THEORIES AND TECHNIQUES IN HOUSING MARKET ANALYSIS

by

Joseph A Langsam

B.Sc. Massachusetts Institue of Technology
(1968)

M.Sc. University of Michigan, Ann Arbor
(1981)

Ph.D. (Mathematics) University of Michigan, Ann Arbor
(1982)

Submitted to the Department of
Urban Studies and Planning
and to the
Department of Economics
in Partial Fulfillment of the
Requirements of the
Degree of

DOCTORATE IN PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MAY 1983

Signature of Author: ______________________
Department of Urban Studies and Planning
May 1983

Certified by: ______________________
Thesis Supervisor

Certified by: ______________________
Second Reader

Certified by: ______________________
Third Reader

Accepted by: ______________________
Head Ph.D. Committee, Department of Urban Studies and Planning

Accepted by: ______________________
Head Ph.D. Committee, Department of Economics

THEORIES AND TECHNIQUES

IN HOUSING MARKET ANALYSIS

by

JOSEPH A. LANGSAM

Submitted to the Department of Urban Studies and Planning and Economics on May 1, 1982 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

ABSTRACT

Housing market analysis whether from the vantage of urban planning or economics presents both methodological and theoretical problems. The housing market is characterized by search, while market data is frequently only available in cross sectional and aggregated formats. This dissertation contains three principle results which should be for use in housing market analysis. In the area of search theory, it is shown that a search model can have an equilibrium price vector where a commodity can have a nondegenerate equilibrium price distribution. A simple one period urn type search model is analysized and the conditions under which buyers or sellers are made better off by market replication are determined. The buyers bid problem is analysized and it is shown that the bid structure need not be monotonic with respect to time.

In the area of estimation and hypothesis testing two results are developed. It is shown that an iterative weighted least squares estimator converges in the sense that for a fixed sample the iterates converge almost surely and also in the sense that the estimator constructed by taking these limit points converges to the true value of the parameters being estimated and possesses other optimal properties. This analysis corrects an error that appears in the article by Oberhoffer and Kmenta. The final result is the analysis of a multistage heteroskedastic estimator which enables the consistent estimation and hypothesis testing on the structure of the heteroskedasticity. This

procedure is a computationally simple procedure for performing estimation and hypothesis testing on both the underlying model and on the parameters generating the heteroskedastic structure. The procedure presented in this essay, unlike that which appears in Glejser and Parks papers leads to consistent estimation and consistent hypothesis tests.

The dissertation begins with a short introduction to the problems in housing to which the theories and methodologics developed in the thesis can be directed. The principle results are presented in the second and third chapters without their proofs. The mathematical proofs are separated out and presented in the fourth chapter so that the thesis can be used by those researchers whose mathematical interests are minimal. The fourth chapter should be of interest to those who are interested in the application of functional analytic tools to regression theory.

Thesis Supervisor: Professor William Wheaton

TO MY PARENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank a number of people who over a ten year span have assisted and supported my efforts to complete this dissertation. My parents, Sid and Helen Langsam gave both moral and financial support to a degree beyond that even expected by a hopeful son. They remained supportive even, when after five years of study in economics and urban studies, I switched fields to earn a Ph.D. in Mathematics. To them, I owe a great debt that cannot be repaid.

I am indebted to my advisor Professors Wheaton and Hausmann and to Professor Diamond. They have been most helpful in assisting me to improve the thesis and to navigate the various rules and regulations associated with earning a joint degree. I would also like to acknowledge kindness and the assistance of Professor Eckaus. That this dissertation is too wordy is my fault; I am indebted to Professor Fisher that it is not even more so.

I would like to gratefully acknowledge the many mathematicians at M.I.T. and the University of Michigan who spent the time to teach me the value of mathematical analysis and rigorous thinking. I would publically like to thank my colleagues in the Department of Mathematics at Case Western Reserve University who have listened with understanding to my stories about writing a second dissertation.

I want to thank and to acknowledge the contributions of my wife, Betty and my children, Daniel and Jessica. Without them, this thesis would never have been written and my life could never have been as happy.

CHAPTER I

INTRODUCTION

1. Introduction. The core of the dissertation is comprised of three distinct investigations. The first of these is directed towards problems arising in search theoretic economic models. A reasonable definition of equilibrium is presented and an existence theorem is proven. A simple one period search model is described and the welfare implications associated with increasing the market size through replication are analysized. The question of optimal bidding in a sequential search model is discussed and an existence theorem for optimal bids is proven. An example is then given which shows that with costly sampling the optimal bid profile need not be monotonic.

The second and third investigations are concerned with estimation and hypothesis testing in a linear model in the presence of heteroskedasticity. In the second investigation an iterative weighted least square estimation for a linear model with block scalar variance convariance matrix is described and analysized. A theorem giving sufficient conditions for the successive iterates associated with a fixed sample to form a Cauchy sequence almost surely as the sample size increases is proven. This theorem corrects an error in the paper by Oberhofer and Kmenta [33] where the argument showing convergence of the successive iterate is faulty. The estimator constructed by taking the limits of the successive iterates is shown to have desirable asymptotic properties.

The final investigation is directed towards estimation and hypothesis testing when the variances follow a linear model. A simple procedure is described and analysized for estimating and performing hypothesis tests on the variance model and then for estimating and performing hypothesis tests on the original model. Unlike the procedure given in Glejser [14], the asymptotic distribution of the estimator for the variance model can be computed. The proofs of the theorems showing consistency of this estimator and hypothesis tests is an extension of the work given in White [59]. It is not an immediately consequence of White's procedure in that his procedure would require inputs that are not observable. This proposed multiple step procedure is then shown to generate an estimator for the underlying linear model with optimal asymptotic properties and with easily computed asymptotic distributions. It is a simple consequence of the theorems proven, that this procedure generates a simple test for heteroskedasticity.

The motivation for these essays comes from problems encountered in urban planning, most particularly in housing and manpower planning. The labor and housing markets are ones in which search plays an important role. They are also markets in which metropolitan and regional data is collected in a cross sectional rather than a time series formating. Thus heteroskedasticity is likely to be a major problem in the analysis of the regional and metropolitan data. This dissertation is intended to be a contribution in both urban studies and economics. For this reason, the dissertation contains a brief discussion of urban problems towards which the theoretical ideas developed in the main body can be applied. A brief discussion of housing discrimination is given since programs directed at housing desegregation tend to aggregate the market and market aggregation or replication is a subject of the first investigation. The attempts to estimate housing demand elasticities is briefly reviewed since this estimation will often be done in models with heteroskedasticity present. Because of the varied purposes towards which the dissertation is directed, the thesis is organized to present the major results first qualitatively, then quanitatively with proofs only sketched, and finally with full mathematical rigour with complete proofs. While this leads to unfortunate redundancy, it does provide the policy maker, technician, and theoretician with the level of generality appropriate for their needs.

The remainder of this chapter contains the above mentioned discourse on housing discrimination and on estimation of housing demand elasticities. The second chapter contains the search related results. It begins with a brief review of known results followed by a description of equilibrium in search models and a theorem giving conditions for the equilibrium to existence. A simple consequence of this theorem is that in a search model identical commodities need not have the same equilibrium price. A simple housing search model is presented and the welfare implications of market replication are analysized. The chapter ends with a discussion of optimal bidding in a search model containing an existence proof and an example showing that the optimal bid profile need not be monotone with time. The third chapter begins with a general discussion of the nature of the heteroskedasticity problem. The iterative least squares estimator for the block scalar variance covariance matrix is then analysized. Theorems giving condtion for the convergence of successive iterates and for the optimal asymptotic properties of the estimator are stated. White's procedure is briefly described followed by the description of a new procedures for estimation and hypothesis tests for models where the variance structure follows a linear model. From this procedure an estimator and hypothesis tests for the variance model are developed. A theorem giving the properties of these tests and estimator is stated. As a corollary a simple new test for heteroskedasticity is presented. Also from this procedure a multiple stage estimator and hypothesis tests for the basic model are developed. A theorem giving conditions for these tests and estimator to have desirable optimal asymptotic properties is then stated. The fourth chapter contains a restatement of the theorem of the first two chapters together with complete proofs. It should be of interest to those interested in the application of functional analytic technique to statistical problems. The last chapter contains concluding remarks and suggestions for future applied and theoretical research.

2. Discrimination in Housing. The nation has adopted as policy goals the desegregation of residential neighborhoods and, the increase of housing consumption by low income families. Federal, state, and local governments have instituted a variety of programs seeking these goals including programs which provide: rent subsidies, subsidized new construction, legislative relief through zoning, and market information services. These programs are designed to relieve some barriers that the planner perceives to generate segregated housing patterns. Underlying the choice of programmatic relief must be a theory of market behavior. Since resources are limited, one naturally tries to select those programs which are most cost effective. To do this requires a sufficient knowledge of economic theory and of empirical techniques. In [103, p.480], Stevens outlines the problem when she states:

Studies of housing demand in the United States have found significant differences between the behavior of white majority and black minority households. In particular, blacks choose housing of a different tenure class mix, quality, and location from white households. These differences in demands have been said to depend on any or all of the following: (1) blacks' preferences differ from those of whites; (2) blacks consume different quantities and qualities of housing than do whites because blacks face price and entry discrimination in the housing market; and (3) blacks' income, both current and permanent, is lower than whites', causing blacks to consume less housing services than do whites.

Identification of the most important cause of the difference between black and white housing consumption patterns is important in devising a housing policy to meet national housing goals. If discrimination is the most important cause of low housing consumption by blacks, then an open housing policy is indicated. If differences in consumption patterns are mainly attributable to differences in current or permanent income, then transfers or policies to increase job skills, labor mobility and employment quality will achieve housing goals. Finally, if tastes differ, housing vouchers or other housing subsidies may be the only feasible way to induce some households to consume an "adequate" level of housing services, however such level of services is defined.

Most previous work has found that price and entry discrimination exists in the housing market. Blacks on the average earn smaller incomes than do whites, whether income is measured on a current or a permanent basis. There is, however, disagreement on the portion of demand differences which can be attributed to each of these factors.

The possible barriers which preclude integrated neighborhoods as a result of market forces include:

differences in tastes between blacks and whites resulting in differing preferences for housing consumption, differing tastes resulting in whites strongly preferring self association, disciminatory practices on the part of sellers and brokers, historically generated endowment differences between blacks and whites, transportation network limitations that result in costly commuting between certain residences and certain job sites, and historically generated housing and work place locational differences between blacks and whites that preclude the free flow of market information. Clearly any subset of these may cause segregated housing patterns. The planner is confronted with the problem of selecting those which make the greatest contribution to market segregation. To choose among these factors those that are dominant, requires both a market structure theory and a means of empirically estimating market parameters. In the area of residential segregation, much effort has been made in determining the importance of these barriers and one can find some of the results in [62], [64], [65], [68], [73], [74], [78], [81], [89], [90], [92], [97], [100], [106], and [108].

A cental issue is planning to desegregate residential neighborhoods is wherer segregation results from pure preference considerations or from discriminatory practices in the market place. Pure preference considerations can generate segregated housing if either there are housing consumption preference and endowment differences between blacks and whites or if racial considerations enter directly into preference structures. Kern in [81] using an equilibrium analysis in a housing market model is able to show: that if whites' preference for whites is stronger than blacks' preference for blacks, an integrated equilibrium is unstable and a segregated equilibrium is stable, if blacks prefer whites greater than whites prefer whites no segregated equilibrium exists and a stable integrated equilibrium exists. In the integrated equilibrium all sites have idential racial composition and therefore racial composition has no effect on equilibrium rent distance function. In the segregated equilibrium where whites prefer whites and blacks prefer blacks equilibrium rents on the white's side of the boarder may exceed, equal or fall short of that on the black side. In the segregated equilibrium where both races prefer white neighborhoods, rents on the white side of boarder exceeds that on the black side provided there are no discriminatory practices. Farley and Bianchi in [73] report a survey that suggests that whites prefer whites while blacks prefer a 50-50 integrated neighborhood. Thus in a segregated equilibrium one expects in the absence of discriminatory practices, that segregated white residential rents will exceed black rents. Miesykowski and Syron in [89] summarize current economic housing market theory and their findings show that income differences are a small factor leading to segregation, pure preference differences are a significant factor and lead to whites paying a premium for segregation, and that whites overtly discriminate against blacks resulting in both higher housing prices and limited job opportunities for blacks. Follain and Malprezzi in [74] attempts to empirically test whether blacks pay a premium. They estimate a hedonic model in which race is a variable and using micro level survey data show that blacks receive a discount of about 15% in owner occupied units and 6% in renter markets. This supports a finding that pure preferences are a dominant factor in determining market segregation.

