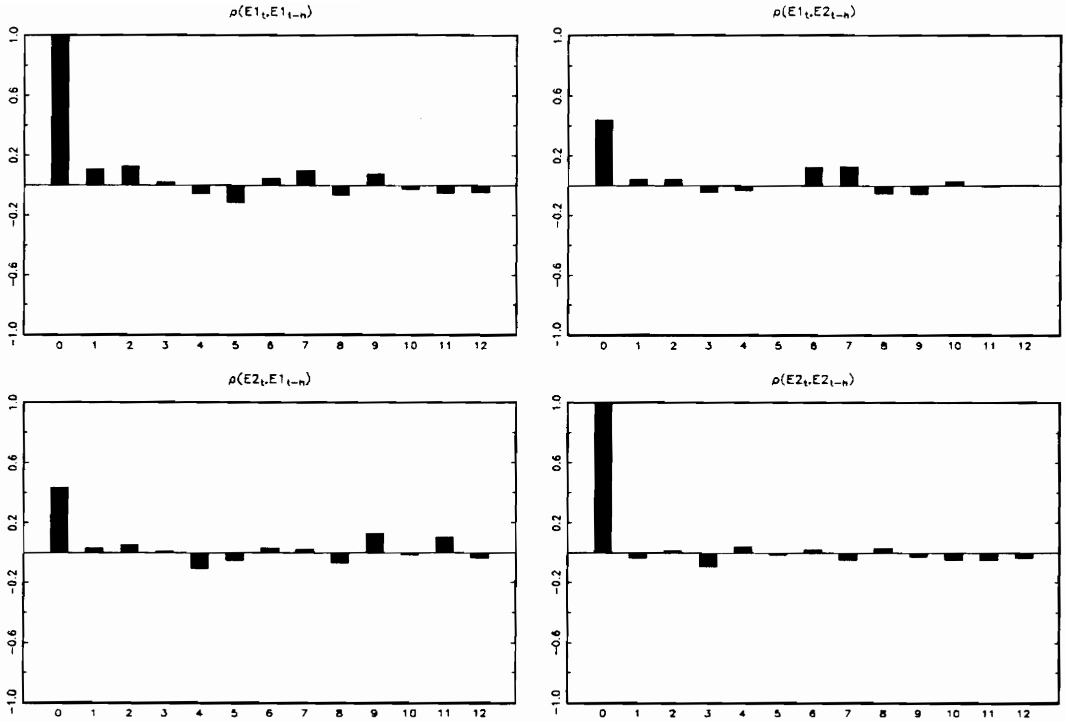
(c) VAR $\hat{u}_{ut}$



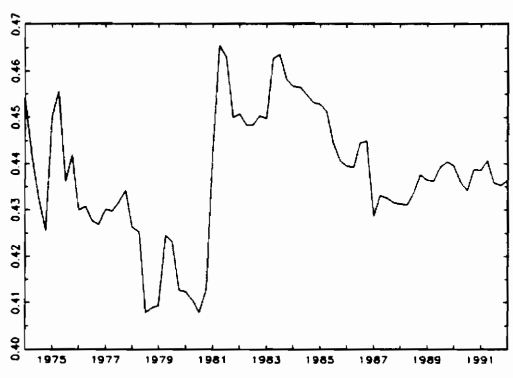


(d) MDLR $\hat{u}_{wt}$ (e) MDLR $\hat{u}_{pt}$

Figure 3.2 - Partial VAR and MDLR Residuals



(a) 12-period cross-correlation functions



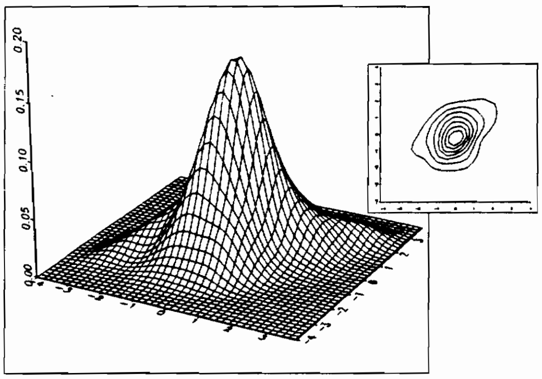
(b) Recursive estimate of $\rho(\hat{u}_{wt}, \hat{u}_{pt})$

Figure 3.3 - Cross-Correlation Function and Recursive Correlation

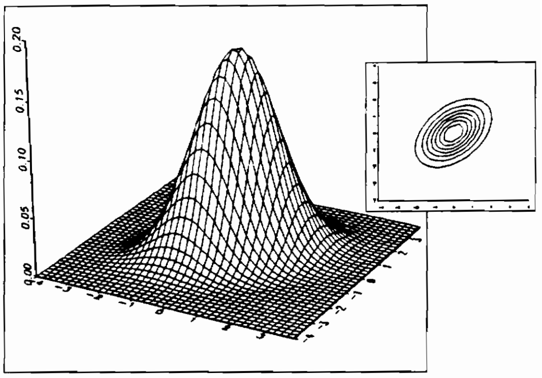
Graphical evidence as to the distributional properties of the bivariate conditional distribution can be examined by considering the bivariate density estimate presented in Figure 3.4. The bivariate density plot for the standardized residuals $\hat{u}_{it}/\hat{\sigma}_i$ (Figure 3.4(a)), is computed using a standard normal kernel with the same correlation structure as the residuals (0.413), and a bandwidth of $h = T^{-1/6} = 0.44$ (see Silverman (1986) for details). These plots indicate that the bivariate distribution is approximately symmetrically bell shaped and unimodal. When compared with the bivariate standard normal density in Figure 3.4(b), we see that the estimate has a similar overall shape, but the tails of the empirical distribution are irregular. However, similar contour patterns were obtained for this sample size using pseudo-random bivariate normal data generated by the program RANDOM.[26]

As an indication of the univariate distributional properties, Figure 3.4(c)-(d) presents probability plots (p-plots) for each residual. By transforming the scale of the empirical distribution function, the normal probability plot is a straight line, which makes inspection for non-normality straightforward in random samples. For the observations $x_1 \leq x_2 \ldots \leq x_T$, a p-plot is a plot of $z_t = G^{-1}(F_T(x_t))$ on x_t, where, $F_T(x_t) = (t-.5)/T$, $t = 1,\ldots,T$, is often used in practice.[27] For platykurtic data the resulting plot will have an S shape, and for leptokurtic data and inverted S appearance (see D'Agostino and Stephens (1986) for details). We can see that neither of the p-plots suggest major deviations from normality.

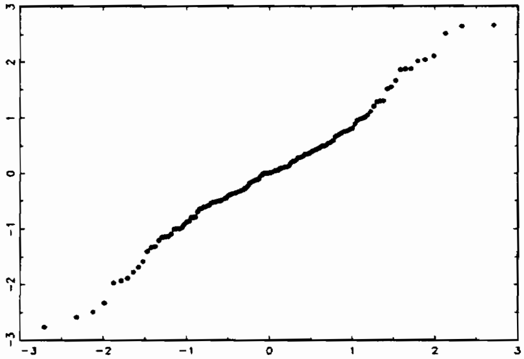
Taken together, these results suggest that the dummy-adjusted MDLR model for w_t and p_t is a reasonable summary of the probabilistic features of the data that can be used for structural modeling purposes. However, the absence of unemployment as a co-determined variable in this model implies that the structural economic models will be theoretically incomplete.



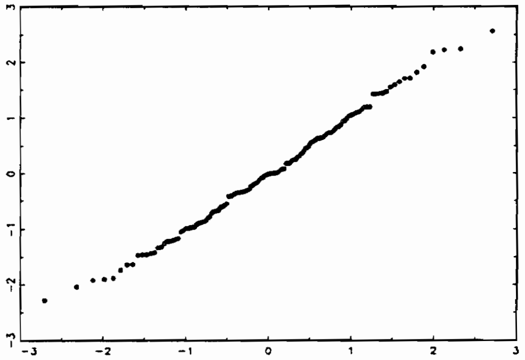
$\hat{D}(u_{wt}, u_{pt} \mid Z_{t-1}; \theta)$



(b) Bivariate Normal



(c) p-plot - $\hat{u}_{wt}$



(d) p-plot $\hat{u}_{pt}$

Figure 3.4 - MDLR Distributional Plots

3.5 Structural Model Specification, Estimation and Testing

In this section, I consider the estimation and testing of specific forms of three of the structural models summarized in Section 3.3.

The first structural model is a modified version of the simple nominal wage adjustment equation of Lipsey (1960). This can be represented as,

$$\Delta w_t = \alpha_0 + \alpha_1 \Delta p_t + \sum_{j=0}^{h} \beta_j u_{t-j} + \epsilon_t, \tag{3.15}$$

where α_1 is an indexation coefficient, and $\beta_0 < 0$ is expected on theoretical grounds.

This model was estimated including the dummy variables, and a second-order lag polynomial in unemployment. Other theoretical restrictions on this equation include: no long-run unemployment effect, $\Sigma_{j=0}^{h}\beta_j = 0$; the zero indexation assumption of Lipsey (1960), $\alpha_1 = 0$; or the perfect adjustment assumption of Friedman (1968), Phelps (1968), $\alpha_1 = 1$. Note that this model is incomplete because there are no explicit structural equations for either prices or unemployment are specified, although prices are treated endogenously for estimation purposes.

An expectations-augmented version of this model was estimated using only lagged information in forming the instruments for p_t. In this form, the model of Alogoskoufis and Smith (1991) (equation (3.10)) arises as a special case. However, unlike Alogoskoufis and Smith, all conditioning variables from the statistical model are used in forming the expectations.

The estimation and testing results for this model are summarized in Table 3.4. The first column reports the results using instrumental variables for p_t based on the variables in the MDLR model. The second column reports the results from using the variables in a partial VAR equation of p_t as instruments.[28] We can see that the signs of the key coefficients are what one would expect a priori, but the restriction that $\alpha_1 = 1$ is rejected at the 0.05 level of significance using heteroskedastic consistent standard errors, while $\Sigma^{h}_{j=0}\beta_j = 0$ is not rejected.

However, these structural inferences are of questionable validity. The over-identifying restrictions for this equation are strongly rejected in terms of the underlying statistical model. Moreover, the single equation misspecification tests reported in Table 3.3, based on the "structural" residuals, show clear evidence of model misspecification.

The second model is a modified version of the two equation wage-price system proposed by Lipsey and Parkin (1970),

$$\Delta p_t = \alpha_{01} + \alpha_{11}\Delta w_t + \sum_{j=1}^{s} \beta_{1j}\Delta p_{t-j} + \alpha_{12}\Delta q_t + \epsilon_{1t}, \tag{3.16}$$

$$\Delta w_t = \alpha_{20} + \sum_{j=0}^{k} \beta_{2j} u_{t-j} + \alpha_{21}\Delta p_t + \alpha_{22} r_t + \epsilon_{2t}. \tag{3.17}$$

The two-stage least squares estimation results are summarized in Table 3.4. Again the signs of the key theoretical variables are as expected a priori, and real wage restriction $\alpha_1 = 1$, and the absence of long-run effects would not be rejected in the wage equation. However, the structural specification is rejected in terms of both the over-identifying restrictions imposed, and its own statistical adequacy.

The third structural model is the real-wage adjustment equation of Sargan (1964). The estimated form is,

$$\Delta w_t = \alpha_0 + \alpha_1\Delta p_t + \alpha_2\Delta p_{t-1} + \alpha_2\Delta u_t + \alpha_3(w\text{-}p\text{-}q)_{t-1} + \epsilon_t, \tag{3.18}$$

which is based on the models used by Alogoskoufis and Smith (1991a). In this model, the equilibrium real wage is approximated by q_t rather than a time trend. The zero response to unemployment in the long-run is imposed, but the rate of acceleration in unemployment is included to capture any short-run unemployment effects. Also, the perfect indexation restriction is not imposed a priori, and lagged inflation is included to capture lagged inflation effects. The estimation results for this model are summarized in Table 3.5. Column A contains the estimates based on the complete set of instruments for p_t, and column B contains the estimates based on the time t-1 forecasting equation for p_t.

We can see that in this case all the coefficients are significant. There is no strong indication of misspecification, apart from some marginal evidence of autocorrelation. However, the over-identifying restrictions are rejected at conventional significance levels, and the coefficient on the error-correction term has the opposite sign as to what would be expected a priori. In view of the discussion of the data in Section 3.4 this latter result is not surprising. The labor share of output has tended to decline over time, while nominal wage inflation has had on overall upward trend. Interestingly, re-estimating this model through 1980 yielded a marginally significant negative coefficient on $(w\text{-}p\text{-}q)_{t-1}$, suggesting that the structural parameters of the model are unstable.[29]

Other modifications of the real wage model were also considered. However, none were found to lead to a more satisfactory model, in terms of the over-identifying restrictions imposed, statistical adequacy, or theoretical interpretation.

Table 3.3 - Phillips Curve Model Estimation Results				
Dependent Variable: Δw_t				
	(A)		(B)	
Δp_t	.8146	[.0497]	.8435	[.0517]
u_t	-1.3446	[.4185]	-1.8442	[.4270]
u_{t-1}	2.0503	[.7490]	2.8523	[.7731]
u_{t-2}	-.7784	[.4068]	-1.1137	[.4193]
R^2	.669		.662	
s	1.5827		1.5900	
Basmann	3.0817	(.0000)		
Anderson-Rubin	69.3262	(.0000)		

P-values for restricted structural equation misspecification tests

	(A)	(B)
DS	.2909	.3240
DK	.3174	.2738
KG2	.0034*	.0089*
AC(1)	.0165*	.0320*
AC(4)	.0028*	.0041*
LB(12)	.0036*	.0146*
ARCH(1)	.2262	.2054
ARCH(4)	.4288	.3711
ML(12)	.8263	.7028

See Appendix 4A for definition of these single equation misspecification tests and Spanos (1986) for details.