The question of whether segregation is a result of pure preferences or discrimination is important in a number of areas. If segregation results from preferences of self association and if residential housing market segregation does not cause disadvantage in other markets, programs to force integration may result in everyone being worse off. There is evidence, see for example [92], that housing segregation leads to blacks being at a disadvantage in the labor market. In this case, one must understand that improvement in blacks' welfare in the labor market resulting from integration of residence are traded off against losses from not being able to self-associate. If, however, segregation is the result of discriminatory practices, programmatic relief might include both legislative and compensatory programs. Both policy goals and programmatic content are affected by the identification of those barriers which generate residential segregation.

Whatever the cause of neighborhood segregation, programs which successfully reduce segregation have the effect of aggregating several nearly independent submarkets into a larger housing market. In many instances a close examination of the housing market reveals that it is comprised of nearly independent submarkets. Some of the factors which lead to this market segmentation include: strong ethnic self-association preferences, transportation networks which makes interzonal commuting costly, physical barriers such as rivers, parklands and large expressways, and historically generated governmental structures. Frequently programs which are directed to these factors have as a secondary effect changes in market segmentation.

In markets in which search is characteristic there are frictional costs associated with search; that is costs required to obtain price information and costs associated with making decisions without full knowledge of price structures. The aggregation ob submarkets into a larger market will change these frictional costs. In the first essay, we perform a partial equilibrium analysis for a specialized type of market aggregation to determine the effect upon frictional costs associated with market aggregate.

An equivalent formulation of the problem of aggregating several identical markets is to consider the replication of a single market. This latter approach is easier to deal with analytically and is the approach taken in the first essay. In the market are m buyers and n sellers. When the market is replicated x times there are, of course, xm buyers and xn sellers. Each seller has one unit for sale, the i^{th} buyer has a potential bid of Y_{ij} for the j^{th} unit. In the replicated market the i_r buyer has a potential bid of Y_{ij} for the j_t unit where $1 \leqq r \leqq x$ and $1 \leqq t \leqq x$. The j^{th} seller has a reservation price or minimum acceptance price of X_j; in the replicated market the reservation price held by the j_t seller is X_j. All analysis for a single fixed time period. In this time period each buyer independent of other buyers selects from a uniform probability distribution exactly one seller to visit. Buyer I_i will purchase the unit owned by seller J_j if buyer i visits seller J_j, $Y_{i,j} \geqq X_j$ and either:

1) $Y_{i,j} > Y_{i',j}$ for all other buyers $I_{i'}$, that visit J_j
2) $Y_{i,j} \geqq Y_{i',j}$ for all buyers $I_{i'}$, that visit J_j

and buyer I_i wins the toss of a fair s sided die, where s is the number of buyers $I_{i'}$ that visit seller J_j with bid $Y_{i',j} = Y_{i,j}$. The same transaction rules apply to the replicated market.

Since we are interested in the effect upon frictional costs associated with market segmentation and aggregation, bid and reservation prices are held fixed. In this model, the frictional cost in the replicated market to the buyer I_i, can be measured by $P(x,i)$, the probability of making a purchase. The frictional cost in the non-replicated market is given by $P(1,i)$. The frictional costs to the seller J_j in the replicated market has two measures: $Q(x,j)$, the probability of making a sale, and $E(x,j)$, the expected value of a sale given that one is made.

The first step in analyzing the impact an frictional costs associated with aggregation is to compute $P(x,i)$, $Q(x,j)$, and $E(x,j)$. Before giving their values, it is necessary to introduce some additional notation. For an arbitrary set S, let $||S||$ denote its cardinality, let $G_j(r)$ be the fraction of buyer whose bid for the j^{th} unit does not exceed r, that is $G_j(r) = ||\{I_i : Y_{i,j} < r\}||/m$.

Let $B_j(r)$ be the fraction of buyer whose bid for the j^{th} unit equals r, and let $H_j(r)$ be the fraction of buyers whose bid for the j^{th} unit is at least r. Let $F(r)$ be the fraction of sellers whose reservation price for this unit does not exceed r, that is $F(r) = ||\{J_j : X_j < r||\}/n$. Let C_j be the set of buyers whose bid for the unit owned by J_j is at least as great as its reservation price. Finally, let $\alpha = m/n$ be the ratio of buyers to sellers and let $u(x) = \frac{nx-1}{nx}$. It is important to note that those measures that are ratios do not change when the market is replicated.

Using the above notation it is shown that

$$P(x,i) = \sum_{j \varepsilon A_i} \frac{1}{mB_j(Y_{i,j})} [\mu(x)^{xm[H_j(Y_{i,j}) - B_j(Y_{i,j})]}$$

$$(x)^{xm[H_j(Y_{i,j})]}$$

$$Q(x,j) - 1 - \mu(x)^{xmH_j(X_j)}$$

$$E(x,j) = \sum_{i \varepsilon C_j}$$

$$\frac{Y_{i,j}}{mB_j(Y_{i,j})} \frac{[\mu(x)^{xm[H_j(Y_{i,j}) - B_j(Y_{i,j})]} - \mu(x)^{xm[H_j(Y_{i,j})]}}{1 - (x)^{xmH_j(X_j)}}$$

While the number of times x that a market may be replicated is integer valued $(x>0)$, the expression $P(x,i)$, $Q(x,j)$ and $E(x,j)$ have for each i and j natural extension to smooth function in the x variable. To determine the effects of frictional costs associated with aggregation one can examine the value of

$\frac{\partial P(x,i)}{\partial x}$, $\frac{\partial Q(x,i)}{\partial x}$, and $\frac{\partial E(x,j)}{\partial x}$ for $x>1$. We see that for a fixed j if there are at least two buyers with different bids for the $j^{th/}$ unit at least as large as X_j and if $n_x>2$, then both $\frac{\partial Q(x,j)}{\partial x} < 0$ and $\frac{\partial E(x,j)}{\partial x} < 0$. Thus in almost every case the sellers frictional costs are increased by aggregating equivalent submarkets. For the buyer I_i, the results are not as conclusive, however, the following can be shown:

1) sufficient conditions for $\frac{\partial P}{\partial X}(x,i) > 0$ are that $nx>2$, $A_i \neq \emptyset$, and $\alpha > \sup[H_j(Y_{i,j}) - B_j(Y_{i,j})]^{-1}$

2) a sufficient condition for $\frac{\partial P}{\partial X}(x,i) < 0$ are that $nx>2$.

   $A_i \neq \emptyset$, and $\alpha < \frac{-1}{nx \ln\frac{(nx-1)}{(nx)}}$

3) if $nx>2$ and $\alpha<5/7$, then $\frac{\partial P(x,i)<0}{\partial x}$.

As a consequence of 2 above in markets with a large number of sellers and in which sellers outnumber buyers it will in almost every case be to the benefit of the buyer not to have market aggregation. In general it appears that maintaining segmented markets helps both the buyers and sellers by restricting the competition that buyers feel from other buyers and sellers from other sellers. The exception occurs for the buyer who tends to be a low bidder for each unit he sees and who is in a market with many more buyers than sellers, in this case the low bid buyer would prefer aggregation. In this case, it appears that aggregation gives the low bidder more opportunities for finding a unit for which he is the only bidder.

The above analysis has clear limitations. In addition to assuming that the segmented markets are identical, it is a single period model without dynamic features. Nevertheless, it provides a framework for illustrating the importance of frictional costs that arise wherever search is a consideration. In particular it alerts the planner and program analyst to welfare factors that should be considered whenever programs directly or indirectly affect the parameter of search including market segmentation.

3. Estimation of Housing Related Elasticities. An essential part of the planning process, especially in housing and manpower planning, is the identification

of the current status of various market parameters. In the field of housing analysis and planning the important parameter includes price and income elasticities of demand and price elasticity of supply. Recall that if f is a function from R^n into R^1, $X^\circ=(X_1^\circ,X_2^\circ,\ldots,X_n^\circ)$ is a point in R^n, and $Y=f(x)$, then the elasticity of Y with respect to X_j at the point X° is denoted by $\eta_{Y,X_j}(X^\circ)$ is given by $\eta_{Y,X_j}(X^\circ) = \frac{\partial f}{\partial X_j}(X^\circ)\,\frac{X_j^\circ}{f(X^\circ)}$. The elasticity is a measure of the percentage change in output for a percentage change in an input.

Much of the current debate among housing policy analysts center on whether programs for low income housing should feature supply or demand side subsidies. Central to this debate are questions concerning supply and demand elasticities. If the income elasticity of demand is high and the price elasticity of supply is near zero, then income transfer programs that raise the incomes of low income families will tend to just raise housing prices. If on the other hand, the price elasticity of supply is very large, such programs will then result in low income families consuming more housing without large increases in the price of housing. If the price elasticity of demand is near zero, supply subsidies designed to lower the price of housing will not result in substantial increases in housing consumption.

Knowledge of these elasticities is also important in designing programs addressed to residential segregation. Suppose higher quality, higher priced housing is found in predominately white neighborhoods and lower quality lower priced housing is found in racially mixed neighborhoods. If low income whites have a higher income elasticity of demand than low income non-whites, an income transfer program that raises the income of low income families will result in white families leaving the integrated neighborhoods to move to the substantially white neighborhoods. If the non-white income elasticity of demand is the larger, then the same program will result in non-whites moving into segregated neighborhoods and thus increase the amount of residential integration.

The usual procedure for estimating part or all of an unknown parameter victor β in R^K is to specify (assume) a linear relation $Y = X\beta+\varepsilon$ where the vector Y and data matrix X is observable. The vector β is estimated by $b_{01s} = (X^TX)^{-1} X^TY$ and the variance of b_{ols} is estimated by

$$s^2 = \frac{(Y-Xb_{ols})^T (Y-Xb_{ols}) (X^TX)^{-1}}{n-K}$$ where Y is a $n\times1$ vector (n is the number of observations).

Among the assumptions which are implicitly made in using this procedure are that the error term has zero mean and follows a homoskedastic distribution, that is $E(\varepsilon\ \varepsilon^T) = \sigma^2\ I$, where σ is a positive scalar. Under the usual OLS assumption, it is not difficult to show that the estimation b of β and s of σ^2 have certain desirable properties including:

1) b_{ols} and s^2 are consistent
2) in the event that the data matrix X is non-stochastic, b_{ols} is a best linear unbiased estimator of $\beta(b_{ols}$ in BLUE).
3) b_{ols} is asymptotically efficient.

If, however, the error term ε has zero mean, but follows a heteroskedastic distribution, $E(\varepsilon\ \varepsilon^t)$ is a non-scalar diagonal matrix, then while b_{ols} is still consistent, efficiency is lost and s^2 is not a consistent estimator of σ^2 so, in particular, the usual hypothesis tests based on s^2 should not be performed.

Heteroskedasticity may be introduced into a model because of both theoretical considerations, and as a result of measurement procedures. The most obvious way in which heteroskedasticity is introduced is through data grouping. If the model is $Y_{ij} = X_{ij}\beta + \varepsilon_{ij}$ where ε_{ij} are homoskedastic, $(1 < i < n,\ 1 < j < m_i)$ and the estimated model uses group averages, that is the estimated model is $Y_i = X_i\beta + \varepsilon_i$ where $Y_i = \frac{1}{m_i}\Sigma_{j=;}^{m_i} Y_{ij}$, $X_i = \frac{1}{m_i}\Sigma_{j=1}^{m_i} X_{ij}$ then the ε_i follows a heteroscedastic distribution. If the numbers m_i are known, one can perform weighted least squares by weighting the vector (Y_i, X_i) by $\sqrt{m_i}$ and thus regain some of the optional propertiesof OLS. If m_i are now known, one should perform a weighted least squares procedure which first estimates the m_i and then use these estimates to weight the observations. Heteroscedasticity can also be present for purely theoretical considerations. The best example is the problems associated with estimating consumption or demand functions. In the simplest case, see [18, 48] consumption is assumed to be a linear function of income and the model $C_t = A_0 + B\,Y_t + \varepsilon_t$ is estimated where C is consumption and Y is income. It is observed that the variance of ε_t varies with Y_t, and it is easy to develop a theoretical explanation for this phenomenom. For an understanding of the consumption income relationship as well as for prediction purposes it is important to analyze and estimate this variance-income relationship (see [12] for an introduction to the relevance of the heteroskedastic structure to the original model).

Heteroskedasticity is likely to be a problem in the estimation of housing demand elasticities both because of grouping and because of the theoretical structure of the error component. There are in the literature a number of attempts to estimate price and income elasticities of demand and own price elasticities of supply with wide ranging results. (See, for example, [66], [70], [72], [80], [82], [84], [85], [87], [88], [94], [35], [101], [102] and [107]. The following table gives an overview of the wise range of estimates that have been presented. It is difficult to analytically compare the estimates of different authors since they use different specifications and different measures for the various variables. To the extent, however, that the concepts of income elasticity and price elasticity has developed into policy and program planning variables, it is interesting to see the variance among these estimates.

	n_y	n_p
Dusenberg and Kristin, 1953, [70][1]	.15 (6)	-.078
Lee, 1968, [84]	.8 owners .58 renters	
Winger, 1968, [107]	1.03	
Maisel, Burnham and Austin, 1971, [88]	.45	-.89
de Leeuw, 1971, [85]	.7 -1.5 owners .8 -1.0 renters	
Carliner, 1973, [66]	.50 renters	
King, 1976, [82][2]	.64	
Polinsky, 1977, [94]	.75	
Stegman and Sumka 1978, [102]	.251 current income .337 -.400 (permanent income) .195 (Black Families)	
Polinsky and Elwood, 1979 [35]	.39 .57	-.67 micro -.72 grouped
McRae and Tuner, 1981 [87]	.25	-.89

1) Their reported elasticity is the coefficient of a linear demand equation. The imputed elasticity would be 0.6.

2) King estimates a Lancastrian demand model and the value 0.6 reported in the table is the imputed

elasticity of demand for space with respect to income.

A central question in the estimation of income demand elasticity is how to measure income. In general micro level household income demand is not collected. Furthermore it is not clear what concept of income, permanent or current should be used. Polinsky and Ellwood [35] attempt an analysis of the specification error associated with various choices of measures of income and with using micro verses grouped data. This paper critiques the earlier work of Carliner [68], Lee [84], Maisel, Burnham, and Austin [88], and Winger [107] and shows that much of the divergence of income and price elasticities can be explained by mispecification of the income variable. While Polinsky and Ellwood observe that heteroskedasicity will occur due to grouping, they fail to adjust for its presence in the micro model estimation and assume in the grouped estimation model that it arises only from grouping.

In the last two technical essays two procedures for estimation and hypothesis testing in the presence of heteroskedasicity are presented and analyzed. In the first of these the variance covariance matrix is assumed to be block scalar with a fixed number of blocks. The model to be estimated could be of the form $Y_{ij} = X_{ij}\beta + \varepsilon_{ij}$ where $1 < i < m$, $1 < j < m_i$ and variance of $\varepsilon_{ij} = \sigma_i^2$ $(\sigma_i^2 > 0)$. The estimation scheme presented is an interactive weighted least square procedure. In the first stage σ_i^2 is estimated by 1 and β_1 is computed using weighted least squares (weights = 1). In the $r+1^{\underline{st}}$ stage σ_i^2 is estimated by $\frac{1}{n_i} \Sigma_{j=1}(Y_{ij} - X_{ij}\beta_r)^2$ and β_{r+1} is computed using the weights derived from the estimated σ_i^2. In the ensuing analysis it is shown that the sequence β_r converges almost sure as $\max_i(n_i) \to \infty$, it is also shown that the estimates for the variance also converge. The resulting estimation of β and σ_i^2 are shown to have the usual desirable asymptotic properties, the estimator for B being asymptotically equivalent to weighted least squares with known true weights. It is also shown that the usual hypothesis tests using the estimates in the weighted model using the weights derived from the estimator σ_i^2 has the usual known asymptotic distribution. This procedure of interactive weighted least squares is appropriate when using group data or when the variance of the error term is thought to be related to the values of a discrete variable.

In the last essay we present a procedure for estimation and hypothesis testing in the model $Y = X\beta+\varepsilon$ where the variance covariance model is diagonal with diagonal vector σ^2, where $\sigma^2 = Z\Gamma$, and where Z is observed and Γ unknown. Starting with the works of Glejser and White we develop an easy to apply multiple step estimation procedure that not only permits hypothesis testing on the estimated Γ but also yields an estimator for β which is asymptotically equivalent to weighted squares with known weights and for which the usual standard hypothesis tests are valid. In the first stage β is estimated by b_{ols}, the ordinary least squares estimator and Γ is estimated by $\gamma = (Z^TZ)^{-1} Z^Te^2$, where e^2 is the vector whose j^{th} element is $(Y_j-X_j\, b_{ols})^2$.

Asymptotically valid hypothesis tests can be performed on Γ using the estimated γ and using $\frac{1}{n}\Sigma_{1=1} W_1^{\,2}\, Z_i^T\, Z_i$ as an estimator of the variance covariance matrix where $W_i^{\,2} = (e_i^{\,2} - Z_i\gamma)^2$. The estimated variances are computed by $\hat{\sigma}^2 = Z\gamma$ and weighed least squares are used to reestimate β. The statistics reported from the ordinary least square regression of the weighted model has the expected known asymptotic distribution.

Heteroskedasticity has been an often ignored problem in estimation and hypothesis testing associated with urban economic analysis. In particular it will be present in the estimation of demand elasticities, which are themselves important variables in the formulation of urban housing policies. This dissertation presents two procedures for handling the heteroskedasticity problem. The first of these is appropriate when the variance covariance matrix of the error term can be put in a block scalar form, for example when the variance is related to the values of a variable with finite discrete range. The second procedure is appropriate whenever the variance of the error term is a linear function of observable variables. This latter procedure is particularly appropriate for use in estimating demand elasticities where the variance of the error term is likely to be a function of income.

There are a number of research directions suggested by the works in the dissertation. In the areas of search theory, one should begin to examine the impact of aggregating non-identical markets and should begin to develop multiple period search models in which there is intertemporal dynamics. In particular one wants to investigate how search affects long run housing patterns. In the area of estimation and hypothesis testing in the presence of heteroskedasticity, both the estimators presented here should be compared to other procedures to determine their relative powers. In particular the interactive scheme presented here can be compared to the simple one interation weighted least squares procedure in a Monte Carlo study. Clearly, the next obvious step in the estimation of housing demand elasticity is to redo the study of Polinsky and Ellwood using the heteroskedastic correct procedure developed in the last essay. It will be interesting to learn whether the divergence among elasticity estimations can in part be the failure to correct for heteroskedasticity.

CHAPTER II

MARKET SEARCH

1. Introduction. In markets with search, either consumers or producers are making decisions with less than full information about commodities and prices. Buyers or sellers are making decisions based upon expected prices and not upon a commonly observed market price. Thus, *a priori*, one need not expect that a commodity in a search model would have a well defined price. Therefore consumers and producers are faced with not only the decision of how much to buy and sell, but also if, when, and where to make these transactions.

Both housing and labor markets are classical examples of search markets. The prospective house buyer usually is unaware of the available stock, its quality and its price while the prospective seller is unaware of the potential market demand. In labor markets, the employer is uncertain of prospective skills of an applicant and of the minimal wage acceptable to the applicant. The applicant is unaware of both job openings and of their potential wage rates. While much of the work done in search theory has drawn from labor market observations for its motivation, many of the results are immediately transferable to analysis of housing markets.

The remainder of the chapter is divided into four sections. In the next section is a brief review of search theory literature with most of the articles concentrating on search in labor markets. The third section contains a discussion of equilibria in search module together with a tentative definition of equilibrium and an existance proof. The fourth section contains a simple one period housing search model which is used to analysize the effects upon buyers and sellers resulting from duplicating the market. The last section contains a discussion of optimal bidding in a search framework.

2. Review of Search Theory Literature. Search is a feature in every market. In almost every market where the cost of search is sufficiently high, transactions are made with the participants possessing less than full information. In some markets, the marginal cost of search is so sufficiently small that behavior in this market is perturbed but slightly

from that in a deterministic market with perfect information. In deterministic full information models, however, it is difficult to support non-degenerate price distribution for homogeneous goods, support un- and underemployment of resources, and support the existence of advertising, while in a search model these phenomenon are natural consequences. Despite the obvious importance of search in economic analysis it has only recently been developed in the literature. Economic search literature seems to owe its origin to the two papers by Stigler [44,45]. In these papers, Stigler argues that non-degenerate price distribution might be supportable if the cost of obtaining price information is high. In the later paper, he presents a job search model that will support the job seekers accepting a wage less than the maximum available in the economy. It has been shown, however, that the search strategy presented in this model is suboptimal.

Much of the literature subsequent to Stigler's premier articles can be divided into two classes; optimal search strategy, and existence of non-degenerate equilibrium price and wage distributions. The optimal search or optimal stopping time theory that appears in economics also appears in sequential analysis, in statistics, and in the study of stopping times and smartingales in probability. It should probably be a meta theorem that any theory that appears semi independently in three fields has relevance. In any event, optimal search theories are making a contribution in explaining labor market behavior.

Much of the optimal search literature has regarded the job seeker's and employer's problem as distinct and separate.*

The job seeker's problem is taken to be some variation of the following. The job seeker samples sequentially from the distribution of wage offers incurring a cost for each sample and seeks a rule to tell him when to stop searching and start working. The employer's problem is taken to be a variation of the following. The employer sequentially observes job seekers with a particular marginal product from a distribution incurring a cost for sampling. The employer is seeking a rule telling him when to stop sampling and make a wage offer w. It is surprising

*For the purpose of simplicity, the language of labor market analysis is adopted throught this section.

that these problems are not consistent. It is assumed that if the employer makes offer, it will be accepted while the job seeker's problem is when to accept an offer. Furthermore in most of the literature, the employer is assumed to have a fixed wage w, so that his decision is only to offer or not offer a job. The complexities of the analysis depends upon the assumptions that are made on the objectives being optimized and the learning process. Some of the variations that have been analyzed in the optimal search strategy literature include assumptions of infinite time horizons no discounting, finite time horizon positive time discount, random number of job offers at each period, underlying wage distribution known, underlying wage distribution learned through a Bayesian process, risk adversion, and wealth constraints with bankruptcy. The literature has also addressed the question of search strategies when one can choose between distributions and search strategy when one can, at a cost, affect the distributions one faces. The former model is used by Wohlstetter and Coleman [60], Kosters and Welch [20], and McCall [25] to explain observed discriminatory behavior in the work place. The latter model has been used by many to explain advertising.

Kohn and Shavell [19], made a substantial contribution by reformulating the optimal search problems in sufficient abstraction and showing that the same analysis applies to both the employer's and job seeker's problem. They are then able to show that in the majority of the variations, the optimal stopping rule is a switch point rule. That is, if under the optimal stopping rule d, one has not stopped after n samples, then there is a number s such that one stops at time $n+1$ if the utility associated with the $(n+1)^{th}$/ observation exceeds s and continues if it is less than s.*

They are then able to determine what happens to the switchpoint s under a variety of conditions. They show that s falls with an increase in time preference, and with an increase in next period's expected search costs. They are also able for special cases to determine the effect upon s of increased risk in the sense of Rothschild-Stiglitz [39] and in the sense of Diamond-Stiglitz [11].

---

*While their analysis holds for other economic problems involving optimal search, Kohn and Shavell have chosen to use the language of the expected utility maximizer.

For the reader interested in the mathematics of optimal stopping, Chow, Robbins, and Siegmund [8] is an excellent, though difficult reference. Good sources for the probability theory necessary to understand the optimal stopping literature are Ash [3], Chung [9], and Feller [13].