Table 3.4 - Two-Equation Phillips Curve Model Estimation Results

	Δw_t		Δp_t	
Δp_t	.9293	[.0720]		
u_t	-1.1198	[.4498]		
u_{t-1}	1.7167	[.8059]		
u_{t-2}	-.5881	[.4432]		
r_t	-.1384	[.0719]		
Δw_t			.8076	[.0722]
Δp_{t-1}			.3426	[.2446]
Δq_t			-18.6766	[6.4473]
R^2	.672		.700	
s	1.5031		1.4407	
Basmann	2.7486	(.0005)	2.2520	(.0049)
Anderson-Rubin	58.5792	(.0000)	47.9945	(.0002)

P-values for restricted structural equation misspecification tests

DS	.3720	.3203
DK	.2656	.4475
KG2	.0114*	.0005*
AC(1)	.1644	.0610
AC(4)	.0282*	.0454*
LB(12)	.2669	.1088
ARCH(1)	.3452	.3757
ARCH(4)	.6181	.5507
ML(12)	.9133	.4065

Table 3.5 - Real-Wage Adjustment Model Estimation Results				
Dependent Variable: Δw_t				
	(A)		(B)	
Δp_t	.6227	[.0766]	.5597	[.0847]
Δu_t	-1.1370	[.2846]	-1.1842	[.3210]
$(w-p-q)_{t-1}$	2.0327	[.4265]	2.2230	[.4312]
Δp_{t-1}	.3008	[.0721]	.2567	[.0981]
R^2	.736		.692	
s	1.4181		1.5460	
Basmann	1.9321	(.0306)		
Anderson-Rubin	35.7809	(.0032)		

P-values for restricted structural equation misspecification tests

DS	.2462	.2240
DK	.1908	.1878
KG2	.7476	.6112
AC(1)	.1131	.0632
AC(4)	.3859	.1701
LB(12)	.7921	.6499
ARCH(1)	.5485	.3209
ARCH(4)	.9782	.4515
ML(12)	.8263	.7238

3.6 Conclusions.

This paper describes the empirical modeling of the inflation-unemployment trade-off in the U.S. using an approach that attempts to integrate the traditional structural equation modeling approach with the largely atheoretical time series literature. The reduced-form is viewed as a convenient summary of the sample information whose specification is determined by both the theory and the data. The theory suggests the relevant variables and the general form of the statistical model, while the final form is chosen so as to adequately account for the probabilistic features of the data. The reduced-form model then forms a basis for estimating and testing the structural models of interest.

The preliminary data analysis suggested that the post-1954 quarterly data on wages, prices and unemployment do not admit an adequate normal, linear, and homoskedastic statistical model; due primarily to the non-linear dependence and non-normality in the wage and unemployment data. Thus, previous empirical findings in the literature based on VAR models for these series may be of questionable validity, since they are based on misspecified statistical models.

However, a MDLR model which included a set of dummy variables was found to be a reasonably adequate summary of the distributional properties of the wage and price series based on both the residual graphical and formal system misspecification test results.

This MDLR model was then used as the general reduced-form of three structural models of wage inflation. It was found that these models generally yielded parameter estimates consistent with the underlying theory (a negative relationship between real wages and unemployment), but were rejected both in terms of the implied over-identifying restrictions, and the statistical adequacy of the estimated equation(s). These results suggest that representations of the basic theories of the aggregate labor market, even when supplemented by straightforward adjustment hypotheses, do not tend to fit the data well.

Appendix 3A

Treatment of Expectations

Consider the SEM,

$$\mathbf{C}\mathbf{y}_t + \mathbf{\Gamma}\hat{\mathbf{y}}_t + \mathbf{D}\mathbf{x}_t = \epsilon_t, \qquad \epsilon_t \,|\, \mathbf{X}_t \sim \mathrm{NIID}(\mathbf{0},\mathbf{\Sigma}), \tag{A3.1}$$

where $\hat{\mathbf{y}}_t$ (hx1) are a set of unobservable (latent) expectations formed at time t-1 on h $\leq$ m endogenous variables based on the information set $\Im_{t-1}$, $\mathbf{x}_t$ is a kx1 vector of observable exogenous variables.[30] The derived "reduced-form" is $\mathbf{y}_t = -\mathbf{C}^{-1}\mathbf{\Gamma}\hat{\mathbf{y}}_t - \mathbf{C}^{-1}\mathbf{D}\mathbf{x}_t + \mathbf{C}^{-1}\epsilon_t$, which we may write as

$$\begin{aligned} \mathbf{y}_{1t} &= \mathbf{\Pi}_{11}\hat{\mathbf{y}}_{1t} + \mathbf{\Pi}_{12}\mathbf{x}_t + \mathbf{e}_{1t}, \\ \mathbf{y}_{2t} &= \mathbf{\Pi}_{21}\hat{\mathbf{y}}_{1t} + \mathbf{\Pi}_{22}\mathbf{x}_t + \mathbf{e}_{2t} \end{aligned} \tag{A3.2}$$

where $\mathbf{\Pi}_{i1}$: mxh, $\mathbf{\Pi}_{i2}$: mxk. The rational expectations assumption is that $\hat{\mathbf{y}}_{1t} = E(\mathbf{y}_{1t} \,|\, \Im_{t-1})$, and assuming that $(\mathbf{I} - \mathbf{\Pi}_{11})$ is non-singular we can take conditional expectations of the reduced-form such that $E(\mathbf{e}_t \,|\, \Im_{t-1}) = 0$.[31] Solving for the unobservable conditional expectation $\hat{\mathbf{y}}_{1t}$ yields $\hat{\mathbf{y}}_{1t} = (\mathbf{I} - \mathbf{\Pi})^{-1}\mathbf{\Pi}_{12}\hat{\mathbf{x}}_t$, where $\hat{\mathbf{x}}_t = E(\mathbf{x}_t \,|\, \Im_{t-1})$, and substituting into the reduced-form we obtain

$$\begin{aligned} \mathbf{y}_{1t} &= \mathbf{\Pi}_{11}(\mathbf{I} - \mathbf{\Pi}_{11})^{-1}\mathbf{\Pi}_{12}\hat{\mathbf{x}}_t + \mathbf{\Pi}_{12}\mathbf{x}_t + \mathbf{e}_{1t}, \\ \mathbf{y}_{2t} &= \mathbf{\Pi}_{21}(\mathbf{I} - \mathbf{\Pi}_{11})^{-1}\mathbf{\Pi}_{12}\hat{\mathbf{x}}_t + \mathbf{\Pi}_{22}\mathbf{x}_t + \mathbf{e}_{2t} \end{aligned} \tag{A3.3}$$

We can see that this approach amounts to transforming the initial latent variable problem in terms of $\hat{\mathbf{y}}_t$ into one in terms of the exogenous variables $\hat{\mathbf{x}}_t$; i.e., a "pass the bucket" strategy.

One estimation method is to substitute the predicted values of $\mathbf{x}_t$ into (A3), based on a choice of $E(\mathbf{x}_t \,|\, \Im_{t-1})$. Alternatively, we could use all the variables defining $\Im_{t-1}$ and any contemporaneous exogenous variables as instruments for $\mathbf{x}_t$.[32]

Notes

1. The empirical modeling is implemented using the computer program SAM described in Chapter 2 because other available interactive programs do not currently offer very extensive multivariate data analysis tools or misspecification for multivariate systems.

2. The VAR is a more general statistical model in the sense that the parameters of the MDLR model can be obtained by reparameterizing those of the VAR (see Spanos (1986, p.523)).

3. The partial VAR can be viewed as a decomposition of the Haavelmo distribution but where inferences are conditional on a k-vector of $\mathbf{Z}_t$ being treated as given.

4. The mapping H(·) corresponds to the inverse mapping used to identify structural model parameters in the traditional approach (see Hsiao (1983)).

5. From the empirical literature we can note that, in general, not only are the statistical assumptions underlying the reduced-form not tested, but the reduced-form is rarely estimated explicitly. When estimation is by IV methods, the implied reduced-form is often not even specified, with the instruments chosen using some other criteria, such as economic theoretic grounds.

6. These conditions are actually only necessary conditions for identification. However, they are generally also sufficient if we exclude the case of phoney restrictions, such as all the equations satisfying the same restriction, or that some other equation satisfying all the restrictions of the *i*th equation.

7. When the model is estimated by full information maximum likelihood the appropriate tests are likelihood ratio tests.

8. This is formulation is equivalent to τ_1 = (RRSS-URSS)/URSS, where RRSS is the restricted residual sum of squares defined as $(\mathbf{y}_1 - \mathbf{Y}_1\hat{\boldsymbol{\gamma}}_{IV} - \mathbf{X}_1\hat{\boldsymbol{\delta}}_{IV})'(\mathbf{y}_1 - \mathbf{Y}_1\hat{\boldsymbol{\gamma}}_{IV} - \mathbf{X}_1\hat{\boldsymbol{\delta}}_{IV})$, and URSS $= (\mathbf{y}_1 - \mathbf{Y}_1\hat{\boldsymbol{\gamma}}_{IV})'\mathbf{M}_X(\mathbf{y}_1 - \mathbf{Y}_1\hat{\boldsymbol{\gamma}}_{IV})$.

9. This is because the restricted form can no longer be estimated by least squares.

10. Actually, there is some evidence that the empirical relationship embodied by the Phillips curve was discovered much earlier by Fisher (1926) (see Donner and McCollum (1972)). An extensive recent survey of the literature is provided by Layard et.al., (1991)).

11. Let y_t be real output, e_t be the level of employment, and m_t be the mark-up of the value of output over wages (all measured in logs). Then we can write $p_t + y_t = m_t + w_t + e_t$, so that $\Delta p_t = \Delta m_t + \Delta w_t + (\Delta y_t - \Delta e_t)$, where $(\Delta y_t - \Delta e_t)$ is the (approximate) growth rate of labor productivity, and $\Delta m_t = 0$ if there is a constant markup of unit labor costs (i.e., the share of profits in the value of output is constant). Often it is also assumed that the rate of productivity growth is constant.

12. Labor productivity is often assumed to grow at a constant rate (see Blanchard and Fischer (1989)). However, as will be seen in Section 3.4, this is not supported by the empirical evidence.

13. The Phillips curve in its original form was postulated as a non-linear relationship such that at high rates of unemployment the relationship to wage inflation would be flat because agents would be unwilling to offer their labor at less than the current rate when the demand for labor is low and unemployment is high.

14. Ideally one would wish to use number of unfilled vacancies and the number of unemployed workers as a measure of excess demand. Lipsey linearized the Phillips curve by using $1/u_t$ and $1/u_t^2$ and Δu_t as regressors.

15. This is also termed the non-accelerating inflation rate of unemployment. However, models based on the concept of a constant natural rate of unemployment ignore the possibility that during a recession, both physical and human capital may deteriorate, reducing the economy's overall productivity; may result in higher unemployment insurance payments; or otherwise raise the social acceptability of unemployment. These phenomena have been summarized in the so-called "hysteresis effect" (Phelps (1972). The importance of this effect is that it suggests that the natural rate will increase during a recession, and reduce the full-employment output of the economy beyond the end of the recession.

16. Ashenfelter and Card (1982) uses quarterly data logarithms of wage rates, GNP deflator, and the unemployment rate, and the Treasury Bill rate from 1956-1980. Blanchard (1986) considers the logarithm of wages, CPI, and civilian employment for 1954-1984. Gordon (1983) has shown that the behavior of a number of these quarterly series is quite different pre and post 1954.

17. Both real wages and productivity have been scaled to fit on the same axis.

18. By way of comparison, I also examined annual data series. These data were obtained from *The Economic Report of the President* for the period 1949-1990 (T-42). These data displayed similar general characteristics as the quarterly data, and lead to the same type of statistical problems in the estimated models.

19. Using the logarithm of u_t, as in Ashenfelter and Card (1982), did not alleviate the skewness in this series.

20. It is important to emphasize that these linear regressions do not take account of any higher-order dependence in the data.

21. The results for the fourth-order Multivariate ARCH test demonstrates the potential for conflicting inferences among the various multivariate test criteria. See Appendix 2B for a discussion.

22. This also raises questions as to the reliability of the inferences from previous studies that have used this basic partial-VAR specification and similar data series (see for example, Ashenfelter and Card (1982), Altonji (1982), Geary and Kennan (1984), Poterba, Rotemberg and Summers (1985), and Blanchard (1986)).

23. The dummy in 1959 corresponds to the effect on wages of the steel workers strike, and that in 1960 accounts for the jump in wages preceding the 1960-1961 recession. The 1971-1974 dummies are for the effects of the Nixon administration wage-price controls. Controls were imposed in August 1971 which froze all but interest rates, taxes and agricultural prices. In November 1971, a standard rate of wage increase of 5.5% was set, and in June 1973 a second freeze was imposed on prices only, while wage constraints where relaxed somewhat. In April 1974 the wage-price controls where abandoned.

24. Because the specification was chosen simply to capture the salient probabilistic features of the data, the estimated parameters of the MDLR model are not reported. However, I note that first difference restrictions on the wage and price series would not be rejected. Also, while the dummy variables where strongly significant, trend terms added little to the adequacy of the specification so were not included in the final form. A partial VAR for Δw_t and Δp_t was also estimated and found to be generally adequate.

25. Because there are too many regressors to form the full quadratic form of the Kolmogorov-Gabor polynomial, only the squared terms were included.

26. RANDOM is a GAUSS program written for generating univariate and multivariate pseudo random numbers from a number of common distributions, and provides considerable flexibility in selecting shape, scale, location and dependence parameters. RANDOM was written jointly by myself and Dr Spanos.

27. For a normal distribution $z_t = \text{sign}[F_T(x_t) - .5][1.238s_t(1 + 0.0262s_t)]$ where $s_t = \{-\ln[4F_T(x_t)(1 - F_T(x_t))]\}^{1/2}$, has been suggested as an accurate approximation (D'Agostino and Stephens (1986)).

28. Including contemporaneous variables to define the (rational) expectation of p_t given information available at time t-1 may be inappropriate (see Appendix 3B). For that reason a forecasting equation for p_t involving only lags of the conditioning variables was also estimated (i.e., a single equation partial VAR). This equation was found to be reasonably adequate, except for some evidence of non-linearity.

29. See Alogoskoufis and Smith (1991) for a discussion of parameter stability as it relates to the Phillips curve.

30. The case of expectations of exogenous variables is straightforward (see Pagan and Wickens (1989)), and methods for dealing with expectations of future endogenous/exogenous variables is discussed in Pesaran (1987).

31. The assumed correspondence between rational expectations and the conditional expectation operator suggests a number of optimal properties in terms of mean square prediction (see Shiryayev (1984, p.235)). In particular, the prediction error is orthogonal to the information in the conditioning information set $\Im_t$. Defining the prediction error as $u_t = y_t - E(y_t \mid \Im_{t-1})$, where $E(u_t \mid \Im_{t-1}) = 0$, and if $E(u_t^2) < \infty$, we have $E(u_t u_{t-i} \mid \Im_{t-i}) = 0$, since $\Im_{t-i} \subset \Im_t$ for all $i \geq 1$, and $E(E(y_t \mid \Im_{t-1})u_t \mid \Im_{t-1}) = 0$.

32. If the variables defined by $\Im_{t-1}$ coincide with the exogenous and predetermined variables in the reduced-form the substitution and instrumental variable methods will coincide. Also, it is necessary to use the reduced-form with $\hat{x}_t$ replaced by the realization of x_t in computing the residual covariance based on the parameter estimates from the two-step procedures (Pagan (1984)). An alternative approach is to estimate the forecasting equations and the actual reduced-form jointly using full information maximum likelihood (see Pesaran (1987, p.166)).

Chapter Four

A Model of the Dynamics of Volatility in U.S. Interest Rates

4.1 Introduction

Modeling the dynamic structure of interest rates represents an important component of both financial and macro-economic analysis. One standard model of interest rate behavior is the linear rational expectations model of the term structure which suggests that long-term yields are a weighted average of expected short-term rates. For example,

$$R_t = \theta + \frac{1}{n}\sum_{k=0}^{n-1} E(r_{t+k} \mid \Im_t) \tag{4.1}$$

where R_t is the yield to maturity on a long-term bond computed from its market price, $E(r_{t+k} \mid \Im_t)$ is the rational expectation of the time t+k short-term rate conditional on information available at time t, and θ is a constant liquidity premium (see Baillie (1988)). The economic rationale underlying this model is quite intuitive. For example, consider a one period bond yielding 10%. If people expect the interest rate on the single period bond to fall to 4% next period, today a two period bond need only yield about 7% to make people essentially indifferent between long and short-term bonds when adjusted for the associated loss in liquidity.

However, there is considerable debate in the literature as to whether long-term yields are too volatile to be consistent with simple averaging models such as (1). For example, Shiller (1979) derived a theoretical measure of excess volatility based on a generalization of (1) and shows that the sample variation of long-term interest rates over time tended to lie outside these bounds (see also Mankiw and Summers (1984)).

Fama (1976,1984) proposed incorporating time-varying risk premia into the term structure as an explanation of the empirical evidence against the rational expectations model. When predictions of volatility change over time, the extra return required for risk-averse agents to hold long-term securities also varies. Typically, these risk premia are assumed to depended on a function of the conditional variance, because the conditional variance provides an obvious measure of systematic risk in this context. Thus, the choice of an appropriate statistical model for interest rate data is an important empirical question.

A number of models of volatility of interest rates are employed in the literature. For example, Weiss (1984) uses the AutoRegressive Conditional Heteroskedasticity (ARCH) model of Engle (1982) to model the clustering effect in monthly bond yields.[1] Engle, Lilien and Robins (1987) employ a weighted ARCH model to describe excess holding yields, and Hong (1988) use the first-order Generalized ARCH (GARCH) model of Bollerslev (1986) for the same variables. Various related studies are summarized in Bollerslev, Chou and Kroner (1992), who note that with few exceptions (e.g., McCulloch (1985)), conditional normality of the error term in these models is satisfactory.[2]

In this paper, I demonstrate the inappropriateness of the conditional normality assumption and the assumption of Markovness implicit in ARCH-type models of interest rate data. As an alternative I apply the Student's t AutoRegressive model with dynamic heteroskedasticity (STAR) of Spanos (1992). For the STAR model, the conditional mean has a linear autoregressive form, and the conditional variance is a quadratic recursive function of all the past history of the series. The STAR model provides a convenient framework for simultaneously analyzing linear and non-linear dependence, and is applied to the modeling of exchange rates by McGuirk et.al., (1992).

The purpose of this paper is to describe the application of the STAR model to monthly long- and short-term bond yields over the post-1960 period. I show that the STAR model:

(a) is a parsimonious, statistically adequate representation of the temporal dependence, and leptokurtosis of all the interest rate series considered;

(b) leads to a different interpretation of systematic interest rate uncertainty than the ARCH-type formulations;

(c) dominates ARCH-type formulations on statistical adequacy grounds; and,

(d) raises questions concerning the appropriateness of previous heteroskedastic models based on the "conditional Student's t" error distribution.

The estimation results suggest that the conditional volatility evolves smoothly over time. The periods of high volatility in bond rates, particularly the late 1960s and the early 1980s, are reflected in uniform increases in the level of predicted volatility. For long-term bonds these estimates have remained relatively high during the 1980s, while volatility predictions in short-term rates have actually declined (but are at a higher level). Interestingly, because all past available information is utilized in deriving the optimal volatility predictions from the model, individual events such as the 1987 stock market crash did not lead to substantial revisions of expectations of future volatility.

The remainder of the paper is organized as follows. In the next section, the probabilistic features of the data are briefly discussed in order to motivate the application of the STAR model to the interest rate data in Section 4.3. In section 4.4, the STAR model is contrasted with the ARCH-type conditional variance formulations based on both normal and "Student's t" conditional error distributions. A summary of the results and implications are presented in Section 4.5.

4.2 Probabilistic Features of the Data

The information provided by various data plots and descriptive statistics relates to the joint and marginal distributions underlying the data. In the context of the Haavelmo reduction framework (Spanos (1989)), the graphical features of SAM are used as a guide in choosing an appropriate conditional model for the data.

The data are log differences of monthly yields on Moody's AAA 20 year bonds (y_{1t}); 10 year treasury notes (y_{2t}); and 3 month U.S. Treasury Bill rates (y_{3t}) from January 1960 to February 1990 (T=362).[3] Each series, standardized by its sample standard deviation is plotted over time in Figure 4.1. From these t-plots a number of probabilistic features of the data can be discerned:

[a] **<u>Leptokurtosis</u>**: The data are reasonably symmetrically distributed (apart from one observation in the y_{3t} series). However, the data are leptokurtic, as indicated by the concentration of observations around the mean, and the large number of "outliers".
[b] **<u>Positive linear dependence</u>**: The series contain long cycles in their level.
[c] **<u>Second-order dependence</u>**: Large and small changes of either sign in each series tend to be systematically clustered together over time. There are differences in these patterns between the long-term and short-term series.

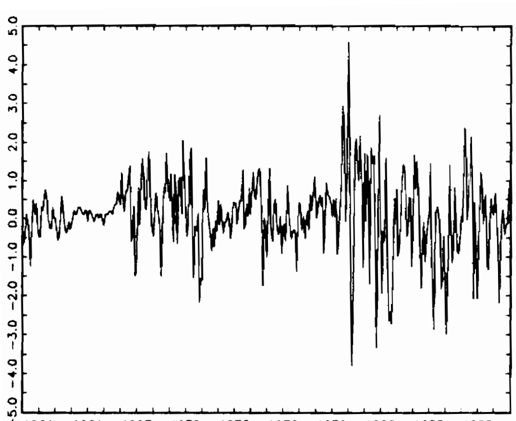
To further examine the probabilistic features of the data non-parametric bivariate densities of $D(y_t,y_{t-1};\phi)$ for each series are estimated, and plotted in Figure 4.2(a)-(c) along with univariate density projections and probability contours.[4] Comparing these estimates with the Normal bivariate density in Figure 4.2(d) we can see that the empirical densities are unimodal, bell-shaped symmetric and leptokurtic; exhibiting peakedness and thick tails.[5] The leptokurtosis is more pronounced in the estimates of $D(y_{3t},y_{3t-1};\phi_3)$ than for $D(y_{1t},y_{1t-1};\phi_1)$ and $D(y_{2t},y_{2t-1};\phi_2)$. The strong positive temporal correlation in each series is also apparent from these estimates. Bivariate estimates for other lags (not reported) yielded similar inferences.

To supplement the graphical evidence, several descriptive statistics for the data are reported in Table 4.1. The sample variance of each series is lower the longer the maturity of the bond, and the sample means are all small and insignificant based on conventional t-tests. The sample skewness and kurtosis coefficients reinforce the graphical evidence of non-normality. The p-values for the various tests for temporal dependence are also reported in Table 4.1. The chi-square LB portmanteau autocorrelation test (Ljung and Box (1977)) indicates the presence of linear temporal dependence, and non-linear dependence is suggested by the ML portmanteau second-order autocorrelation test (McLeod and Li (1983)).[6]

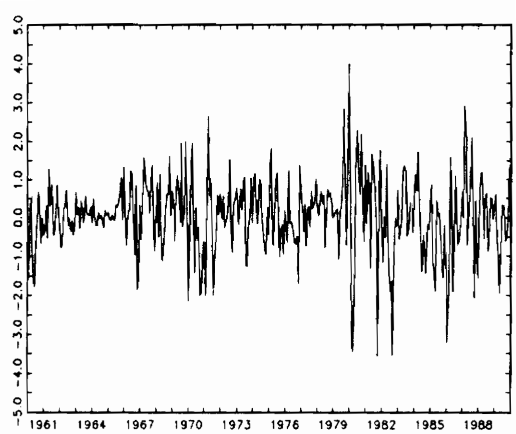
In view of this preliminary evidence, I take as a working hypothesis that each series can be characterized as a stationary, weakly dependent Student's t stochastic process $\{y_t, t \in \mathbb{N}\}$, with $\nu > 2$ degrees of freedom. In the next section I examine the implications and validity of these modeling assumptions in detail.

Table 4.1 - Interest Rate Sample Statistics (T = 361)

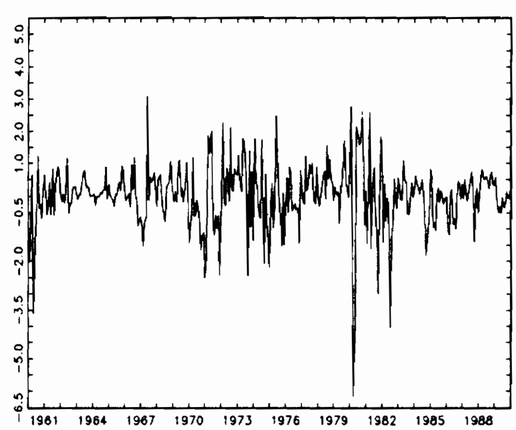
	y_{1t}	y_{2t}	y_{3t}
Mean	0.192	0.162	0.155
Variance	5.754	12.040	47.986
Skewness	0.230	0.295	1.005
Kurtosis	4.813	4.715	8.412
P-values for Sample Test Statistics			
LB(24)	0.0000*	0.0001*	0.0000*
ML(24)	0.0000*	0.0000*	0.0001*



(a) y_{1t} - 20 year AAA bonds

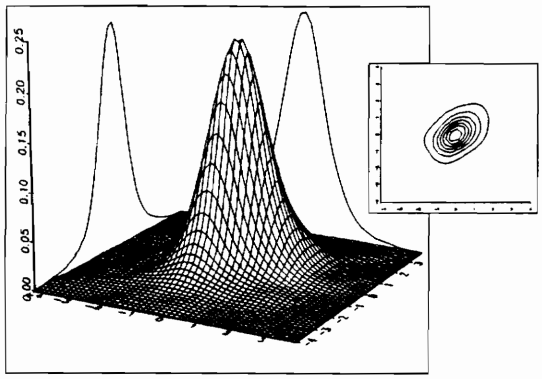


(b) y_{2t} - 10 year Govt. Bonds

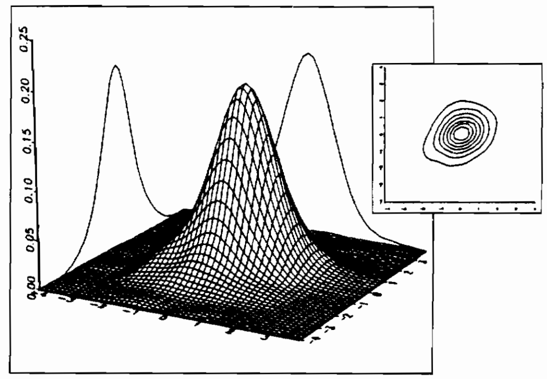


(c) y_{3t} - 3 month Govt. Bills

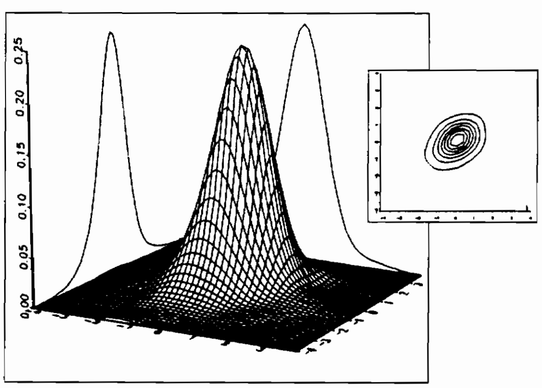
Figure 4.1 - First Differences of (log) Interest Rates



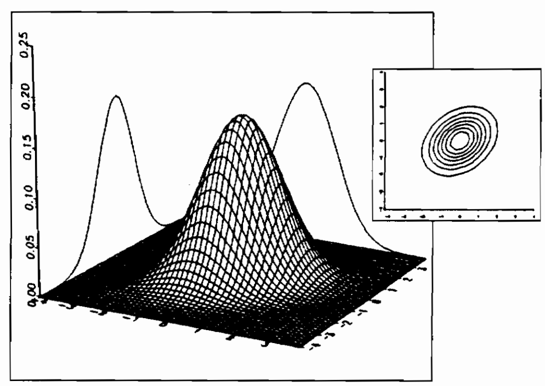
(a) y_{1t}



(b) y_{2t}



(c) y_{3t}



(b) Normal, $\rho = 0.35$

Figure 4.2 - Bivariate Density Estimates of $D(y_t, y_{t-1}; \phi)$

4.3 The STAR Model.

The Student's t AutoRegressive model with dynamic heteroskedasticity (STAR(ℓ,p;ν)) is proposed by Spanos (1992a) as a possible statistical model for stationary, thick tailed, and non-linearly dependent data. The STAR model is based on the first two moments of the Student's t conditional distribution of $\{y_t \mid \mathbf{Y}^{\circ}_{t-1}, t\in\mathbf{N}\}$, and takes the form,

$$y_t = \beta_0 + \sum_{i=1}^{\ell} \beta_i y_{t-i} + u_t, \quad t\in\mathbf{N} \tag{4.2}$$

$$\omega_t^2 = \left[\frac{\nu}{\nu+t-3}\right]\sigma^2\left[1+\sum_{i=1}^{t-1}\sum_{j=-p}^{p}\delta_{|j|}[y_{t-i}-\mu][y_{t-j-i}-\mu]\right], \quad 0\leq p\leq \ell, \tag{4.3}$$

where $u_t \equiv y_t - E(y_t|\mathfrak{F}_{t-1}), \sim St(0,\omega_t^2;\nu)$, ω_t^2 is the conditional variance, $\mathfrak{F}_{t-1} = \sigma(\mathbf{Y}^{\circ}_{t-1})$ is the conditioning information set, $\mathbf{Y}^{\circ}_{t-1} \equiv (y_{t-1}, \ldots, y_1)'$, and $\mu = E(y_t)$. This model is directly related to other Student's t dynamic models when the conditioning information set is expanded via alternative Haavelmo reductions.