Perhaps the greatest motivation for search theory research has been in the analysis of non-degenerate wage distributions and in the analysis of unemployment. The optimal search strategy has attacked these problems from a partial equilibrium analysis, that is regarding either the employer's or job seeker's behavior as exogenous. Early equilibrium wage models were generally unsuccessful in supporting non-degenerate wage distributions. Indeed, in many of the early models, the wage distribution collapsed to the single monopsony wage. This has been shown to be a result of assuming that: there is a single market, the number of employers in the market is large, the cost of search in positive, employers maximize profits, employees maximize discounted net wages, and the equilibrium distribution is known by all. In the early 1970's a number of authors presented models in which some of the above assumptions were relaxed and which sustained non-degenerate wage distributions. Mirman and Porter [28], Lucas and Prescott [23], Mortensen [30], Diamond [10], and Telser [47] have each presented equilibrium models explaining wage distributions.

More recently Varian [52] has shown that the search structure can explain the existence of sales, Butters [7], uses a search structure for analyzing advertising, and numerous authors have used the search structure for analyzing effects of government policy on unemployment.

3. Equilibrium in Search Models. In this section a definition of equilibrium in search models is presented, and for certain elementary models this equilibrium is shown to exist. Much of the notation and many of the concepts in this section are taken from Arrow and Hahn [2].

In an elementary general equilibrium model there might be n distinguished goods, F firms, with firm f possessing a set of feasible production allocation Y_f in $\mathbb{R}^n$, and H households with households h having an initial endowment $\bar{x}_h$ in $\mathbb{R}^n$, a utility function $U_h : \mathbb{R}^n \to \mathbb{R}$, and a share $d(h,f)$

of firm f. Here $d(hf) \geq 0$ and for each f, $\Sigma_h d(h,f) = 1$. A price vector p^*, a consumption allocation $x^* \varepsilon \bigoplus_{h=1}^{H} \mathbb{R}^n$, and a production allocation $y^* \varepsilon \bigoplus_{f=1}^{F} Y_f$ constitute a general equilibrium if

(a) $p^* > 0$, where $p^* \varepsilon \mathbb{R}^n$, and $p^* > 0$, that is $p^*(j) \geq 0 \quad j = 1,2,\ldots,n$ and for some j, $p^*(j) > 0$.

(b) $\Sigma_h x_h^* \leq \Sigma_f y_f^* + \Sigma_h \bar{x}_h$

(c) y_f^* maximizes $p^* y_f$ subject to $y_f \varepsilon Y_f$.

(d) x_h^* maximizes $U_h(X_h)$ subject to $p^* X_h \leq p^* \bar{X}_h + \Sigma_f d(h,f) \, p^* y_f^*$

In the general equilibrium framework households and firms have full market information. No utility maximizing household will make a purchase of good i from firm f if firm f does not post the lowest price for good i. Since any firm would capture the entire market demand by any undercutting of the market, it is easy to show that in a full information competition market model all firms and households face the same prices. In a search model, price information is not universally distributed. Households or firms act upon their expectation of prices.

Even though a household's decision of how much to buy might be based upon an observed price, its decision of where to shop is usually based upon price expectation and not upon a full set of observed prices. A single n commonent price vector, since it need not exist in a search market economy, cannot be expected to erradicate excess demand as it does in a full information general equilibrium model. In the search model we have for each household h and firm f a price vector $p(h,f)$ and a vector $p(f)$, where $p(h,f)$ represents the price household h expects firm f to change and $p(f)$ is the price that firm f posts. In this model, only households are searchers. In equilibrium it is reasonable to expect that a household shops where it expects to maximize utility and that for this firm the expected and posted price should agree. This leads to the following definition of equilibrium for a search model.

<u>DEFINITION</u>: A price profile $p^*(h,f)$, $p^*(f)$, a consumption vector x^* allocation vector y^*, and a choice function $C:H F$ is a <u>competitive</u> <u>search</u> equilibrium if:

(a) $p^*(h,f) > 0$, $p^*(f) > 0$

(b) y_f^* maximizes $p^*(f)\ y_f$ subject to $y_f \varepsilon Y_f^*$

(c) x_n^* maximizes $U_h(x)$ over all x such that there exists $f \varepsilon F$ with

$$xp(f) \leqq \bar{x}_n\ p(f) + \sum_{k \varepsilon f} d(h,k)\ p^*(k)\ y_k^*$$

(d) $c^*(h) = f$ implies that

$$x_h^*\ p^*(f) \leq \bar{x}_n\ p^*(f) + \sum_{k \varepsilon F} d(h,k)\ p^*(k)\ y_k^*$$

(e) $$\sum_{\substack{h \\ c^*(h)=f}} (x_h^* - \bar{x}_n) \leq y_f^*$$

(f) $p^*(h,c^*(h)) = p^*(c^*(h))$.

Conditions a-e have obvious interpretations. Condition a is that expected and posted prices satisfy the standard notions of a price, that is they are non-negative and that the price of some good is positive. Condition b is that firms are profit maximizers while conditions c and d are the conditions that individuals are expected utility maximizers. Condition e is that in equilibrium there is not excess demand felt by any firm. Condition f is that the expected price held by a household agrees with the posted price at the firm where the household has transactions.

The obvious question is whether such an equilibrium exists for a search model. It will be shown below that the answer is affirmative if we have sufficient continuity conditions on the household demand and production supply functions and if we have a Walras' law type assumption on each demand. We begin by letting c be a function from H into F and listing our assumptions.

Assumption 1. For each $p \in \mathbb{R}^n$ $p > 0$, and each firm f, there is a choice of $y(p)$ in Y_f such that $py_f(p) \geq py$ all $y \in Y_f$. Further more the max $p \to y_f(p)$ is a continuous map from $\{p:p>0\}$ into Y_f.

Assumption 2. For each $p \in \bigoplus_{f=1}^{F} \mathbb{R}^n$ with $p(f) > 0$ and each household h, there is a choice of $x_h(p)$ in $\mathbb{R}^n$ such that $U_h(x_h(p) \geq \max\{U_h(x):$ all x such that

$$p(c(h))(x-\bar{x}_h) \leq \sum_{k\in F} d(h,k)\ p(k)\ Y_k(p(k))\},$$

$$p(c(h))\ (x_n-\bar{x}\ ) \leq \sum_{k\in F} d(h,k)\ p(k)\ y_k(p(k));\ \text{and}$$

further the map $p \to x_h(p)$ is continuous from $\{p:p \in \bigoplus_f \mathbb{R}^n,\ p(f) > 0\}$ into $\mathbb{R}^n$.

For each firm f and price function p in $\bigoplus_f \mathbb{R}^n$, let $Z_f(p) = \sum_{\substack{h \\ c(h)=f}} (x_h(p)-\bar{x}_n) - Y_f(p)$.

Walras' law states that $\Sigma_{f\varepsilon F}\, p(f)\, z_f(p) = 0$, we need a somewhat stronger assumption which is as follows.

<u>Assumption 3</u>. For no p in $\bigoplus_f \mathbb{R}^n$ with $p(f)\varepsilon S_n$ (the unit simplex) is it the case that $Z_f(p)(i) > 0$ implies that $p(f)(i) = 1$.

This is a condition on the function c, In essence it states that if for some price function p if there is excess demand in the system then there is some firm f experiencing excess demand for some good i where the price of good i charged by firm f is not the highest price the firm could charge. In a general equilibrium model assumption 3 is a consequence of Walras' law that $pz(p) = 0$.

We prove in Chapter IV the following theorem:

<u>Theorem</u>: Under assumptions 1, 2, and 3 a competitive search equilibrium exists with $C^* = C$.

The proof of this theorem is a direct application of a Browner fixed point theorem. It is similar to the proof for the existence of a competitive general equilibrium appearing in Chapter 2 of Arrow and Hahn [2].

An interesting corollary on price distributions can be obtained by appending two search models together.

Corollary: In a search equilibrium two different firms may post different prices for identical commodities.

4. A Housing Search Model. In the next section I present a simple one period search model which has significance in housing analysis. In the hosing market we find that potential buyers visit (according to some process) sellers to gain information about the characteristics of the unit the seller is offering. The potential buyer, without full information of the housing market and with knowledge of the seller's asking price but not of his reservation price makes a bid on the unit. The potential seller must await bids from buyers and must decide the level of his asking price as well as when to accept a bid which might be below the asking price.

It should be clear that the individual search processes in the housing market do not follow any simple model. The housing market is a dynamic market with buyers and sellers learning as they sample. Buyers do not sample at random but rather develop a search strategy. Sellers need not wait for buyers but may and do advertise. Furthermore, market brokers (Realtors) exist to facilitate the exchange of price and quantity information. However, the data transmitted by realtors need not always be accurate. A search model which attempts to incorporate each of these factors will be intractible to mathematical analysis. The obvious hope is that as with labor market analysis, a simplified model will capture enough of the behavior to yeild valid analysis.

The seller's problem is much the same as that of the job seeker's in labor market models. The seller is faced with a sequence of offers and must decide when to stop sampling and sell. This problem is well researched and the optimal strategy under a wide range of assumptions concerning the seller's objectives is known. The potential buyer's problem is not well developed in the literature. The buyer does not know whether or not a given bid will be accepted and thus this problem is not covered by the optimal search literature for employer's strategies. A simple model of the buyer's problem can be stated as follows:

At time n, the buyers samples from among m classes of units. Associated with each unit is an unobserved reservation price below which it will not be sold. Within class j, the reservation price is a random variable with distribution $F_j(\cdot)$. For each sample, the buyer incurs a fixed cost c. If he purchases a unit from class j at period n for price $P_j(n)$, he then enjoys the payoff, $U(X_j, Y-P_j(n)-nc)$, where Y = individual income. We assume that the probability of drawing a unit from class j at the n^{th} draw is p_j constant for all n. Let $P_j(n)$ be the buyers bid at time n for unit j and let P be the function with values $p_j(n)$. Let $i(n)$ be the class of unit sampled at the n^{th} period and let $Z(n)$ be the actual reservation price for the unit sampled at the n^{th} draw. If the buyer has income Y and bid structure P, he will enjoy pay off.

$B(Y,P) = U(X_{i(n)}, Y - P_{i(n)}(n) - nc)$ if and only if

$P_{i(n)}(n) \geq Z(n)$ and $P_{i(j)}(j) < Z(j)$ for $j<n$.

The buyer's problem is to find bid structure P which optimizes EB(Y,P).

Let $W(Y,P) = E[B(Y,P)]$, then it follows that if P* optimizes W(Y,P), then LP* defined by $LP^*(n) = P_j^*(n+1)$ optimizes W(Y-C,P).

It is important to note that P does not define a stopping rule for the process i(n) but does define a stopping rule for the process [i(n), Z(n)]. It is the unobservability of the random variables Z(n), that distinquishes this problem from that solved in the literature. To the best of my knowledge this problem, even under simplified assumptions, has yet to be solved. In the model presented below this problem is finessed by simply assuming the buyer has a bid structure.

A number of factors will affect the welfare of individuals in the housing market. One factor which is either affected directly by housing market policies or is indirectly affected by transportation policies is the size of the market. In the next section, I analyze how expanding the market affects through the search process the welfare of buyers and sellers. In particular I show that sellers are made worse off by expansion, and whether or not buyers are made better off depends upon the ratio of buyers to sellers.

The basis for the analysis is a one period market model with m buyers and n sellers. Each buyer I_i is assumed to have a bid structure $P_{i(\cdot)}$, where P_{ij} is the bid of individual i for sellers j's unit. Each seller J_j is assumed to have a reservation price X_j for his unit. At the beginning of the period, the buyers are distributed independently of each other among the sellers such that probability that indivudal I_i visits seller J_j is $1/n$. I_i buys unit J_j with probability $1/k$ if and only if I_i visits J_i, $P_{i,j} \geq X_j$, $P_{i,j} \geq P_{s,j}$ for all

---

*See appendix to this essay for further results on this problem.

individuals I_s visiting J_j and the number of individuals going to seller J_j with bid $P_{s,j} = P_{i,j}$ is exactly k. This one period model can be thought of as an equilibrium model where the market process is such that buyers and sellers are rpelaced as they are successful in the market and in which bids and reservation prices are independent of experience.*

Before continuing with the analysis it is necessary to introduce additional terminology.