From the equations (4.2) and (4.3), we can note that the conditional mean function is linear in the conditioning variables, and the conditional variance is a quadratic function of all past information, rendering the conditional process $\{y_t \mid \mathbf{Y}^{\circ}_{t-1}, t\in\mathbf{N}\}$ non-Markov. In particular, ω_t accumulates locally smoothed squared deviations (from -p to p) of y_t around its unconditional mean over time. In other words, ω_t is a smoothed version of the unconditional variance, and is parameterized with only $m = p+1$ unknown δ_j's.[7]

The statistical parameterization of the STAR model in terms of the underlying joint distribution of $\{y_t, t\in\mathbf{N}\}$ can be most readily illustrated in the context of the static "independent" Student's t linear regression model developed in Spanos (1991),

$$y_t = \beta_0 + \beta' X_t + u_t,$$

$$\sigma_t^2 = \left[\frac{\nu}{\nu+k-2}\right]\sigma^2\left[1 + \left[\frac{1}{\nu}(X_t-\mu_x)'\Sigma_{22}^{-1}(X_t-\mu_x)\right]\right] \tag{4.4}$$

where $Z_t \equiv (y_t, X_t')'$ has a $(1+k)$-dimensional multivariate Student's t distribution (see Zellner (1971)) with mean vector $(\mu_y, \mu_x')'$, and covariance $\nu\Sigma/(\nu-2) > 0$, partitioned as

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12}' & \Sigma_{22} \end{bmatrix} \tag{4.5}$$

Using the linear form of the conditional mean and the properties of the expectations operator (Spanos (1986, p.126)) we can deduce that,

$$\beta_0 = \mu_y - \beta'\mu_x, \text{ and } \beta = \Sigma_{22}^{-1}\sigma_{12}'. \tag{4.6}$$

Moreover, from the identity,

$$\mathrm{Var}(y_t) \equiv E[\mathrm{Var}(y_t \mid \sigma(X_t))] + \mathrm{Var}[E(y_t \mid \sigma(X_t))], \tag{4.7}$$

we have,

$$E[\mathrm{Var}(y_t \mid \sigma(X_t))] = [\nu/(\nu-2)]\sigma^2$$

where $E[(X_t-\mu_x)'\Sigma_{22}^{-1}(X_t-\mu_x)] = E[\mathrm{tr}(\Sigma_{22}^{-1}(X_t-\mu_x)'(X_t-\mu_x)] = \nu k/(\nu-2)$, and

$$\mathrm{Var}[E(y_t \mid \sigma(X_t))] = [\nu/(\nu-2)]\sigma_{12}\Sigma_{22}^{-1}\sigma_{12}'.$$

That is,
$$\sigma^2 = \sigma_{11} - \sigma_{12}\Sigma_{22}^{-1}\sigma_{12}'. \tag{4.8}$$

This result suggests that the conditional mean and variance should not be modeled separately because the coefficients of the model are interrelated through the joint distribution.[8]

The non-Markovness of $\{y_t \mid \mathbf{Y}^{\circ}_{t-1}, t\in\mathbb{N}\}$ raises a number of interesting issues for estimation of the STAR model. Under stationarity, the likelihood function can be written in terms of a recursive decomposition of $D(\mathbf{y};\phi)$, and takes the form (excluding the p initial conditions),[9]

$$\begin{aligned}\ln L \propto & -\frac{T}{2}\ln(\pi) + \ln\left\{\Gamma\left[\frac{1}{2}(\nu+T)\right]\right\} - \ln\left\{\Gamma\left[\frac{\nu}{2}\right]\right\} \\ & - \frac{T}{2}\ln\left(\nu\sigma^2\right) - \frac{1}{2}\sum_{p+1}^{T}\ln\left(c_t^2\right) - \frac{1}{2}\sum_{p+1}^{T}(\nu+t)\ln\left(\gamma_t^2\right)\end{aligned}, \qquad (4.9)$$

where $$\gamma_t^2 = \left[1 + \frac{u_t^2}{\nu\sigma^2 c_t^2}\right], \quad c_t^2 = \left(1 + \left[\mathbf{Y}_{t-1} - \mathbf{1}_{t-1}\mu\right]' \boldsymbol{Q}_{t-1}\left[\boldsymbol{Y}_{t-1} - \mathbf{1}_{t-1}\mu\right]\right). \qquad (4.10)$$

$\boldsymbol{Q}_{t-1} \equiv \boldsymbol{V}_{t-1}^{-1}$ is a p-banded persymmetric matrix, $\boldsymbol{V}_{t-1}$ is the corresponding temporal "covariance" matrix of $\mathbf{Y}^{\circ}_{t-1}$, and is a t-1 dimensional positive-definite, symmetric Toeplitz matrix.[10] If the elements of $\boldsymbol{V}_{t-1}$, $v(|t-s|)$, die out "sufficiently quickly" with $|t-s|$, the model will be operational for small values of p and ℓ. Spanos (1992b, Proposition 1) shows that a memory restriction of the form $|v(k)| \le c\lambda^k$, $0 < c < \infty$, $0 < \lambda < 1$, implies that $\boldsymbol{Q}_T$ has elements $\delta_{ij} = 0$ for $|i-j| > p$, and $\delta_{ij} = \delta_{km}$ for $|i-j| = |k-m| > 0$.

Under these conditions, $\boldsymbol{Q}_T$ can be partitioned in a (symmetric) block tri-diagonal form, with blocks of dimension p, and the off-diagonal matrices are lower (upper) triangular. This suggests that the model can be estimated by imposing Markovness of order 2p and employing the Maximum Likelihood (ML) methods developed in Spanos (1991) for the static Student's t regression model. When the parameters of the lower (upper) triangular blocks of the off-diagonal sub-matrices of $\boldsymbol{Q}_T$ are zero, the ML estimates of the STAR model based on the recursive decomposition in (4.9) will coincide with the Markov-form estimates, and the log-likelihood function will be as in (4.9).

Zellner (1976) shows that there is no ML estimator for υ (see below). However, different values of ν can be chosen for each model using the evidence in the bivariate density estimates (see Figure 4.2) and the sample kurtosis coefficient α_4, where $\alpha_4 = 3 + 6/(\nu\text{-}4)$ for the marginal Student's t distribution as guides. Also, the relationship between the parameters of the model and the unconditional moments provides obvious starting values for the estimation algorithm based on the sample moments.[11]

The estimation algorithms were written as part of an interactive menu-driven Student's t modeling program in the GAUSS matrix language. This program was written by Dr Spanos, Dr McGuirk, and myself, and contains a user interface similar to that of SAM (see Section 2.6). Currently, versions of this program for estimating Student's t linear regression, autoregressions, and dynamic linear regressions are available. The analytical first derivatives of Spanos (1991) were used for the gradient function, and the inverse of the final Hessian matrix from the optimization gave the covariance matrix for the estimators (see the Applications Manual, GAUSS 3.0 for details). Other optimization and covariance methods yielded identical parameter estimates and the same inferences respectively. The convergence tolerance for the estimation algorithm was set at $1\text{x}10^{-12}$ and was achieved in less than three minutes of a 25MHz 486 PC.

The final choice of ℓ, p, and ν was made on the basis of the model's ability to account for the probabilistic features of the data (statistical adequacy), with the significance of the coefficients of the conditional mean and variance providing additional guidance in the selection process. A STAR(2,2;9) model was chosen for y_{1t} and y_{2t}, and a STAR(1,1;7) for y_{3t}. The STAR parameter estimates and misspecification tests based on the weighted residuals proposed by Spanos (1992b) are summarized in Table 4.2.[12]

Table 4.2 - STAR(ℓ,p;υ) Estimation Results			
	y_{1t}	y_{2t}	y_{3t}
$\hat{\beta}_0$	0.155 (.027)*	0.225 (.086)*	0.420 (.200)*
$\hat{\beta}_1$	0.489 (.045)*	0.350 (.050)*	0.324 (.054)*
$\hat{\beta}_2$	-0.213 (.027)*	-0.192 (.051)*	_.___ (.__)
$\hat{\delta}_0$	0.048 (.004)*	0.018 (.001)*	0.008 (.0008)*
$\hat{\delta}_1$	-0.022 (.004)*	-0.006 (.001)*	-0.002 (.0004)*
$\hat{\delta}_2$	0.010 (.003)*	0.004 (.001)*	_.___ (.__)
$\hat{\mu}$	0.214 (.028)*	0.267 (.086)*	0.621 (.195)*
$\hat{\sigma}^2$	2.418 (.211)*	6.374 (.504)*	17.814 (1.626)*
υ	9	9	7
LogL	-803.361	-933.406	-1191.686
P-values for Misspecification Tests			
DS	0.168	0.088	0.036*
DK	0.000*	0.000*	0.000*
KG	0.509	0.949	0.453
AC(2)	0.561	0.958	0.625
AC(4)	0.491	0.980	0.268
LB(24)	0.774	0.896	0.761
HC(2)	0.596	0.598	0.428
HC(4)	0.648	0.326	0.612
ML(24)	0.934	0.741	0.813

(·) are asymptomatic standard errors, and * indicates significant at the 5% level - see Appendix 4A for details of these tests.

Table 4.3. - OLS Estimation Results			
	y_{1t}	y_{2t}	y_{3t}
$\hat{\beta}_0$	0.153*	0.160	0.185
	(0.113)	(0.169)	(0.346)
$\hat{\beta}_1$	0.502*	0.387*	0.321*
	(0.051)	(0.052)	(0.053)
$\hat{\beta}_2$	-0.257*	-0.236*	-0.114*
	(0.051)	(0.051)	(0.052)
$\hat{\sigma}_0^2$	4.558	10.227	43.040
P-values for Misspecification Tests			
DS	0.314	0.261	0.000*
DK	0.000*	0.000*	0.000*
LB(24)	0.046*	0.122	0.066
ML(24)	0.000*	0.000*	0.000*

From Table 4.2 we can note that the conditional mean parameter estimates for the STAR model are quite different from the corresponding LS estimates in Table 4.3. This result is not surprising in view of the homoskedastic assumption underlying the LS models. A comparison of the sample variance, the LS $\hat{\sigma}_0^2$, and the STAR $\hat{\sigma}^2$ provides a relative measure of the effects of modeling all the systematic information in the data. These estimates suggest that the residual noise from the STAR models is more than a third lower than that implied by LS, due primarily to the modeling of the information in the tails of the distribution and the second-order dynamics. Moreover, the estimated means in the STAR model are significantly different from zero and lower for longer-term bonds. All the coefficients of $\hat{\omega}_t$ are significant based on the asymptotic standard errors, and the scaled parameters $\hat{\sigma}^2\hat{\delta}_j$ are similar across all three models.

Misspecification tests based on a typical draw of the weighted residuals suggest that there are serious departures from the underlying model assumptions (see Appendix 4A). In particular, there is only marginal evidence of skewness, and no indication of non-linearity of the conditional mean, unmodeled heteroskedasticity, autocorrelation, or non-linear dependence.

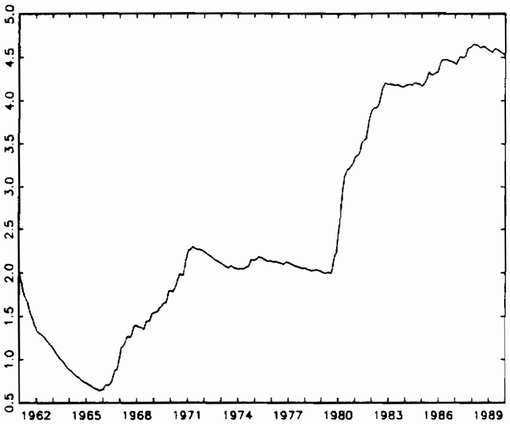
The STAR conditional variance estimates are plotted in Figure 4.3 (a)-(c) and can be easily interpreted from an historical perspective. We can see that initially, the estimates were very sensitive to any variability in the data. However, the relative stability of bond markets during the early 1960s was indicated by correspondingly lower volatility predictions. In the long-term bond market the estimated conditional variance rose between 1966 and 1971, possibly reflecting inflationary pressures associated with the Vietnam war. However, price controls in the early 1970s and active interest rate targeting by the Federal Reserve stabilized variability in long-term bond rates until 1979, as shown by the relatively stable volatility estimates in Figure 4.3.[13] The anti-inflationary monetary policies enacted in 1979 represented another major change in economic conditions, and this was reflected in a second substantial increase in the predicted level of volatility. For long-term rates this higher level is maintained through the 1980's, while the estimated conditional variance for short-term bonds has declined steadily since 1982 (but remains higher in magnitude than for long-term bonds). This latter phenomena possibly reflected the close relationship between the behavior of short-term rates and monetary policy. The information structure of the model is also evidenced by the fact it gives relatively low weight to short-run high-frequency phenomena such as the stock market crash. In particular, the short-run variability in long-term bond rates during this period did not provide systematic new information that led to any real revision of predictions of the overall level of volatility.