BASIC Model.

$I = \{I_i \mid i = 1,2,\ldots,m\}$ set of Buyers.

$J = \{J_j \mid j = 1,2,\ldots,n\}$ set of units.

For an arbitrary set S, let $\|S\|$ denote its cardinality. $\mu_x(n) = \frac{nx-1}{nx}$ with $\mu = \mu(1)$

Y_{ij} be the bid of buyer I_i for unit J_j

X_j be the reservation price for unit J_j

$G_j(r) := \|\{I_i \mid Y_{ij} \leq r\}\|/m$ this is the fraction of buyers whose bid for the unit J_j falls below r

$B_j(r) := \|\{I_i \mid Y_{ij} = r\}\|/m$

---

*In general, this assumption will not be consistent with optimal search with positive search costs.

$H_j(r) := \|\{I_i \mid Y_{ij} = r\}\| \;/m$

$F(r) = \|\{J_j \mid X_j \leq r\}\| \;/n$

$C_j := \{i \mid Y_{ij} \geq X_j\}.$

$A_i := \{j \mid Y_{ij} \geq X_j\}$; if $j \in A_i$ then individual I_i has a bid for unit J_j at least as large as its reservation price.

PROPOSITION 4.1. The probability that individual I_i will make a successful bid is given by:

$$\sum_{j \in A_i} \frac{1}{mB_j(Y_{ij})} \left[\mu^{m[H_j(Y_{ij}) - B_j(Y_{ij})]} - \mu^{mH_j(Y_{ij})} \right]$$

Proof.[†] 1) The probability that I_i is successful and I_i visits unit J_j, and the number of bidders t for unit J_j with $Y_{tj} = Y_{ij}$ equal to k is given by

$$\left[\left\{ \frac{1}{n} \frac{1}{k} \binom{mB_j(Y_{ij})-1}{k-1} \left(\frac{1}{n}\right)^{k-1} \left(\frac{n-1}{n}\right)^{mB_j(Y_{ij})-k} \right\} ; \right.$$

$$(*) \qquad \left. \left\{ \left(\frac{n-1}{n}\right)^{m[H_j(Y_{ij}) - B_j(Y_{ij})]} \right\} \right] .$$

Provided $Y_{ij} \geq x_j$, it is 0 otherwise.

Let $P(j,k)$ denote the expression given in (*), then the probability that I_i is successful is given

---

[†]Detailed proofs appear in Chapter IV.

by

$$\sum_{j \varepsilon A_i} \sum_{k=1}^{mB_j(Y_{ij})} P(j,k) \quad \text{which equals}$$

$$\sum_{j \varepsilon A_i} \frac{1}{mB_j(Y_{ij})} \left[\mu^{m[H_j(Y_{ij}) - B_j(Y_{ij})]} - \mu^{mH_j(Y_{ij})} \right] .$$

PROPOSITION 4.2. The probability that J_i is sold is given by $1 - \mu^{mH_j(X_j)}$. If unit J_j is sold, the expected value of the sale is given by

$$\sum_{i \varepsilon C_j} \frac{Y_{ij}}{MB_j(Y_{ij})} \frac{\mu^{m[H_j(Y_{ij}) - B_j(Y_{ij})]} - \mu^{mH_j(Y_{ij})}}{1 - \mu^{mH_j(X_j)}} .$$

Proof. 1) Unit J_j will <u>not</u> be sold only if each bidder with bid for J_j at least as great as X_j visits a unit other than J_j. There are $mH_j(X_j)$ bidders with bids for unit J_j at least as big as its reservation price X_j. The probability of going to a unit other than J_j is given by .

2) If $Y_{ij} \geq X_j$, then the unit J_j will be sold for Y_{ij} if some individuals with bid $Y_{sj} = Y_{ij}$ visits J_j and all individuals I_t with bid $Y_{tj} > Y_{ij}$ go elsewhere. Hence probability that J_j is sold for exactly Y_{ij} (where $Y_{ij} \geq X_j$) is given

by $[1 - \mu^{mB_j(Y_{ij})}]\ \mu^{m[H_j(Y_{ij}) - B_j(Y_{ij})]}$ which equals

$$\sum_{k=1}^{mB_j(Y_{ij})} \frac{1}{mB_j(Y_{ij})} [1-\mu^{mB_j(Y_{ij})}]\mu^{m[H_j(Y_{ij})-B_j(Y_{ij})]} .$$

3) The probability that it is sold for Y_{ij} $(Y_{ij} \geq X_j)$ given that it is sold is simply the probability it is sold for Y_{ij} divided by the probability that it is sold.

4) To obtain the result stated, we need only observe that the set of individual with bids for J_j at least as great as X_j is the disjoint union of classes of individuals whose bid for J_j equals r over all $r \geq X_j$.

In a full information deterministic market model a seller can affect the share of the market captured by varying his prices relative to that of other sellers. Proposition 2. tells us that it is a construct of this model that the welfare of any particular seller is independent of the reservation prices of other sellers. Proposition 1. states that buyers in this market compete with each other and that the buyer's problem is a game theory problem involving the action of both the other buyers and the sellers.

Suppose now each of the individuals in the model is replicated x times to give an expanded market. Our goal is to determine the effects upon the buyer's and seller's welfare from such an expansion. This is a partial equilibrium analysis in that we assume that price structures are not affected. If buyers and sellers determine their bids and reservation prices upon the distribution of bids and reservation prices and independently of the market size then the price structures will not change. Behavior of buyers and sellers independent of market size is suboptimal, however, since changing the size affects the probabilities of being visited and of having competition in a bid.

The expanded market is the x time disjoint union of the market in the original model. Thus the new set of sellers can be denoted by $\hat{I} := \{I_{i,s}$ $i = 1,2,\ldots,m,\ \ s = 1,\ldots,x\}$. If we denote the state variables in the new model by a $\hat{}$, we observe the following relations:

$$\hat{J} = \{J_{j,s} \mid j = 1,\ldots,n \ \ s = 1,\ldots,x\}$$

$$\hat{m} = xm$$

$$\hat{n} = xn$$

$$\hat{\mu} = \mu(n) = \frac{nx-1}{nx}$$

$$\hat{G}_{(j,s)}(r) = G_j(r)$$

$$\hat{B}_{(j,s)}(r) = B_j(r)$$

$$\hat{H}_{(j,s)}(r) = H_j(r)$$

$$\hat{F}(r) = F(r)$$

$$\hat{C}_{(j,s)} := \{(i,s) \quad i \in C_j \quad s = 1,\dots,x\} = \dot{U}_{s=1}^{x} C_j$$

$$\hat{A}_{(i,s)} := \{(j,s) \quad j \in A_i \quad s = 1,\dots,x\} = \dot{U}_{s=1}^{x} A_i .$$

PROPOSITION 4.3. In the replicated market model, the probability that indivudal $I_{(i,s)}$ will make a successful bid is given by

$$\sum_{j\varepsilon A_i} \frac{1}{mB_j(Y_{ij})} \mu(x)^{xm[H_j(Y_{ij}) - B_j(Y_{ij})]} - \mu(x)^{xmH_j(Y_{ij})} .$$

Proof. Use the fact that $\hat{A}_{(i,s)} = \dot{U}_{s=1}^{x} A_i$ and then use Proposition 4.3. replacing all the variables with their values in the replicated market model.

PROPOSITION 4.4. The probability that $J_{(j,s)}$ will be sold is $1 - \mu(x)^{xmH_j(X_j)}$. The expected value of the sale of $J_{(j,s)}$ given that a sale occurs is

$$\sum_{i\varepsilon C_j} \frac{Y_{i,j}}{mB_j(Y_{i,j})} \frac{\mu(x)^{xm[H_j(Y_{i,j})-B_j(Y_{i,j})]} - \mu(x)^{xmH_j(Y_{ij})}}{1 - \mu(x)^{xmH_j(X_j)}} .$$

Proof. Use Proposition 2. replacing variables by their values in the replicated market model and use the fact that $C_{(j,s)} = U_{s=1}^{x} C_j$.

We are now able to determine how expending the market through replication affects the welfare of buyers and sellers.

THEOREM 4.5. Let $P(x,i) = \sum_{j\varepsilon A_i} \frac{1}{mB_j(Y_{i,j})}$ $\mu(x)^{xm[H_j(Y_{i,j}) - B_j(Y_{i,j})]} - \mu(x)^{xmH_j(Y_{i,j})}$

(the probability that buyer $I_{(i,1)}$ buys a unit in the market replicated x times. Let $\alpha = m/n$.

a) Sufficient conditions for $\frac{\partial P(x,i)}{\partial x} > 0$ are that $\alpha > \sup_{j\varepsilon A_i} [H_j(Y_{i,j}) - B_j(Y_{i,j})]^{-1}$, $nx > 2$ and $A_i \neq \emptyset$.

b) A sufficient condition for $\frac{\partial P(x,i)}{\partial x} \leq 0$ are that $\alpha > \frac{-1}{nx \ln (\frac{nx-1}{nx})}$ $nx > 2$ and $A_i \neq \emptyset$. In particular if $\alpha < 5/7$, $\frac{\partial P(x,i)}{\partial x} \leq 0$ $(nx>2)$.

One can reasonably interpret α as a measure of congestion among buyers or equivalently for a fixed market size, α is a rough measure of the competition between buyers. Theorem 4.5 states that when this competition is low, when there are proportionately more sellers than buyers, buyers are made worse off by market expansion. Theorem 4.5 states that if a buyer tends to be a low bidder on each unit whose reservation price doesn't exceed his bid, then market expansion makes the buyer better off. Market expansion affects the buyers by increasing the number of competitiors and by increasing the number of opportunities. Theorem 4.5 gives sufficient conditions for one of these effects to dominate.

THEOREM 4.6. Let $Q(x,j) = 1 - \mu(x)^{xmH_j(X_j)}$ be the probability that $J_{(j,s)}$ is sold when the market is replicated x times. Let $E(x,j) = \sum_{i \varepsilon C_j} \frac{Y_{i,j}}{mB_j(Y_{i,j})}$

$$\frac{\mu(x)^{xm[H_j(Y_{i,j}) - B_j(Y_{i,j})]} - \mu(x)^{xmH_j(Y_{i,j})}}{1 - \mu(x)^{xmH_j(X_j)}}$$

be the expected value of the sale of unit $J_{(j,s)}$

given that a sale takes place in the market replicated x times. Then if $nx \geq 2$.

a) Necessary and sufficient conditions for

$\frac{\partial Q(x,j)}{\partial x} < 0$ are that $C_j \neq \emptyset$ and

b) Necessary and sufficient conditions for

$\frac{\partial E(x,j)}{\partial x} < 0$ are that there are i i' with

$Y_{i,j} > Y_{i',j} \geq x_j$.

Proof. See Chapter IV.

The interpretation of Theorem 4.6 is straightforward. Expanding the market never is beneficial to the seller. In particular, if there are at least two buyers with different bids exceeding the reservation price, the seller is made worse off by expansion.

5. An Optimal Bidding Problem in a Search Model.

In some markets in which search is a prevalent feature, there is sufficient flow of information that buyers and sellers have full knowledge of price distribution. Search still occurs since prospective buyers, although they know price distributions, do not know which seller is posting the lower prices. An example of this is the residential housing market in urban areas where realtors maintain extensive records of past transactions and make these available to prospective home buyers. The potential buyer visits a unit and then may tender a bid for that unit. The decision of how much to bid is in part determined by the bidders expectation of the seller's reservation price and upon the bidder's wealth. In this section it is shown that an optimal bidding strategy exists and that if search is costly this strategy need not result in a bid pattern that is monotonic with respect to time.