The STAR conditional variance estimates also tend to parameterize the description of volatility provided by the recursive LS variance estimates (Brown et. al., (1975)), which are plotted in Figure 4.4.[14] However, unlike these "parameter-free" descriptions, the STAR model provides a parsimonious functional form useful for predictive purposes. An auxiliary regression of the recursive and STAR estimates yielded R^2 values of 0.991, 0.979, 0.974, and regression coefficients of 1.048, 1.001, and 1.101, respectively.

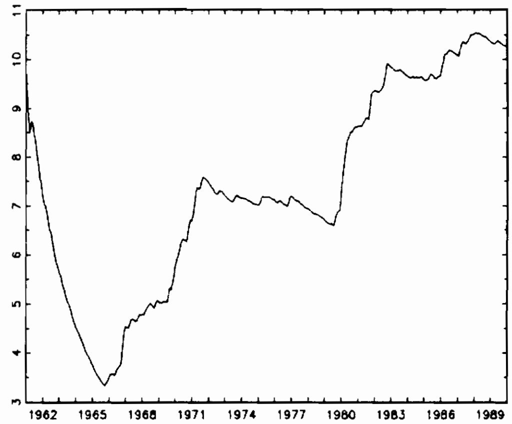
One-step-ahead volatility forecasts can be computed via the one-sided form of (4.3)

$$\hat{\omega}^2_{T+1} = \left[\frac{\nu}{\nu+T-2}\right]\sigma^2\left[1+\sum_{i=1}^{T}\sum_{j=0}^{p}\hat{\delta}_j\left[y_{T-i+1}-\hat{\mu}\right]\left[y_{T-j-i+1}-\hat{\mu}\right]\right] \tag{4.11}$$

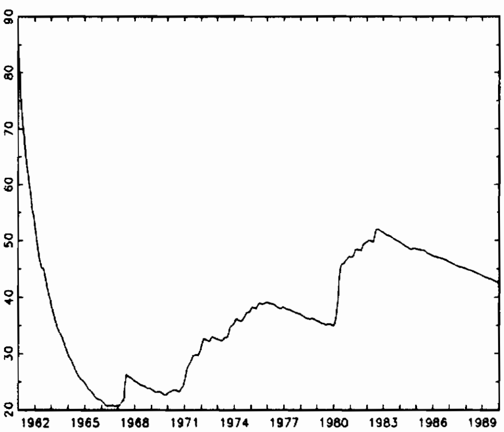
To illustrate these predictions, a sequence of one-step-ahead forecasts of the level of volatility in y_{2T+s} are computed for the period March 1990 to June 1991 ($s=1,\ldots,17$). These variance forecasts are shown in Figure 4.3(d) together with the OLS recursive variance (solid line). An auxiliary regression between the forecast and recursive series yields an R^2 of .947, and a significant regression coefficient of 0.884.



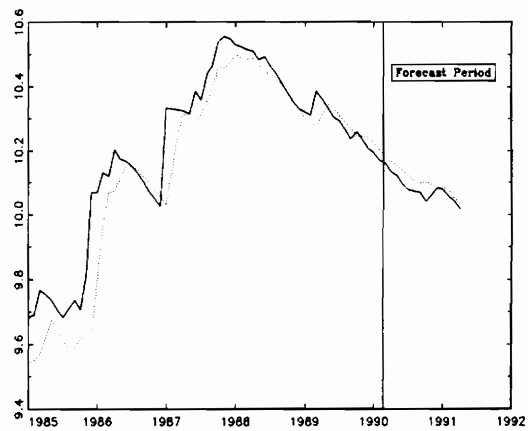
(a) y_{1t}



(b) y_{2t}

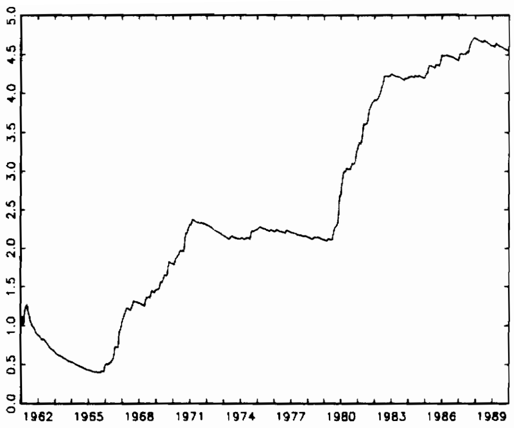


(c) y_{3t}

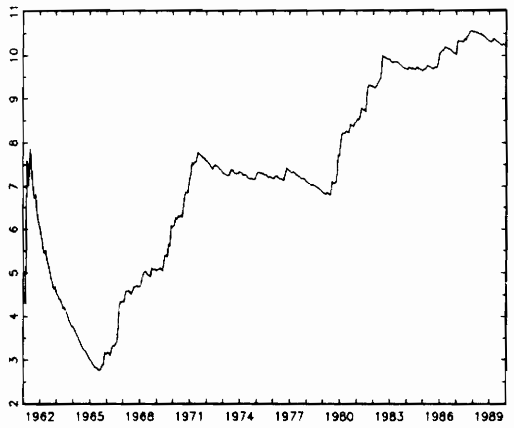


(d) $\hat{y}_{2t}$ - Forecast

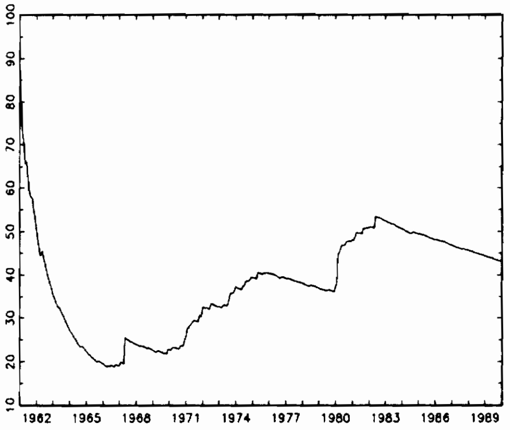
Figure 4.3 - STAR Conditional Variance Estimates and Volatility Forecast



(a) y_{1t}



(b) y_{2t}



(c) y_{3t}

Figure 4.4 - Recursive LS Conditional Variance Estimates

4.4 Statistical Comparisons

In this section I compare the performance of the STAR model with models of dynamic heteroskedasticity based on the ARCH formulation of Engle (1982). While the literature on ARCH-type models is extensive (see Bollerslev et.al., (1992) for a survey), it is instructive to briefly re-consider the main features of these formulations, so that the main differences from the STAR model can be highlighted. In particular, I examine the plausibility of the ARCH and GARCH formulations in terms of the implicit restrictions imposed on the moments of the observable process $\{y_t, t \in \mathbb{N}\}$, and the distributional assumption of conditional normality.

The ARCH conditional variance formulation is,

$$h_t^2 = \gamma_0 + \sum_{i=1}^{p} \gamma_i u_{t-i}^2 \tag{4.12}$$

where $\gamma_0 > 0$, $\gamma_i \geq 0$, $\Sigma_{i=1}^{p}\gamma_i < 1$.[15] The first two parameter restrictions ensure that the conditional variance is positive, and the third ensures the existence of the unconditional variance, and "stability" of the conditional variance (Engle (1982)).[16]

In empirical papers, the number of lags in the ARCH function tends to be large. Engle (1983) suggested an Almon lag form where $\gamma_i = \gamma[(p+1-i)/p(p+1)]$, and Bollerslev (1986), Engle and Bollerslev (1986) proposed a Generalized ARCH (GARCH) formulation,

$$h_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \gamma_i h_{t-i}^2, \tag{4.13}$$

for $\alpha_0 > 0$, $\alpha_i \geq 0$, $\gamma_j \geq 0$, and $(\Sigma_{i=1}^{p}\alpha_i + \Sigma_{j=1}^{q}\gamma_i) < 1$. Equation (4.13) can be viewed as a parsimonious approximation to a higher-ordered ARCH function with exponentially decaying parameters, using a ratio of two lower order lag polynomials.

The statistical parameterization of the ARCH formulation can be illustrated using the results of Spanos (1992a). Let $(y_t, y_{t-1}, \ldots, y_{t-p})' \equiv (y_t, \mathbf{Y}^{p\prime}_{t-1})'$ have a stationary joint distribution with mean $\boldsymbol{\mu} = (\mu_1, \boldsymbol{\mu}_2')'$, covariance $\boldsymbol{\Sigma} > 0$, $\sigma_{11} = \mathrm{Var}(y_t)$, $\boldsymbol{\sigma}_{12} = \mathrm{Cov}(y_t, \mathbf{Y}^p_{t-1})$, $\boldsymbol{\Sigma}_{22} = \mathrm{Cov}(\mathbf{Y}^p_{t-1})$, $\mu_1 = \mu_{2j}$, and $\mathrm{Cov}(y_i, y_j) = h(|i-j|)$ for all i,j.[17] From the linearity of the conditional mean function, we can deduce that $\beta_0 = \mu_1 - \boldsymbol{\beta}'\boldsymbol{\mu}_2$, and $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{12}'$, while the ARCH form of the conditional variance implies

$$E[\mathrm{Var}(y_t \mid \sigma(\mathbf{Y}^{\circ}_{t-1}))] = \gamma_0 + \boldsymbol{\gamma}'E[\mathbf{u}^2_{t-1}] = \gamma_0 + \sigma^2\boldsymbol{\iota}'\boldsymbol{\gamma}. \tag{4.14}$$

Substituting (4.14) into the variance identity (4.6), we obtain,

$$\sigma_{11} = \gamma_0/(1-\boldsymbol{\iota}'\boldsymbol{\gamma}) + \boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}, \tag{4.15}$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)'$, $\boldsymbol{\iota}$ is a px1 vector of ones, and $\mathbf{u}^2_{t-1} = (u^2_{t-1}, \ldots, u^2_{t-p})'$. Thus, the ARCH-GARCH conditional variance parameters do not have a unique statistical representation in terms of the unconditional moments because of the restrictions imposed by their definitions on the probabilistic structure of $\{y_t, t \in \mathbf{N}\}$.[18]

The unconditional distributional properties for constant conditional mean ARCH-type formulations are considered, *inter alia,* by Engle (1982), Milhøj (1985). The result in (4.15) differs because it is not assumed that $E(y_t \mid \sigma(\mathbf{Y}^{\circ}_{t-1}))$ is constant. The potential non-existence of the unconditional variance implied by (4.15) is problematic, since the conditional mean parameters (β_0 and $\boldsymbol{\beta}$) can not even be defined in that case.

These issues do not arise for the STAR model, however, because it is specified directly in terms of the distributional properties of $\{y_t, t \in \mathbf{N}\}$. The assumption of stationarity implies stability of both the conditional mean and variance, and ω^2_t has a well defined limit $E[\mathrm{Var}(y_t \mid \sigma(\mathbf{Y}^{\circ}_{t-1}))] \xrightarrow{} \nu\sigma^2/(\nu-2)$, ensuring that $\mathrm{Var}(y_t) \xrightarrow{} \nu\sigma_{11}/(\nu-2)$, $\nu > 2$ (Spanos (1992b)).

The distributional assumption of conditional normality for the ARCH-type model is shown to be inappropriate by Spanos (1992a). Following Nimmo-Smith (1979), and Cambanis, Huang, and Simmons (1981), Spanos (1992a) proves that the linearity of the implied reverse regression (under stationarity) for an autoregressive model is both necessary and sufficient for the underlying joint distribution to be a member of the elliptic family. Spanos (1922a) then notes that Kelker (1970) shows that the only elliptic distribution with a homoskedastic conditional variance is the normal. Hence, the joint (and conditional) distribution for a stationary, dynamic model of heteroskedasticity must be a non-normal member of the elliptical family.

Bollerslev (1987), Baillie and Bollerslev (1989), and Hsieh (1989), suggest estimating ARCH-type models based on non-normal conditional distributions, such as the Student's t. The formula used in these papers for the Student's t distribution of u_t is,

$$D(u_t | Y_{t-1}^{p}; \theta_1) = \frac{\Gamma\left[\frac{1}{2}(\nu+1)\right]}{\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}\nu\right]}[(\nu-2)h_t^2]^{-\frac{1}{2}}\left[1+\frac{u_t^2}{(\nu-2)h_t^2}\right]^{-\frac{1}{2}(\nu+1)}, \tag{4.16}$$

which is not the same as the Markov Student's t conditional distribution formula,

$$D(u_t | Y_{t-1}^{p}; \theta_2) = \frac{\Gamma\left[\frac{1}{2}(\nu+p+1)\right]}{\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}(\nu+p)\right]}[\nu\sigma^2 d_t^2]^{-\frac{1}{2}}\left[1+\frac{u_t^2}{\nu\sigma^2 d_t^2}\right]^{-\frac{1}{2}(\nu+p+1)}, \tag{4.17}$$

(Zellner (1971)), or the non-Markov form in (4.9).

The formula in (4.16) can be obtained by substituting h_t^2 into the marginal Student's t distribution and re-arranging the scale parameter (Raiffer and Schlaifer (1961)). In contrast, (4.17) is derived from the joint distribution $D(y_t, \ldots, y_{t-p}; \phi)$. Comparing the two formulae we can see that the degrees of freedom parameter, ν, enters (4.17) separately in the gamma functions, but as a product with σ^2 in the other terms. Thus, estimation of both ν and σ^2 is precluded (Zellner (1976)). Any attempt to estimate ν, ignoring σ^2, will result in the estimation of an inappropriate mixture of both (see McGuirk et.al., (1992)).

For comparison purposes, a number of alternative ARCH-type formulations were estimated using both the normal and "conditional Student's t" error distribution assumptions (see also McGuirk et.al., (1992)). The results are largely invariant to the choice of functional form, distribution, and lag structure, so I will only report the estimation results from a first-order GARCH model under the assumption of conditional normality, and using two lags in the conditional mean. The first-order GARCH formula corresponds to the form typically used in literature (see Bollerslev et.al., (1992)). The estimates for this model and misspecification tests based on the standardized residuals $\hat{u}_t/\hat{h}_t$ are summarized in Table 4.4.