The bidder samples sequentially and at time n samples the pair $(X_{i(n)}, Z(n))$ where $i(n) \in \{1,2,\dots,m\}$ and $Z(n) \in \mathbb{R}$. $X_{i(n)}$ or equivalently $i(n)$ is observed but $Z(n)$ is not, however the distribution of $Z(n)$ given $i(n)$

is known and denoted by $F_{i(n)}(\cdot)$. The probability that $i(n) = j$ is fixed and denoted by λ_j $(j \in \{1,2,\ldots,m\})$. The bidder starts with income Y and incurs a fixed cost $c > 0$ per each draw of the sample. If the individual bids $P_{i(n)}(n)$ at time n for $X_{i(n)}$, he will enjoy one time payoff of $U(X_{i(n)}, Y - nc - P_{i(n)}(n))$* provided $P_{i(n)}(n) \geq Z(n)$ and $P_{i(j)}(j) < Z(j)$ for $j < n$. Let P be a bid profile, that is $P_{i(n)}$ is the bid for X_i at time n, then associated with the bid profile P is the expected payoff $W(Y,c,P)$. The bidder's problem is to find a bid profile p^* that maximizes $W(Y,c,P)$.

The bidder's problem is not too difficult to understand. If he bids too low, he will fail to make a buy and then must incur the cost of additional search. If, on the other hand, his bid is in excess of that needed to make a buy, then the difference of his bid and the minimum needed to secure the buy is lost opportunity. In the model described above, it is assumed that search requires little time and so utility is not time discounted. To study the bid profile, the shift operation L is introduced where if P is a bid profile, LP is a bid profile with $(LP)_i(n) = P_i(n+1)$.

PROPOSITION 5.1. Let L_p^j be the bid profile with $(L_p^j)_i(n) = P_i(n+j)$. Then

$$W(Y,c,P) = \sum_{i=1}^{m} U(X_i, Y_i - c - P_i(1)) \lambda_i F_i(P_i(1)) +$$

$$+ [1 - \sum_{i=1}^{m} \lambda_i F_i(P_i(1))] W(Y - c, c, LP).$$

Proof. 1) Let $h(P,n) = \sum_{i=1}^{m} \lambda_i F_i(P_i(n))$ $\{= \text{Prob } P_{i(n)}(n) \geq Z(n) \geq Z(n)\}$ $H(Y,P,n) =$

$$\sum_{1=1}^{m} \lambda_i U(X_i, Y - nc - P_{i(n)}(n)) F_i(P_i(n))/h(P,n)$$

$= \{$expected payoff given $P_{i(n)}(n) \geq Z(n)$ and $P_{i(j)}(j) < Z(j)\}$.

---

*Assume $U(X_i,Z)$ is nondecreasing in the second argument for each i.

Then $W(Y,c,P) = h(P,1)H(Y,P,1) + [1-h(P,1]h(P,2)H(Y,P,2]+$

$[1-h(P,1)][1-h(P,2)]h(P,3)H(Y,P,3)+\ldots =$

$\Sigma_{j=1}^{\infty} \Pi_{i=o}^{j-1} \quad (1-h(P,i)]h(P,j)H(Y,P,j) \quad [h(P,0) \equiv 0]$

2) Note. $\Pi_{i=o}^{n-1} [1-h(P,i)]h(P,n) = \text{prob}\{P_{i(n)}(n) \geq Z(n)$

and $P_{i(j)}(j) < Z(j) \ j < n$.

3) $h(P,n+1) = h(LP,n)$ for $n=1,2,\ldots$ (by def $h(LP,0) = 0$)

$H(Y,P,n+1) = H(Y-c,LP,n) \quad n=1,2,3\ldots$

4) $W(Y,c,P) = h(P,1)H(Y,P,1) + [1-h(P,1)]\{h(P,2)H(Y,P,2)$

$+\Sigma_{j=3}^{\infty}\Pi_{i=2}^{j-1} (1-h(P,i))h(P,j)H(Y,P,j)\}$

$= h(P,1)H(Y,P,1) + [1-h(P,1)] \{ h(LP,1)H(Y-c,LP,1) +$

$+ \Sigma_{j=3}^{\infty}\Pi_{i=1}^{j-1} (1-h(LP,i-1)]h(LP,j-1)H(Y-c,LP,j-1)\} =$

$= h(P,1)H(Y,P,1) + [1-h(P,1)]\Sigma_{j=1}^{\infty} \Pi_{i=o}^{j-1}$

$[1-h(LP,i)]h(LP,j)H(Y-c,LP,j)$

5) $W(Y,c,P) = \Sigma_{i=1}^{m}\lambda_i \ U(X_i,Y-c-P_i(1)F_i(P_i(1))$

$+ [1 - \Sigma_{i=1}^{m} \lambda_i F_i(P_i(1))]W(Y-c,c,LP)$.

PROPOSITION 5.2.

Suppose P* optimizes W(Y,c,P), if for some i $F_i(P_i^*(1)) \neq 1$, then LP* optimizes W(Y-c,c,P).

Proof.

1) It suffices to show any bid profile Q.

$W(Y-c,c,LP^*) \geq W(Y-c,c,Q)$.

2) Suppose $W(Y-c,c,Q) > W(Y-c,c,LP^*)$

Let P be the bid profile with $P_i(1) = P_i^*(1)$

$P_i(n) = Q_i(n-1)$ for $n \geq 2$

Then

$$W(Y,c,P) = \Sigma_{i=1}^{m} \lambda_i \; U(X_i, Y-c-P_i^*(1) F_i(P_i^*(1)) +$$

$$[1-\Sigma_{i=1}^{n} \lambda_i \; F_i(P_i^*(1))] W(Y-c, cQ)$$

and

3) $$W(Y,c,P) > \Sigma_{i=1}^{m} \lambda_i \; U(X_i, Y-c-P_i^*(1)) F_i(P_i^*(1)) +$$

$$[1-\Sigma_{i=1}^{m} \lambda_i \; F_i(P_i^*(1))] W(Y-c,c,LP^*)$$

and

4) $W(Y,c,P) > W(Y,c,P^*)$ contradiction

Proposition 2 states that if we know the optimal bid structure from time n+1 onward or if we know $W(Y-(n+1)c,c,L^nP^*)$, we can compute $P^*_{(1)} \ldots P^*_n(n)$. In particular $P_i^*(n)$ must be the bids that maximizes.

$$\text{(A)} \quad \sum_{i=1}^{m} U(X_i,\ Y - nc - P_i)\lambda_i F_i(P_i)$$

$$+ \left[1 - \sum_{i=1}^{m} \lambda_i F_i(P_i)\right] W(Y - nc,\ c,\ L^n P^*) \quad .$$

The maximum value that (A) takes on is then $W(Y-(n-1)c, c, L^{(n-1)}P^*)$ and the value P_i that maximize (A) become $P_i^*(n)$. The problem of solving for $P^*(n)$ (and hence for $P^*(1), \ldots, P^*(n)$) reduces to solving the simpler problem of finding p that maximizes $[U(X_i, Y-nc-p) = W(Y-nc, c, P^n P^*)]F_i(p)$. It is perhaps helpful to note that if $W(Y,c) = \sup_P W(Y,c,P)$: then $W(Y,c)$ is increasing in Y and decreasing in c as expected. In the event that the distribution F_i is associated with a probability measure with finitely many atoms, the search for optimal P_i can be restricted to the atoms. In the case that F_i are continuously differentiable we can develop further results.

Let us now assume that each F_i is twice continuously differentiable in interval $(0,\infty)$ and that $\lim_{p \downarrow 0} F_i(p) = 0$. We further assume that for each i $U(X_i, Z)$ is twice continuously differentiable in Z.

$$\frac{\partial U(X_i, Z)}{\partial Z} > 0 \ \Big|_{Z>0} \quad \frac{\partial^2 U}{(\partial Z)^2}(X_i, Z) < 0 \ \Big|_{Z>0}$$

$\lim_{Z \downarrow 0} U(X_i, Z) = 0.$

THEOREM 5.3. Under the conditions above an optimum bid profile exists.

Proof. 1) Since $U(X_i, Z) \leq 0$ for $Z \leq 0$. $W(Y-nc, c) = 0$ for all n such that $nc \geq Y$. Let N be the least n such that $Y-nc \leq 0$. Then we can let $P_i^*(n) = 0$ for all i and $n \geq N$, and of course $W(Y-(N-1)c, c) = 0$.

2) To find $P_i^*(N-1)$ we need to find a solution to maximize $U(X_i, Y-(N-1)c-p_i)F_i(p_i) = m(p_i)$. Now for $p_i < 0$ $m(p_i) = 0$ and for $p_i \geq c$ $m(p_i) \leq 0$; furthermore $m(p_i)$ is continuous, so there exists p_i that maximizes $U(X_i, Y-(N-1)c-p_i)F_i(p_i)$.

3) Thus we can find $P_i^*(N-1)$ and we can compute $W(Y-(N-2)c, c) = \Sigma_{i=1}^{n} \lambda_i\, U(X_i, Y-(N-1)c -$

$- P_i^*(N-1))F_i(P_i^*(N-1))$.

4) To find $p_i^*(N-2)$ we need to find p_i that maximizes $m(p_i) = [U(X_i, Y-(N-2)c-p_i) - W(Y-(N-2)c, c)]F_i(p_i)$. For $p_i < 0$, $m(p_i) = 0$ and for $p_i > 2c$, $m(P_i) \leq 0$; since $m(P_i)$ is continuous we can find $0 \leq p_i \leq 2c$, that maximizes $m(P_i)$ and thus can compute $p_i^*(N-2)$ and $W(Y-(N-3)c, c)$.

5) In general, let $W(n) = W(Y\text{-}nc,c)$; to find $p_i{}^*(n)$ we need only find p_i that maximizes $m(p_i) = [U(X_i,Y\text{-}nc\text{-}p_i) - W(n)]F_i(p_i)$. Now $m(p_i)$ is continuous with $m(p_i) = 0$ for $p_i < 0$ and $m(p_i) \leq 0$ for $p_i > Y\text{-}nc$, so a solution is always possible.

In general we seek solutions to the problem of maximize $M(p_i) = [U(X_i,Y\text{-}nc\text{-}p_i) - W(n)]F_i(p_i)$. If $F_i(p_i) = 0$ for all p_i such that $p_i < Y\text{-}nc$ then we can choose for our optimal p_i, $p_i = 0$ and $m(0) = 0$. If $m(p_i) > 0$ for some p_i, then our optimal solution must satisfy the differential condition.

$0 = m'(p) = [U(X,Y\text{-}nc\text{-}p)-W(n)]F'(p) - F(p)\,U_2(X,Y\text{-}nc\text{-}p)$. Thus at our optimal solution p, we must satisfy,

$$\frac{F'(p)}{F(p)} = \frac{U_2(X,Y\text{-}nc\text{-}p)}{U(X,Y\text{-}nc\text{-}p) - W(n)} \quad ** \quad .$$

It is somewhat surprising that the optimal bid profile, $P_i{}^*(n)$, need not be monotonic in n.

---

*We have dropped the subscript i for convenience of notation.

$$^{**}U_2(X,Z) = \frac{\partial U}{\partial Z}(X,Z).$$

$P_i^*(n)$ maximizes the generic function $H(Y-p)F(p) + (1-F(p))W = M(p)$. As n goes from n to $n+1$ both Y and W fall. At the optimal $p = P$, $M'(P) = 0 = [H(Y-P)-W]F'(P)-F(P)H'(Y-P)$. If $F''(\cdot)<0$, then $m''(P) = [H(Y-P)-W]F''(P) - F'(P)H'(Y-P) - F'(P)H'(Y-P) + F(P)H''(Y-P)$ so $M''(P) < 0$ and there is a unique P such that $M(P) = 0$. Now let $P = P(Y,W)$ solve $M'(P) = 0$. Taking $\frac{\partial M'(P(Y,W))}{\partial Y} = \frac{\partial 0}{\partial Y} = \frac{\partial 0}{\partial W} = \frac{\partial M'(P(Y,W))}{\partial W} = 0$, we find that

$$\frac{\partial P}{\partial Y} = \frac{F(P)H''(Y-P) - H'(Y-P)F'(P)}{[H(Y-P)-W]F''(P) - 2F'(P)\,H'(Y-P) + F(P)H''(Y-P)} =$$

$$\frac{\partial P}{\partial Y} = \frac{F(P)H''(Y-P) - H'(Y-P)F'(P)}{M''(P)} > 0$$

$$\frac{\partial P}{\partial W} = \frac{F'(P)}{M''(P)} < 0 \quad .$$

Since as n increases both Y and W fall there is no conclusive determination of what happens to the optimal bid $P_i^*(n)$. One might suspect and it is easy to construct examples where $F_i(P)$ is a discrete probability distribution and $p_i^*(n)$ is falling as n increases. The following example reveals that it is possible for $p_i^*(n) < p_i^*(n+1)$.