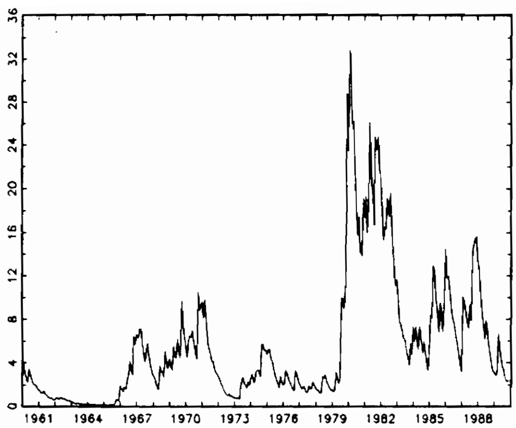
The estimation results indicate that most of the conditional mean parameters are significant at conventional levels. The conditional variance parameter estimates satisfy the constraint $\alpha_1 + \gamma_1 = 1$, suggesting the presence of so-called "integrated-GARCH" effects (Engle and Bollerslev (1986)). Unit roots in the conditional variance is viewed as an important issue because it implies that shocks to the conditional variance persist, such that m-period-ahead conditional variance forecasts depend on $h_t^2(\alpha_1+\gamma_1)^m$, and does not vanish as $m \rightarrow \infty$ (see Engle and Bollerslev (1986), Nelson (1991a)). Hence, the GARCH conditional variance will have a very long memory and be sensitive to the initial conditions.

As noted in Section 4.3, the question of a unit root and persistence in the conditional variance of the STAR model did not arise, and both the conditional mean and variance had well defined probability limits (see Spanos (1992b)).

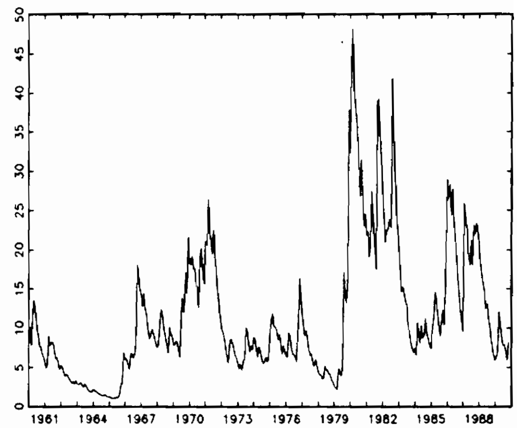
Table 4.4 - Normal GARCH (1,1) Estimation Results			
	y_{1t}	y_{2t}	y_{3t}
$\hat{\beta}_0$	0.077 (.049)	0.222 (.123)	0.615 (.233)*
$\hat{\beta}_1$	0.499 (.058)*	0.339 (.058)*	0.296 (.060)*
$\hat{\beta}_2$	-0.175 (.057)*	-0.134 (.056)*	-0.011 (.059)
$\hat{\alpha}_0$	0.007 (.008)	0.115 (.098)	1.017 (.457)*
$\hat{\alpha}_1$	0.225 (.042)*	0.171 (.042)*	0.220 (.055)*
$\hat{\gamma}_1$	0.815 (.026)*	0.837 (.031)*	0.777 (.043)*
LogL	-704.140	-898.483	-1124.833
P-values for Misspecification tests			
DS	0.086	0.042*	0.023*
DK	0.000*	0.001*	0.000*
KG	0.541	0.338	0.304
AC(2)	0.379	0.891	0.572
AC(4)	0.006*	0.056	0.597
LB(24)	0.089	0.507	0.741
A(2)	0.227	0.090	0.899
A(4)	0.021*	0.014*	0.836
ML(24)	0.609	0.491	0.994

The misspecification tests for non-correlation and non-linear dependence, reported in Table 4.4, indicate that the GARCH models for y_{1t} and y_{2t} have not adequately accounted for the dynamics in the data. This problem was not alleviated by increasing the lag length in the conditional mean. One possible explanation for this behavior may be the failure of the GARCH model specification to explicitly incorporate the interdependence between the distributional parameters and the coefficients of the model, and between the conditional mean and the conditional variance. The GARCH residuals also indicate unmodeled excess kurtosis and some skewness for all the series. Interestingly, the "conditional Student's t" GARCH estimates of the degrees of freedom parameter were 5.645, 7.432, and 3.636 respectively, but the corresponding residual kurtosis estimates were even greater than for the conditional normal GARCH residuals.

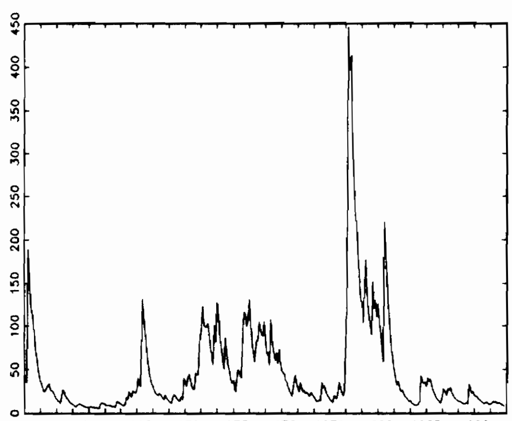
The GARCH conditional variance estimates are presented in Figure 4.5. The similarity between these estimates and the corresponding sequences of 12-period window (rolling) LS estimates for each series in Figure 4.6 is apparent. The GARCH model necessarily assigns a very high weight to large squared deviations in the residuals because the variability predictions are made only in terms of local information. Thus, any cumulative learning in the decision making process underlying the data is ruled out by construction (see Diebold (1988)). Consequently, the 1979-81 period is attributed only short-term importance relative to the rest of the 1980's, and the stock market crash is given a large weight in the long-term bond estimates. Also, the GARCH conditional variances are always substantially larger and more variable than the corresponding STAR model estimates. The disparity between the conditional variance estimates of the two models occurs because it is the smoothed, cumulative effects, that determine volatility predictions from the STAR model.



(a) y_{1t}

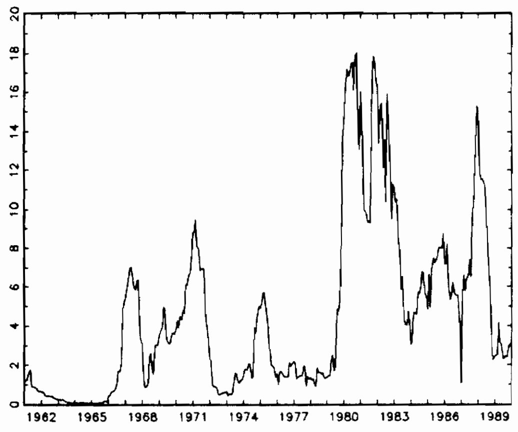


(b) y_{2t}

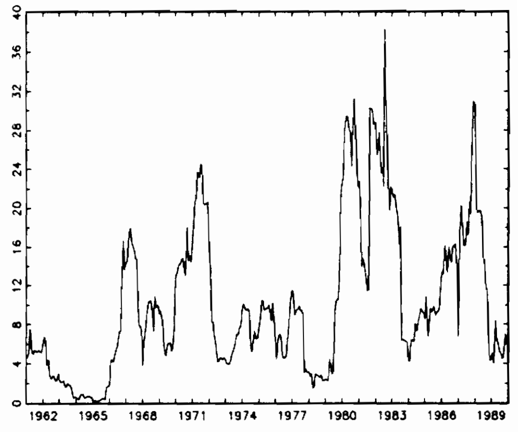


(c) y_{3t}

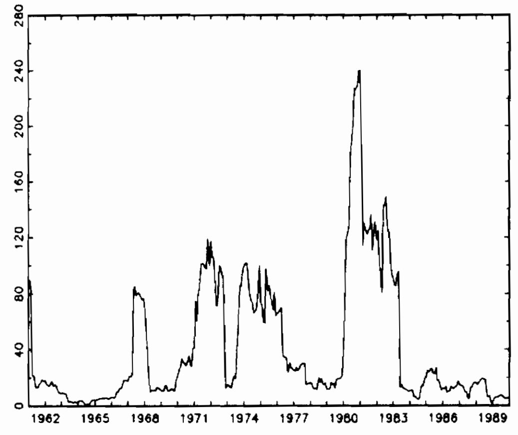
Figure 4.5 - Normal GARCH(1,1) Conditional Variance Estimates



(a) y_{1t}



(b) y_{2t}



(c) y_{3t}

Figure 4.6 - 12-Period Window LS Conditional Variance Estimates

4.5 Conclusions.

In this paper I considered the modeling of volatility in U.S. interest rates. Modeling volatility is an important topic because, for example, the question of whether long-term yields are too volatile to be consistent with standard models of the term structure depends, *inter alia*, on the way volatility is statistically characterized.

Based on evidence conveyed by preliminary data analysis, the STAR model was estimated for the log differences in three interest rate series. The estimation and misspecification testing results suggested that the STAR model adequately accounted for the probabilistic features of the data: bell-shape symmetry; leptokurtosis; first and second-order temporal dependence.

The STAR model was then compared with various ARCH-type formulations. It was found that these alternative models gave rise to apparent long memory (unit roots) in the conditional variance, and failed a number of diagnostic tests. It was also noted that the "Student's t" ARCH-type model was based on an inappropriate formula.

From an economic perspective, the STAR model estimates indicate that risk premia in the term structure may not be as large or variable as suggested by ARCH-type formulations. The information structure of the STAR model also implies that the ability to distinguish between systematic and unsystematic volatility increases over time as more observations became available. High frequency variation (relative to the given information set) will always be essentially unpredictable. However, the effects of systematic variability are built into the optimal volatility predictions through the recursive up-dating scheme. For example, the change in monetary policy in 1979-80 had a quite different effect on volatility predictions than the 1987 stock market crash. An implication of this learning behavior is that changes in the financial environment can be modeled by allowing the conditional moments to adapt over time as a function of the conditioning information.

Appendix 4A

STAR Model Misspecification Tests

The basic assumptions underlying the STAR model are:

(i) $D(y_t | \mathbf{X}_t; \Psi)$ is Student's t;

(ii) $E[y_t | \sigma(\mathbf{X}_t)] = \beta'\mathbf{X}_t$ (linear in $\mathbf{X}_t$);

(iii) $Var[y_t | \sigma(\mathbf{X}_t)] = \omega_t$ (heteroskedastic);

(iv) v_t is a martingale difference process;

where $\mathbf{X}_t = (y_{t-1}, \ldots, y_{t-\ell})'$, $v_t = (y_t - \beta'\mathbf{X}_t + \omega^2{}_t\epsilon_t)/\omega_t$ is the generalized error, $\epsilon_t \sim$ Stiid(0,1), and $\sigma(\mathbf{X}_t)$ denotes the sigma-field generated by $\mathbf{X}_t$ (see Spanos (1992b)). Most of the misspecification tests for this model are based on auxiliary regressions as discussed in the multivariate context in Chapter 2. In particular, KG refers to an F-form of the second-order Kolmogorov-Gabor polynomial test for conditional mean non-linearity (Spanos (1986), p.460). AC are F-forms of the modified Lagrange multiplier tests for autocorrelation (Spanos (1986), p.521)), and HC is a modification of the Engle (1982) ARCH test, which includes the squares and cross-products of the regressors under the null, to be consistent with the form of heteroskedasticity implied by the Students t distribution. In general, F-forms of these tests are preferred because they attempt to account for the degrees of freedom used in estimation. The LB and ML statistics are reported as additional tests for temporal dependence in the residuals, as described in the text. DS and DK denote the D'Agostino skewness and kurtosis tests respectively, which are approximately distributed N(0,1) under the null (see Appendix 2C).

Notes

1. The popularity of modeling non-linear dynamics through the conditional variance arises in part from the relationship in economic theory between conditional moments and the determination of asset prices. Additive, non-linear stochastic generating mechanisms, such as the Granger and Anderson (1978) bi-linear model represents an alternative approach (Hsieh (1989a)).

2. Pagan and Schwert (1990) compare several formulations of dynamic heteroskedasticity, including a number of non-parametric variance estimates such as kernel estimates and flexible Fourier forms.

3. The data are from the *CitiBank Economic Database* tapes. The non-stationary nature of U.S. interest rates over the sample period has been documented elsewhere (see Engle and Granger (1987), Campbell and Shiller (1990)). In terms of the statistical adequacy of the resulting models, differencing, rather than including time trends, was found to be a preferable method of achieving stationarity. Taking logarithms of interest rates is not uncontroversial but has been frequently used in approximations to linear term structure models (see Shiller (1989, p.291)).

4. The bi-variate estimate uses a standard normal kernel with the same correlations as the data, a bandwidth of 0.45, and evaluated on a 51 by 51 grid; see Silverman (1986) for details.

5. The reference densities use a correlation of 0.35.

6. The sample moment estimates should be interpreted with some caution because of the memory in the data.

7. The form of the scaling factor $(\upsilon/\upsilon+t-3)$ in the conditional variance suggests that the issue of non- existence of the conditional moments does not arise if $\nu>2$. For the multivariate Student's t distribution, all the even conditional moments are multiples of the conditional variance, and all of the odd conditional moments are zero (see Spanos (1992b)).

8. See Spanos (1992b) for the corresponding statistical parameterization in the dynamic case.

9. For ease of notation it is assumed here that $p = \ell$.

10. Actually, the covariance matrix is $\boldsymbol{\Omega} = \nu\boldsymbol{\Sigma}/(\nu-2)$, $\boldsymbol{V} = \nu\boldsymbol{\Sigma}$ (compare with the static case). As discussed in Spanos (1992), the dimensionality of the V_T matrix, and the relative spareness of Q_T renders direct estimation of the joint likelihood function infeasible in general. However, in principle, the joint likelihood function can be used to obtain estimates of the unconditional moments, and the coefficients of the STAR model found be using the implied statistical parameterization, and the invariance property of ML.