Example.

$M = 1$ and hence $\lambda_1 = 1$

$Y = 14 \quad C = 5 \quad Q = 3 \quad P = 2\text{-}1/2$

$F(\cdot)$ is discrete with $F(t) =$

0	$t<2.5$
.571	$2.5\ t\leq 3$
.6	$3\leq t<100$
1	$100\leq t$

$U(X,1) = 1$

$U(X,1.5) = 1.05$

$U(X,6) = 1.49$

$U(X,6.) = 1.538$.

Note $U(X,\cdot)$ is chosen such that the above values could be generated from a function $U(X,t)$ with $\frac{\partial}{\partial t} U(X,t) > 0 \quad \frac{\partial^2 U}{\partial t^2} (X,t) < 0$.

1) Since $Y-3c = 14-15 = -1$, it follows that $P^*(n) = 0$ for $n \geq 3$. Furthermore, by the nature of $F(\cdot)$, $p^*(n) \varepsilon \{2.5, 3\}$ for $n = 1,2$.

2) $U(Y-2c-Q)F(Q) = U(X,1)(.6) = .6 > .59955 = (1.05)(.571) = U(X, 1.5)(.571 = U(Y-2c-P)F(P)$ it follows that $P(2) = Q = 3$ and $W(1) = .6$.

3) $F(P)U(X,Y-c-P) + (1-F(P))W(1) = 1.135598$ $1.134 - F(Q)U(X,Y-c-Q) + (1-F(Q)]W(1)$. So $P(L)=P=2.5$.

4) $P(1) < P(2)$.

Search is naturally a part of the urban economy and particularly of the housing market. There are numerous directions that search theory research might take with respect to urban economics. One area of research which seems especially fruitful, is to develop rational bid rent search models. A rational bid rent search model is one in which buyer's bids and seller's reservation prices are consistent with optimal search strategies given some rule for determining how buyers and sellers are brought together.* It is interesting to note that if in a given bid rent model,** the equilibrium bids satisfy the following conditions:

1) For each I_i there is a unique J_{j_0} such that B_{i,j_0} B_{s,j_0} $i \neq s$. (denote $j(i) = j_0$).

---

*Examples of such roles are:

1) Naive rule each buyer independent of the action of other buyers visits a given seller at random with equi probabilities of visiting one seller.

2) Maximal expected utility: Buyer I_i chooses from random with equi probability among those sellers J_j that maximize expected utility.

**Sell Alonzo [1] and Wheaton [56].

2) For each J_{j_o} there is an I_i such that $j(i) = j_o$.

Then if let $X_j = \sup_i B_{i,j}$, the bids $B_{i,j}$ and reservation prices X_j are consistent with the search rule that each buyer visits the seller which maximizes his expected utility.

A natural consequence of the full information bid rent model is that individuals with the same income and tastes will end up enjoying the same level of utility. This fact is often exploited in empirical studies to estimate parameters of individuals utility functions and to estimate marginal rates of substitution between various housing characteristics.*** In a stochastic search model the hypothesis of constant utility for individuals with some preferences and initial income is not supported and hence parameter estimates based upon the assumption of indifference need not be consistent. It remains to determine, however, the degree of inconsistency that this introduces.

***See Wheaton [57].

CHAPTER III

ESTIMATION AND HYPOTHESIS TESTING IN THE PRESENCE OF HETEROSKEDASTICITY.

1. Introduction. The often encountered one equation liner model can be written as either:

$$y_n = \sum_{i=0}^{K} x_{n,i}\,\beta_i + \varepsilon_n \quad \text{or}$$

$$Y = X\beta + \varepsilon$$

where Y and ε are $N\times 1$ vectors, X is an $N\times K$ matrix, β a $K\times 1$ vector, Y and X are observed, β and ε are unobservable, and ε is a vector of random variables with $E(\varepsilon) = 0$. The usual analysis involves the estimation of β, testing of hypothesis on β, or predicting y_{n+1}, when $x_{n+1,1},\ldots,x_{n+1,k}$ are given or predicted. The usual naive assumption that $E(\varepsilon\varepsilon^T)$ is a scalar diagonal matrix cannot usually be supported. When the observations are generated by time series data one would expect serial correlation, $E(\varepsilon_s\varepsilon_t) \neq 0$ when $s \neq t$, while cross sectional and grouped data frequently imply problems of heteroskedasticity, $E(\varepsilon_i^2) \neq E(\varepsilon_j^2)$ when $i \neq j$.

The naive assumption of $E(\varepsilon\varepsilon^T) = \sigma_0^2 I$ is hard to give up. Under this assumption, together with assumptions on the data matrix X, one has the well known Gauss Markov theorem which states that the ordinary least squares estimator $\hat{\beta}$, given by $\hat{\beta} = (X^TX)^{-1} X^TY$ is a minimum variance estimator. Furthermore, the predictor $y_* = x\hat{\beta}$ is a best linear unbiased predictor of y given x_*. Under additional assumptions on the data matrix X, the asymptotic distribution of $\hat{\beta}$ can be computed. If Q is given by the $\lim_N N^{-1}(X^TX)$, then $\sqrt{N}(\hat{\beta}-\beta)$ has a normal limiting distribution with mean 0 and covariance matrix $\sigma_0^2 Q^{-1}$. Furthermore if $S^2 = \frac{\varepsilon^T\varepsilon}{N-K}$, where $\hat{\varepsilon} = Y - X\hat{\beta}$, S^2 is a consistent estimator of σ_0^2. Under the naive assumption of $E(\varepsilon\varepsilon^T)$ scalar, one can easily compute the limiting distributions for $\hat{\beta}$, compute asymptotic confidence intervals for $\hat{\beta}$, and test, at least using asymptotic theory, linear hypothesis on $\hat{\beta}$. Indeed, most regression packager automatically report all these statistics which of course are valid when $E(\varepsilon\varepsilon^T)$ is scalar. If $E(\varepsilon\varepsilon^T) = \Omega$ is not a scalar multiple of the identity, then the statistics reported by the usual regression packages do not have their usual meaning. To examine the loss resulting from using ordinary least squares. When Ω is not a scalar multiple of the identity, we consider the case where the data matrix X is non-stochastic and the $\lim_N N^{-1} X^T X$ exists and equals some $K\times K$ matrix Q. The ordinary least squares estimator $\hat{\beta}$ is given by $\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta+\varepsilon) = \beta + (X^T X)^{-1} X^T \varepsilon$. The OLS estimator $\hat{\beta}$ is un unbiased estimator of β and $p \lim_N \hat{\beta} = p \lim_N \beta + (\frac{1}{N} X^T X)^{-1} (\frac{1}{N} X^T \varepsilon) = \beta$, so $\hat{\beta}$ is a consistent estimator of β. If Ω is positive definite then, since Ω is symmetric, we can find a diagonal operator D and unitary operator U such that $\Omega = U^{-1} D^2 U$; let $S = D^{-1} U$, then $SY = SX\beta + S\varepsilon$. Now, $E(S\varepsilon\varepsilon^T S^T) = D^{-1} U \Omega U^T D^{-1} = D^{-1} D^2 D^{-1} = 1$, so by the Gauss Markov theorem, $\hat{\beta}_\omega = (X^T S^T S X)^{-1} X^T S^T S Y = (X^T (\Omega^{-1}) X)^{-1} X^T \Omega^{-1} Y$ is the least linear unbiased estimator for β and thus $\hat{\beta}$ is not efficient. In the case of heteroskedasticity, Ω is diagonal and S is diagonal with $(S)_{ii} = (\sigma_{ii}{}^2)^{-\frac{1}{2}}$. The covariance matrix for $\hat{\beta}$ is given by $E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T] = E[(X^T X)^{-1} X^T \varepsilon\varepsilon^T X (X^T)^{-1}] = (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}$ and not by $\sigma_0{}^2 (X^T X)^{-1}$.

Therefore $\sqrt{N}(\hat{\beta}-\beta)$ is not asymptotically distributed with mean zero and covariance matrix $\sigma_0^2(N^{-1}X^TX)^{-1} = \sigma_0^2 Q$. Furthermore the commonly printed statistic of $\frac{\varepsilon^T\varepsilon}{N-K}(N^{-1}X^TX)^{-1}$ is not a consistent estimator of the covariance matrix for $\sqrt{N}(\hat{\beta}-\beta)$. It is clear that the computation of a consistent estimator for this covariance matrix requires a consistent estimator for $\frac{1}{N}X^T\Omega X$.

In the case of heteroskedasticity, Ω is diagonal the operator S is diagonal with $(S)_{ii} = (\sigma_{ii}^2)^{-\frac{1}{2}}$, and $\hat{\beta}_\omega$ is simply weighted least squares. If Ω is known then S can be computed and the model $Y = X\beta+\varepsilon$ can be transformed to the model $SY = SX\beta+S\varepsilon$. Not only does OLS on the transformed model give optimal linear estimator for β, but the usual test statistics computed by the standard regression packages for this transformed model can be correctly interpreted.

The identification of the heteroskedastic structure has importance beyond that of statistical consideration. In many instances the data has a natural grouping such that within each group the variances are constant. This may suggest to the investigator that while the group share the same structural parameters β, the processes generating the structure are not the same across groups. The idea that the heteroskedastic structure conveys theoretical information is explored in [12].

In general, Ω is not known and $\frac{1}{N} X^T \Omega X$ is difficult to estimate. The solution, at least for special forms of Ω, is to either develop sufficiently good estimators for Ω and for S, so that the transofrmed model using these estimators has desirable asymptotic properties or to develop consistent estimator for $N^{-1} X^T \Omega X$ so that hypothesis tests can be done using the OLS' estimator $\hat{\beta}$ on the original model.

In the case of heteroskedasticity, Halbert-White has proposed a consistent estimator for $N^{-1} X^T \Omega X$ so that asymptotically valid confidence intervals and tests can be developed using $\hat{\beta}$. This procedure is explored in the next section.

2. <u>A Theorem of Halbert White</u>. One approach to the problem of hypothesis testing when Ω is diagonal has been to look for a simply computable consistent estimator of the convariance matrix associated with

the OLS estimator $\hat{\beta}$. White in [59], proposes such an estimator and as a corollary develops asymptotically valid confidence intervals and hypothesis tests based upon $\hat{\beta}$. Before stating White's results, it is useful to motivate his approach. For this purpose, assume that X is non-stochastic, although this assumption is not necessary to obtain White's results.

As we have already seen, the covariance matrix $X(\hat{\beta})$ for $\hat{\beta}$ is given by $(X^TX)^{-1}(X^T\Omega X)(X^TX)^{-1}$ and that for $\sqrt{N}(\hat{\beta}-\beta)$ is given by $(N^{-1}X^TX)(N^{-1}X^T\Omega X)(N^{-1}X^TX)^{-1}$. Since X is observable, the problem reduces to developing an estimator for $(N^{-1}X^T\Omega X)$. The matrix X can be written as

$$\begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{pmatrix} \qquad \text{where}$$

X_i is a now vector with $(X_i)_j = X_{ij}$. The matrix X^T can be written as $(X_1^T, X_2^T, \ldots, X_N^T)$ where X_i^T is a column vector with $(X_i^T)_j = X_{ji}$. [Please note that X is a real matrix so $X^T = X^*$.] In the case of heteroskedasticity, Ω is a diagonal matrix of the form

$$\Omega = \begin{bmatrix} \sigma_{11}^2 & & & & 0 \\ & \sigma_{22}^2 & & & \\ & & \ddots & & \\ 0 & & & & \sigma_{NN}^2 \end{bmatrix} .$$

The critical matrix $N^{-1} X^T \Omega X$ reduces to

$$N^{-1} \sum_{i=1}^{N} \sigma_{ii}^2 X_i^T X_i .$$

Under fairly general restrictions on X and ε, the strong law of large numbers can be evoked to show that

$$\lim_N \| N^{-1} \sum_{i=1}^{N} (\sigma_{ii}^2 - \varepsilon_i^2) X_i^T X_i \| = 0, \quad \text{where} \quad \| \cdot \|$$

can be taken to be any norm on $\mathbb{R}^{K\times K}$. This suggests that $N^{-1} \sum_{i=1}^{N} \varepsilon_i^2 X_i^T X_i$ would be a good estimator of $N^{-1} X^T \Omega X$, but ε_i^2 like σ_{ii}^2 is not observable. However, $\hat{\varepsilon}_i$ which equals $Y_i - X_i \hat{\beta}$ is observable, which leads one to speculate that since $\hat{\beta}$ is strongly consisted estimator of β,

$$N^{-1} \sum_{i=1}^{N} \hat{\varepsilon}_i^{\,2} X_i^{\,T} X_i$$

might be a good estimator for $N^{-1}X^T\Omega X$. It seems likely that White followed a similar line of reasoning to obtain the following results.