11. As shown in Spanos (1991), standard ML theory applies for this estimation problem.

12. The weighted residuals are defined as $\hat{u}_t/\hat{\omega}_t - \hat{\omega}_t\epsilon_t$, where $\hat{u}_t = y_t - \hat{\beta}_0 - \Sigma_{i=1}^{\ell}\hat{\beta}_i y_{t-i}$, and $\epsilon_t \sim St(0,1;\upsilon)$ is a simulated standard i.i.d. Student's t series. These weighted residuals are a modification of the usual (Pearson) residuals, where the additional term purports to account for the non-linear effects of the conditional variance on y_t. Since the long-run conditional variance is $[(\nu/(\nu-2)]\sigma^2$, the additional term will be asymptotically negligible (see Spanos (1992b)). The Student's t (pseudo) random numbers were generated using the Kinderman and Monahan's ratio method (Algorithm 1.7 (Dagpunar (1988)), and the GAUSS uniform number generator. This algorithm is part of a collection of univariate and multivariate random number generators contained in the menu-driven program RANDOM.

13. Homer and Sylla (1991) provide an excellent detailed historical summary of interest rates over the sample period.

14. The recursive OLS estimate is based on a sequence of up-dated one-step-ahead mean-square errors, and its use in describing volatility can be traced back to Mandelbrot (1963). Both the recursive OLS and τ-period window (rolling) OLS estimates presented below utilize a second-order conditional mean specification.

15. Engle (1982) notes that the ARCH(p) function may be viewed as an approximation to a more complex model with a non-ARCH conditional variance function. He mentions, but does not pursue, possible generalizations of this function that would also allow lagged regressors to enter the conditional variance function.

16. ARCH-type formulations are often referred to as: (a) models of heteroskedasticity (Engle and Bollerslev (1986)); (b) models of time varying conditional variances (Baillie and Bollerslev (1992)); or (c) as models of non-linear dependence, where the non-linearity enters via $u_t = h_t\epsilon_t$, $\epsilon_t \sim \text{IID}(0,1)$ (Hsieh (1989)). However, it is important to distinguish between non-constancy of the conditional variance over time due to functional dependence on the conditioning variables (heteroskedasticity) or non-stationarity of the unconditional moments (Spanos (1990a)).

17. Recall, stationarity refers to the absence of time dependence of the moments of the joint distribution of the observable processes involved.

18. See Spanos (1992b) for a further discussion of the nature of these restrictions in the context of the STAR model.

References

Alogoskoufis, G.S., and R. Smith. (1991): "The Phillips Curve, the Persistence of Inflation, and the Lucas Critique: Evidence from Exchange Rate Regimes," *Amer. Econ. Rev.*, 81(5), 1254-1275.

Alogoskoufis, G.S., and R. Smith. (1991a): "On Error Correction Models: Specification, Interpretation, and Estimation," *J. Econ. Surveys*, 5(1), 97-128.

Altonji, J. (1982): "The Intertemporal Substitution Model of Labor Market Fluctuations," *Rev. Econ. Stud.*, 49, 783-824.

Amemiya, T. (1985): *Advanced Econometrics*. Cambridge, MA, Harvard U. Press.

Anderson, T.W. (1971): *The Statistical Analysis of Time Series*. N.Y., John Wiley.

Anderson, T.W. (1984): *An Introduction to Multivariate Statistical Analysis*. N.Y., John Wiley.

Anderson, T.W., and H. Rubin. (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Ann. Math. Statist.*, 20(1), 46-63.

Andrews, D. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817-858.

Anscombe, F.G., and W.J. Glynn. (1983): "Distribution of the Kurtosis Statistic b_2 for Normal Statistics," *Biometrika*, 70(1), 227-234.

Aptech Systems (1992): *GAUSS 3.0 Manual*. Kent, Aptech Systems Inc.

Ashenfelter, O., and D. Card. (1982): "Time Series Representations of Economic Variables and Alternative Models of the Labor Market," *Rev. Econ. Stud.*, 49, 761-782.

Baillie, R.T. (1989): "Econometric Tests of Rationality and Market Efficient," *Econometric Rev.*, 8, 151-186.

Balanda, K.P., and H.L. MacGillivray. (1988): "Kurtosis: A Critical Review," *Amer. Statistician*, 42(2), 111-119.

Barra, J.R. (1981): *Mathematical Basis for Statistics*. N.Y., Academic Press.

Bartlett, M.S. (1947): "Multivariate Analysis," *J. Roy. Stat. Soc.*, A., 2, 176-197.

Basmann, R.L. (1960): "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistic." *J. Amer. Statist. Assoc.*, 55(3), 650-659.

Bera, A.K., and C.M. Jarque. (1982): "Model Specification Tests: A Simultaneous Approach," *J. Econometrics*, 20, 59-82.

Bera, A.K., and S. John. (1983): "Tests for Multivariate Normality with Pearson Alternatives," *Comm. Statist. - Theor. Meth.*, 12(1), 103-107.

Bera, A.K., and C.R. McKenzie. (1986): "Tests for Normality With Stable Alternatives," *J. Statist. Comput. Simul.*, 25, 37-52.

Beran R. (1977): "Adaptive Estimates for Autoregressive Processes," *Ann. Inst. Statist. Math.*, 28(1), 77-89.

Berndt, E.R., and N.E. Savin. (1977): "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," *Econometrica*, 45(5), 1263-1277.

Blanchard, O.J. (1986): "Empirical Structural Evidence on Wages, Prices and Employment in the U.S.," Working Paper 2044, National Bureau of Economic Research.

Blanchard, O.J., and S. Fischer. (1989): *Lectures on Macroeconomics*. Cambridge (Mass), M.I.T. Press.

Breusch, T.S., and A.R. Pagan. (1979): "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, 47(5), 1287-1294.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *J. Econometrics*, 31, 307-327.

Bollerslev, T. (1987): "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Rev. Econ. Statist.*, 542-547.

Bollerslev, T., R. Chou, and K. Kroner. (1992): "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *J. Econometrics*, 52, 5-60.

Bowden, R.J., and D.A. Turkington. (1984): *Instrumental Variables*. Cambridge (U.K.): Cambridge U. Press.

Bowman, K.O., and L.R. Shenton. (1975): "Omnibus Test Contours for Departures From Normality Based on $\sqrt{b_1}$ and b_2," *Biometrika*, 62(2), 243-249.

Box, G.E.P. (1949): "Generalized Distribution Theory for a Class of Likelihood Criteria," *Biometrika*, 36(3), 317-346.

Brown, R.L., J. Durbin and J.M. Evans. (1975): "Techniques for Testing the Constancy of Regression Relationships Over Time," *J. Roy. Statist. Soc.*, B, 49-92.

Cambanis, S., Huang, S., and G. Simmons. (1981): "On the Theory of Elliptically Contoured Distributions," *J. Mult. Anal.*, 10(2), 343-350.

Cameron, A.C., and R.R. Trivedi. (1990): "Conditional Moment Tests and Orthogonal Polynomials," unpublished manuscript, Dept. Econ., U. of California, Davis.

Campbell, J.Y., and R.J. Shiller. (1990): "Yield Spreads and Interest Rate Movements: A Bird's Eye View," Unpublished Paper, Princeton University.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *J. Econometrics*, 34, 305-334.

Chesher, A. (1983): "The Information Matrix Test: Simplified Calculation via a Score Test Interpretation," *Econ. Letters*, 13(1), 45-48.

Chesher, A., and R. Spady. (1991): "Asymptotic Expansions of the Information Matrix Test," *Econometrica*, 55(5), 787-815.

Christiano, L.J., and M. Eichenbaum. (1992): "Current Real Business Cycle Theories and Aggregate Labor-Market Fluctuations," *Amer. Econ. Rev.*, 82(3), 430-450.

Cox, D.R., and N.J.H. Small. (1978): "Testing Multivariate Normality," *Biometrika*, 65(2), 263-272.

Cox, D.R., and D.V. Hinkley. (1974): *Theoretical Statistics*. London, Chapman Hall.

Dagpunar, J. (1988): *Principles of Random Number Generation*, Oxford, Oxford U. Press.

D'Agostino, R.B. (1970): "Transformation to Normality of the Null Distribution of g_1," *Biometrika*, 57, 679-681.

D'Agostino, R.B., A. Belanger, and R.B. D'Agostino, Jr. (1990): "A Suggestion for Using Powerful and Informative Tests for Normality," *Amer. Statistician*, 44(4), 316-321.

D'Agostino, R., and E.S. Pearson. (1973): "Tests for Departures from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$," *Biometrika*, 60(3), 613-622.

D'Agostino, R., and E.S. Pearson. (1974): "Correction", *Biometrika*, 61(3), 647.

D'Agostino, R.B., and M.A. Stephens. (1986): *Goodness of Fit Techniques*. New York: Marcel Dekker.

Davis, P.J. (1975): *Interpolation and Approximation*. Waltham, Blaisdell.

Desai, M. (1984): "Wages, Prices and Unemployment a Quarter of a Century after the Phillips Curve," in *Econometrics and Quantitative Economics*, ed. D.F. Hendry, and K.F. Wallis. Oxford, Basil Blackwell.

Dhrymes, P.J. (1978): *Introductory Econometrics*. N.Y., Springer Verlag.

Diebold, F.X. (1988): *Empirical Modeling of Exchange Rate Dynamics*, N.Y., Springer Verlag.

Donner, A., and J.F. McCollum. (1972): "The Phillips Curve: An Historical Note," *Economica*, 39(155), 322-323.

Durbin, J. (1970): "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38(3), 410-421.

Engle, R.F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987-1008.

Engle, R.F., and C.W.J. Granger. (1987): "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251-276.

Engle, R.F., D.F. Hendry, and J.F. Richard. (1983): "Exogeneity," *Econometrica*, 51(2), 277-304.

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

Fama, E.F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Evidence," *J. Finance*, 25, 319-338.

Fama, E.F. (1976): "Inflation Uncertainty and the Expected Returns on Treasury Bills," *J. Pol. Econ.*, 84, 427-448.

Fama, E.F. (1984): "Term Premiums in Bond Returns," *J. Financial Econ.*, 13, 529-546.

Fischer, S. (1977): "Long Term Contracts, Rational Expectations and the Optimal Money Supply Rule," *J. Pol. Econ.*, 85(1), 191-205.

Fisher, R.A. (1930): "The Moments of the Distribution for Normal Samples of Measures of Departure from Normality," *Proc. Roy. Soc., A*, 130, 16.

Friedman, M. (1966): "Comments," in *Guidelines, Informal Controls, and the market Place: Policy Choices in a Full Employment Economy*, ed. G.P. Schultz and R.Z. Aliber. Chicago, U. Chicago, Press.

Friedman, M. (1968): " The Role of Monetary Policy," *Amer. Econ. Rev.*, 58(1), 1-17.

Giri, N.C. (1977): *Multivariate Statistical Inference*. N.Y., Academic Press.

Godfrey, L.G. (1988): *Misspecification Testing in Econometrics*. Cambridge, Cambridge U. Press.

Godfrey, L.G., and M.R. Wickens. (1981): "Testing Linear and Log-Linear Regressions for Functional Form," *Rev. Econ. Stud.*, 48(3), 487-496.

Gordon, R. (1983): "A Century of Evidence on Wage and Price Stickiness in the United States, The United Kingdom, and Japan," in *Macroeconomics, Prices and Quantities*, ed. J. Tobin. Washington D.C., Brookings Inst.

Gregory, A.W., and M.R. Veall. (1985): "Formulating Wald Tests of Nonlinear Restrictions," *Econometrica*, 38(3), 682-704.

Hall, A.D. (1987): "The Information Matrix Test for the Linear Model," *Rev. Econ. Stud.*, 54(2), 257-263.

Hall, S.G. (1986): "An Application of the Granger and Engle Two-step Estimation Procedure to United Kingdom Aggregate Wage Data," *Oxf. B. Econ. Stat.*, 48(3), 229-240.

Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50(4), 1029-1054.

Hansen, L.P., and R.J. Hodrick. (1980): "Forward exchange rates as Optimal Predictors of Future Spot Rates: An Econometric Investigation," *J. Pol. Econ.*, 88, 829-853.

Hansen, L.P., and K.J. Singleton. (1982): "Generalized Instrumental Variables Estimation of Non-linear Rational Expectations Models," *Econometrica*, 50(4), 1269-1286.

Hart, M.L., and A.H. Money. (1976): "On Wilks Multivariate Generalization of the Correlation Ratio," *Biometrika*, 63(1), 59-67.

Hatton, T.J. (1988): "A Quarterly Model of the Labor Market in Interwar Britain," *Oxf. B. Econ. Stat.*, 50(1), 1-25.

Hausman, J.A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251-1270.

Hendry, D.F. (1991): *PC-GIVE: An Interactive Econometric Modelling System*. Institute of Economics and Statistics and Nuffield College, University of Oxford.

Hendry, D.F., and G.E. Mizon. (1978): "Serial Correlation as a Convenient Simplification Not a Nuisance: A Comment on the Study of the Demand for Money by the Bank of England," *Econ J.*, 88, 549-563.

Hendry, D.F., A.J. Neale, and F. Srba. (1988): "Econometric Analysis of Small Linear Systems Using PC-FIML," *J. Econometrics*, 38, 203-226.

Homer, S., and R. Sylla. (1991): "*A History of Interest Rates*," New Brunswick, Rutgers U. Press.

Hong, C. (1988): "The Integrated Generalized Autoregressive Conditional Heteroskedastic Model: The Process, Estimation and Monte Carlo Experiment," Unpublished Paper, Department of Economics, UCSD.

Hotelling, H. (1931): "The Generalization of Student's ratio," *Ann. Math. Statist.*, 2(2), 360-378.

Hooper, J.W. (1959): "Simultaneous Equations and Canonical Correlation Theory," *Econometrica*, 27(3), 245-256.

Hsiao, C. (1983): "Identification," in *Handbook of Econometrics, 1*, ed. Z. Griliches, and M.D. Intriligator. N.Y., North-Holland.

Hsieh, D.A. (1989a) "Testing for Nonlinear Dependence in Daily Foreign Exchange Rates," *Journal of Business*, 62, 339-368.

Ivakhnenko, A.G. (1984): "Past, Present, and Future of GMDH", in *Self-organizing Methods in Modeling*, Ed.: S.J. Fallow. N.Y., Marcel Dekker.

Jarque, C.M., and A.K. Bera. (1987): "A Test for Normality of Observations and Regression Residuals," *Int. Statist. Rev.*, 55(2), 163-172.

Johnson, N.L. (1949): "Systems of Frequency Curves Generated by the Methods of Translation," *Biometrika*, 36(1), 149-176.

Johnson, N.L., and S. Kotz. (1972): *Distributions in Statistics: Continuous Multivariate Distributions*. N.Y., John Wiley.

Johnston, J. (1984): *Econometric Methods*. N.Y., McGraw-Hill.

Kelker, D. (1970): "Distribution Theory of Spherical Distributions and a Location-scale Parameter Generalization," *Sankhyā*, 32(3), 419-430.

Kendall, M., A. Stuart, and J.K. Ord. (1987): *Advanced Theory of Statistics, Vol 1*. Oxford, Oxford U. Press.

Kennan. J., and G.R. Neumann. (1988): "Why Does the Information Matrix Test Reject Too Often? A Diagnosis of Some Monte Carlo Symptoms," unpublished manuscript, Hoover Institution, Stanford U.

Kiefer, N.M., and M. Salmon. (1983): "Testing Normality in Econometric Models," *Econ. Letters*, 11(1), 123-127.

Kres. H. (1983): *Statistical Tables for Multivariate Analysis*. N.Y., Springer Verlag.

Kydland, F.E., and E.C. Prescott. (1980): "A Competitive Theory of Fluctuations and the Feasibility and Desirability of Stabilization Policy," in *Rational Expectations and Economic Policy*, ed. S. Fischer, Chicago, U. Chicago Press.

Lawley, D.N. (1938): "A Generalization of Fisher's z Test," *Biometrika*, 30, 180-187.

Layard, R.G., Nickell, S., and R. Jackman. (1991): *Unemployment*. Oxford, Oxford U. Press.

Lipsey, R.G. (1960): "The Relationship Between Unemployment and the Rate of Change of Money Wage Rates in the U.K., 1862-1957," *Economica*, 27(105), 1-31.

Lipsey, R.G., and J.M. Parkin. (1970): "Incomes Policy: A Reappraisal," *Economica*, 37(146), 115-138.

Lucas, R.E. (1973): "Some International Evidence on Output-Inflation Tradeoffs," *Amer. Econ. Rev.*, 63(3), 326-334.

Lucas, R.E., and L.A. Rapping. (1969): "Price Expectations and the Phillips Curve," *Amer. Econ. Rev.*, 59(3), 342-351.

Lung, G.M., and G.E.P. Box. (1978): "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65, 297-303.

Lütkepohl, H. (1991): *Introduction to Multiple Time Series Analysis*. N.Y., Springer Verlag.

MacKinnon, J.G. (1992): "Model Specification Tests and Artificial Regressions," *J. Econ. Lit.*, 30(1), 102-146.

Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *J. Business*, 394-419.

Mankiw, N.G., and L.H. Summers. (1984): "Do Long-Term Interest Rates Overreact to Short-Term Interest Rates," *Brookings Papers on Economic Activity*, 223-247.

Manski, C.F. (1988): *Analog Estimation Methods in Econometrics*. N.Y., Chapman and Hall.

Mardia, K.V. (1970): "Measures of Multivariate Skewness and Kurtosis With Applications," *Biometrika*, 57(3), 519-530.

Mardia, K.V. (1974): "Applications of Some Measures of Multivariate Skewness and Kurtosis for Testing Normality and Robustness Studies," *Sankhyā*, A, 36, 115-128.

Mardia, K.V. (1980): "Tests for Univariate and Multivariate Normality," in *Handbook of Statistics, Vol. 1*, Ed.: P.R. Krishnaiah. Amsterdam, North-Holland.

Mardia, K.V., and K. Foster. (1983): "Omnibus Tests for MultiNormality Based on Skewness and Kurtosis," *Comm. Statist. - Theor. Meth.*, 12(2), 207-221.

McCulloch, J.H. (1985): "Interest Rate Sensitive Deposit Insurance Premia: Stable ACH Estimates," *J. Banking and Finance*, 9, 137-156.

McGuirk, A., J. Robertson, and A. Spanos. (1991): "Modeling Exchange Rate Dynamics: Nonlinear Dependence and Thick Tails," unpublished manuscript, Dept. Econ., Virginia Tech.

McLeod, A.I., and W.K. Li. (1983): "Diagnostic Checking ARMA Time Series Models Using Squared-residual Autocorrelations," *J. Time series Anal.*, 4(3), 239-273.

Myers, R.H. (1990): *Classical and Modern Regression with Applications*, Boston, PWS-Kent.

Muirhead, R.J. (1982): *Aspects of Multivariate Statistical Theory*. New York, John Wiley.

Nanda, D.N. (1950): "Distribution of the Sum of Roots of a Determintal Equation Under a Certain Condition," *Ann. Math. Statist.*, 21(3), 432-439.

Nelson, D.B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 318-334.

Newey, W.K. (1985): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53(5), 1047-1070.

Newey, W.K. (1985a): "Generalized Method of Moments Specification Testing," *J. Econometrics*, 29, 229-256.

Newey W.K., and K.D. West. (1987): "A Simple, Positive Semi-definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703-708.

Nimmo-Smith, I. (1979): "Linear Regressions and Sphericity," *Biometrika*, 13(1), 25-45.

Nymoen, R. (1988): "Modelling Wages in a Small Open Economy: An Error-Correction Model of Norwegian Manufacturing Wages," *Oxf. B. Econ. Stat.*, 5193), 239-258.

Olson, C.L. (1974): "Comparative Robustness of Six tests in Multivariate Analysis of Variance," *J. Amer. Statist. Assoc.*, 69(348), 894-908.

Ord, J.K. (1968): "Approximations to Distribution Functions Which are Hypergeometric series," *Biometrika*, 55, 243-250.

Pagan, A.R. (1984): "Model Evaluation by Variable Addition," in *Econometrics and Quantitative Economics*, Eds.: Hendry, D.F., and K.F. Wallis. Oxford, Basil Blackwell.

Pagan, A.R., and A.D. Hall. (1983): "Diagnostic Tests as Residual Analysis," *Econometric Rev*, 2, 159-218.

Pagan, A.R., and Y. Pak. (1991): "Testing for Heteroskedasticity," unpublished manuscript, Dept. Econ., U. Rochester.

Pagan, A.R. and G.W. Schwert. (1990): "Conditional Stock Volatility," *J. Econometrics*, 45, 267-290.

Pagan, A.R., and F. Vella. (1989): "Diagnostic Tests for Models Based on Individual Data: A Survey," *J. Applied Econometrics*, 4(Supplement), S29-S59.

Pagan, A.R., and M.R. Wickens. (1988): "A Survey of Recent Advances in Econometrics," *Econ. J.*, 99(398), 962-1026.

Pearson, E.S. (1930): "A Further Development of Tests of Normality," *Biometrika*, 22, 239-249.

Pearson, E.S. (1965): "Tables for Percentage Points of $\sqrt{b_1}$ and b_2 in Normal Samples: A Rounding Off," *Biometrika*, 52, 282-285.

Pearson, E.S., R.B. D'Agostino, and K.O. Bowman. (1977): "Tests for Departures From Normality: Comparison of Powers," *Biometrika*, 64(2), 231-246.

Pesaran. M.H. (1987): *The Limits to Rational Expectations*. Oxford, Basil Blackwell.

Pesaran, M.H., and B. Pesaran. (1990): *Microfit: An Interactive Econometric Software Package*. Oxford, Oxford U. Press.

Phelps, E.S. (1968): "Money Wage Dynamics and Labor Market Equilibrium," *J. Pol. Econ.*, 76(4), 678-711.

Phelps, E.S. (1972): *Inflation Policy and Unemployment Theory*. London, MacMillan.

Phillips, A.W. (1954): "Stabilization Policy in a Closed Economy," *Econ. J.*, 64(2), 290-323.

Phillips, A.W. (1958): "The Relation Between Unemployment and the Rate of Change of Money Wages in the United Kingdom; 1861-1957," *Economica*, 24(100), 283-299.

Pierce, A.D. (1982): "The Asymptotic Effect of Substituting Estimators for Parameters in Certain types of Statistics," *Ann. Statist.*, 10, 475-478.

Pillia, K.C.S. (1955): "Some New Test Criteria in Multivariate Analysis," *Biometrika*, 43(2), 122-127.

Poterba, J., Rothemberg, J., and L. Summers. (1985): "A Tax-based Test for Nominal Rigidities," working paper 1627, National Bureau of Economic Research.

Ramsey, J.B. (1969): "Tests for Specification Errors in Classical Linear Least-squares Regression Analysis," *J. Roy. Statist. Soc.*, B, 31(2), 350-371.

Rao, C.R. (1951): "An Asymptotic Expansion of the Distribution of Wilk's Criterion," *Bull. Int. Statist. Inst.* 33(2), 177-180.

Rivlin, T.J. (1969): *An Introduction to the Approximation of Functions*. Waltham, Blaisdell.

Rudd, P.A. (1984): "Tests of Specification in Econometrics," *Econometric Rev.*, 3(2), 211-242.

Sargan, J.D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26(3), 393-415.

Sargan, J.D. (1964): "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in *Econometric Analysis for National Economic Planning*, ed. Hart, P.E., Mills, G., and J.K. Whitaker. London, Butterworth.

Sargent, T.S. (1987): *Macroeconomic Theory*. N.Y., Academic Press.

Schatzoff, M. (1966): "Sensitivity Comparisons Among Tests of the General Linear Hypothesis," *J. Amer. Statist. Assoc.*, 64, 415-435.

Shenton, L.R., and K.O. Bowman. (1975): "Johnson's S_U and the Kurtosis Statistics," *J. Amer. Statist. Assoc.*, 70(2), 220-228.

Shiller, R.J. (1979): "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," *J. Pol. Econ.*, 87, 1190-1219.

Shiller, R.J. (1989): *Market Volatility*, Cambridge (Mass.), MIT Press.

Shiryayev, A.N. (1984): *Probability*. N.Y., Springer Verlag.

Silverman, B.W. (1986): *Density Estimation for Statistics and Data Analysis*. London, Chapman and Hall.

Sims. C.A. (1980: "Macroeconomics and reality," *Econometrica*, 48(1), 1-48.

Small, N.J.H. (1980): "Marginal Skewness and Kurtosis in Testing Multivariate Normality," *Appl. Statist.*, 29(1), 85-87.

Spanos, A. (1986): *Statistical Foundations of Econometric Modelling*. Cambridge, Cambridge U. Press.

Spanos, A. (1987): "Error Autocorrelation Revisited: The AR(1) Case," *Econometric Reviews*, 6(2), 285-294.

Spanos, A. (1989): "On Re-reading Haavelmo," *Econometric Theory*, 5, 405-429.

Spanos, A. (1990): "The Simultaneous-Equations Model Revisited: Statistical Adequacy and Identification," *J. Econometrics*, 44, 87-105.

Spanos, A. (1990a): "Unit Roots and Their Dependence on the Conditioning Information Set," *Advances in Econometrics*, 8, 271-292.

Spanos, A. (1991): "A Parametric Approach to Heteroskedasticity: The Student's t and Elliptic Linear Regression Models," unpublished manuscript, Virginia Tech.

Spanos, A. (1992): *An Introduction to Modern Econometrics*. Cambridge, Cambridge U. Press.

Spanos, A. (1992a): "Normality and Linear Regression Models" Karl Pearson's Thesis Revisited," unpublished manuscript, Virginia Tech.

Spanos, A. (1992b): "The Student's t Autoregressive Model with Dynamic Heteroskedasticity," unpublished manuscript, Virginia Tech.

Tauchen, G.E. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *J. Econometrics*, 30, 415-443.

Taylor, J. (1979): "Staggered Price setting in a Macro Model," *Amer. Econ. Rev.*, 69(2), 108-113.

Taylor, J. (1980): "Aggregate Dynamics and Staggered Contracts," *J. Pol. Econ.*, 88, 1-24.

Weiss, A.A. (1984): "ARMA Models with ARCH Errors," *J. Time Series Anal.*, 5, 129-143.

White, H. (1980): "A Heteroskedastic-Consistent Covariance Matrix estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817-838.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1), 1-25.

White, H. (1987): "Specification Testing in Dynamic Models," in *Advances in Econometrics - Fifth World Congress, 1*. Ed.: T. Bewley. Cambridge, Cambridge U. Press.

White, H., and G. MacDonald. (1980): "Some Large Sample Tests for Nonnormality in the Linear regression Model," *J. Amer. Statist. Assoc.*, 75(369), 16-28.

Wilks, S.S. (1932): "Certain Generalizations in the Analysis of Variance," *Biometrika*, 24(4), 471-494.

Wilson, E.B., and M.M. Hilferty. (1931): "The Distribution of Chi-square," *Proc. Nat. Acad. Sci.*, 17, 684-688.

Wooldridge, J.M. (1990): "A Unified Approach to Robust, Regression-based Specification Tests," *Econometric theory*, 6(1), 17-43.

Zeldes, S.P. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation," *J. Pol. Econ.*, 97(2), 305-346.

Zellner, A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregated Bids," *J. Amer. Statist. Assoc.*, 57, 348-368.

Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics*, N.Y., John Wiley.

Zellner, A. (1976) "Bayesian and Non-Bayesian Analysis of the Regression Model With Multivariate Student-t Error Terms," *J. Amer. Statist. Assoc.*, 71, 400-405.

Vita

John Robertson was born in Dunedin, New Zealand in 1962, the third of four children of James and Dawn Robertson. John completed his Bachelor and Masters degrees in Economics at Lincoln College of the University of Canterbury; the latter under the supervision of Bert D. Ward. In 1986 John was employed by the Agribusiness and Economics Research Unit at Lincoln College, and in 1987 was awarded a Reserve Bank of New Zealand Fellowship that allowed him to begin Ph.D. studies at Virginia Tech, beginning in the Fall of 1988. A further Reserve Bank Fellowship was granted in 1989. In June 1991 John married Debra Thunberg of Milford, New Hampshire, and in July 1992 he will take up a lecturing position in the Department of Statistics at the Australian National University.