Before stating White's results it is necessary to introduce several new definitions and notation and to enumerate the formal assumptions of his theorems. This we now do.

A1) The model is known to be

$$Y_i = X_i \ \beta_0 + \varepsilon_i \qquad i=1,2,\ldots,n$$

where (X_i, ε_i) is a sequence of independent (not necessarily identically) distributed random vectors, such that X_i (a $1\times K$ vector) and ε_i (a scalar) satisfy $E(X_i^T \varepsilon_i) = 0$. The scalar valued ε_i are unobservable while Y_i and X_i are observable. The parameter vector β_0 is a finite unknown $K\times 1$ vector to be estimated.

A2) (a) There exists positive finite constants S and Δ such that for all i, $E(|\varepsilon_i^2|^{1+s}) < \Delta$ and $E(|X_{ij} \ X_{ik}|^{1+s}) < \Delta$, $j,k = 1,2,\ldots,K$.

(b) $\overline{M}_n = n^{-1} \sum_{i=1}^{n} E(X_i^T, X_i)$ is nonsingular for all n sufficiently large and for n sufficiently large let $\overline{M}_n > S > 0$.

A3) (a) There exist positive finite constants S and Δ such that for all i

$$E(|\varepsilon_i{}^2 X_i, X_{ik}|^{1+s}) < \Delta \qquad j,k = 1,2,\ldots,K \quad .$$

(b) The average covariance matrix is $\overline{V}_n := \frac{1}{n} \sum_{i=1}^{n} E(\varepsilon_i{}^2 X_i{}^T X_i)$ and for n sufficiently large $\overline{V}_n$ is nonsingular with $\det \overline{V}_n > S > 0$.

Let $\hat{\beta}_n = (X^T X)^{-1} X^T Y$ be the OLS estimator of β_0, let $\hat{\varepsilon}_{in} = Y_i - X_i \hat{\beta}_n$ and let $\hat{V}_n = \frac{1}{n} \Sigma_{i=1}^n \hat{\varepsilon}_{in}^2 X_i^T X_i$. Finally let R be a known fixed $q \times k$ matrix and let r be a fixed known $q \times 1$ vector.

A4) There exists positive constants S and Δ such that for all i $E(|X_{ij}^2 X_{ik} X_{i\ell}|^{1+S}) < \Delta$ $j,k,\ell = 1,2,\ldots,K$.

With the notation developed above, White then proves the following:

LEMMA 2.1. Given A1 and A2, $\hat{B}_n$ exists almost surely for n sufficiently large and $\hat{\beta}_n \overset{a.s.}{\to} \beta_0$.

LEMMA 2.2.

Under A1 - A3,

$$\sqrt{n}\ \bar{V}_n^{\frac{1}{2}}\ M_n(\hat{\beta}_n-\beta_0) \overset{A}{\sim} N(0,I_K) \quad .$$

THEOREM 2.3.

(i) $|\hat{V}_n - \bar{V}_n| \overset{a.s.}{\to} 0$ under A1, A2, A3(a) and A4.

(ii) $|(X^TX/n)^{-1}\ V_n(X^TX/n)^{-1} - \bar{M}_n^{-1}\ \bar{V}_n\ M_n^{-1}| \overset{a.s.}{\to} 0$.

(iii) $n(R\hat{\beta}_n-r)^T[R(X^TX/n)^{-1}\ \hat{V}_n(X^TX/n)^{-1}R^T]^{-1}(R\hat{\beta}_n-r) \overset{A}{\sim} \chi_q^2$ given the null hypothesis $H_0:R\beta_0 = r$ and under A1 - A4.

3. Block Scalar Covariance Matrix. In some circumstances in which heteroskedasticity is present, the observations can be grouped into a small number of groups such that variances are constant within each group. If the data is so grouped, the covariance matrix Ω will be of a block scalar from where the blocks may have unequal sizes. In general this is equivalent to having a model of the form $Y_{ij} = X_{ij}\beta+\varepsilon_j$, $(j=1,2,\ldots,J,\ i=1,2,\ldots,N_j)$ with $E(\varepsilon_{ij}{}^2) = \sigma_j{}^2$. In this case one tries to estimate the $\sigma_j{}^2$ and then to use these estimates as weights to transform the data. If the ε_{ij} are independently

distributed with normal distribution of zero mean and variance σ_j^2, one can use a maximum likelihood procedure to jointly estimate β and Ω. However, even in this case the computations requires the solution of a nonlinear equation. An alternative procedure, the one proposed in this essay, is to iteratively estimate β and Ω, and then to take as our estimator of β and Ω the limit of their iterations. In the case ε is normally distributed this becomes the procedure proposed in Oberhofer and Kmenta [33]. They however, use an erroneous argument to show that such a limit exists. In this section, it is shown that this iterative procedure, regardless of the form of the likelihood function does converge and the resulting estimator have the usual desirable asymptotic properties. Proofs for the theorems in this section are given in Chapter IV.

Consider the linear model $Y_{ij} = X_{ij}\beta_0 + \varepsilon_{ij}$ $(j = 1,2,\ldots,J;\ i = 1,2,\ldots,m_j)$ which can also be written as $Y_j = X_j\beta_0 + \varepsilon_j$ $(j = 1,\ldots,J)$ or $Y = X\beta_0 + \varepsilon$ where:

1) Y_{ij} and ε_{ij} are real valued;
2) X_{ij} is a $1\times K$ vector with real entries;
3) β_0 is a $K\times 1$ fixed vector with real entries;
4) Y_j, ε_j are $m_j\times 1$ vectors with real entries whose i^{th} elements are Y_{ij} and ε_{ij} respectively;
5) X_j is a $m_j\times K$ matrix with real entries whose i^{th} row is given by X_{ij};
6) Y and ε are $N\times 1$ vectors formed by stacking Y_j and ε_j, $j = 1,\ldots,J$, and where $N = \Sigma_{j=1}^{J} m_j$; and
7) X is a $N\times K$ matrix with real entries formed by stacking X_j, $j = 1,\ldots,J$.

The problem is to estimate β_0 where Y and X are observed, and where

$$E(\varepsilon_{ij}) = 0 \quad \text{and} \quad E(\varepsilon_{ij},\varepsilon_{r,s}) = \begin{cases} 0 & (i,j) \neq (r,s) \\ \sigma_j^2 & (i,j) = (r,s) \end{cases} .$$

If the σ_j^2 are all known, then the appropriate linear estimator is weighted least squares, where $(Y_{ij},\ X_{ij})$

are weighted by $1/\sqrt{\sigma_j^2}$. If there is no knowledge about $\{\sigma_j^2 | j=1,\ldots,J\}$ but the likelihood function of $(Y|X)$ is known, then one may use a maximum likelihood estimator which in general requires the solution of nonlinear equations. An alternative to the maximum likelihood estimator which is also often used when the likelihood function is unknown is a weighted least squares procedure. Estimates of σ_j^2 replace the true values of σ_j^2 in the first mentioned weighted least squares procedure above.

In this paper we discuss an iterative weighted least square procedure that can be described as follows.

Step 1. Select any J positive numbers and denote them by $\sigma_{11}^2,\ldots,\sigma_{1J}^2$. Let Σ_1 be the block scalar matrix where

$$\Sigma_1 = \begin{bmatrix} \overbrace{\sigma^2_{i,1} \quad \sigma^2_{1,1}}^{m_1} & & & O \\ & \overbrace{\sigma^2_{\;,2} \quad \sigma^2_{1,2}}^{m_2} & & \\ & & \ddots & \\ O & & & \overbrace{\sigma^2_{1,J} \quad \sigma^2_{1,J}}^{m_J} \end{bmatrix}$$

Step 2. Let $B_1(\sigma^2_{1,1}, \sigma^2_{1,2}, \ldots, \sigma^2_{1,J}, X, Y) = (X^T \Sigma_1^{-1} X)^{-1} X^T \Sigma_1^{-1} Y$.

Step 2n+1. Let $\sigma^2_{n+1,j} = \frac{1}{m_j} (Y_j - X_j B_n)^T (Y_j - X_j B_N)$ and let

$$\Sigma_{n+1} = \begin{bmatrix} \sigma^2_{n+1,1} & & & & & \\ & \ddots_{m_1} & & & & 0 \\ & & \sigma^2_{n+1,1} & & & \\ & & & \ddots & & \\ & & & & \sigma^2_{n+1,J} & \\ & 0 & & & & \ddots_{m_j} \\ & & & & & \sigma^2_{n+1,J} \end{bmatrix}$$

Step 2n+2. Let $B_{n+1} (\sigma^2_{11}, \ldots, \sigma^2_{1J}\ X, Y) = (X^T \Sigma_{n+1}^{-1} X)^{-1} X^T\ \Sigma_{n+1}^{-1}\ Y$.

Choose as the estimate of β_0 any limit point of $\{B_n(\sigma^2_{11}, \ldots, \sigma^2_{1J}, X, Y)\ n \geq 1\}$.

In this paper we shall show that under fairly general assumption:

(1) $\{B_n(\sigma^2_{11}, \ldots, \sigma^2_{1J}, Y, X) | n \geq 1\}$ has a limit point for all pairs (Y,X) satisfying the assumptions.

(2) For any $\varepsilon > 0$, there exists $N_\varepsilon < \infty$ such that $N \geq N_\varepsilon \Rightarrow \text{prob}\{\{B_n(\sigma^2_{11}, \ldots, \sigma^2_{1J}\ X, Y)\}_{n \geq 1}$ converges (to a unique limit point) is greater than $1-\varepsilon$. Where $N = \Sigma_{j=1}^J m_j$.

(3) If an estimator Θ satisfies that for some fixed $\sigma_{11},\ldots,\sigma_{iJ}$, $\Theta(X,Y)$ is a limit point of $\{B_n(\sigma_{11},\ldots,\sigma_{1J},X,Y)\}_{n\geq 1}$ then $\underset{N\to\infty}{\text{p lim}}\ \Theta(X,Y) = \beta_0$; i.e., Θ is a consistent estimator of β_0, and

(4) Under additional assumption on X and ε_j, the estimator described above has the properties that

a) $\underset{N\to\infty}{\text{p lim}}\ (\Theta-\beta_0) = 0$ and

b) $\sqrt{N}(\Theta-\beta_0)$ is asumptotically distributed $N(0,Q^{-1})$ where Q is a fixed positive definite matrix.

c) If we let σ_j^2 be given by $\frac{1}{m_j}(Y_j-X_j\Theta)^T(Y_j-X_j\Theta)$ and let Σ be given by

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & & & & & & & & 0 \\ & \ddots_{m_1} & & & & & & & & \\ & & \sigma_1^2 & & & & & & & \\ & & & \sigma_2^2 & & & & & & \\ & & & & \ddots_{m_2} & & & & & \\ & & & & & \sigma_2^2 & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & \sigma_J^2 & & \\ & & & & & & & & \ddots_{m_j} & \\ 0 & & & & & & & & & \sigma_J^2 \end{bmatrix}$$

then $\hat{Q} = \frac{1}{N} X^T\Sigma^{-1}X$ is a consistent estimator of Q and $(\hat{Q})^{-1}$ is a consistent estimator of Q^{-1}.

Definition and Conventions.

The model:

$$Y_{ij} = X_{ij}\beta_0 + \varepsilon_{ij} \qquad (j = 1,\ldots,J, \quad i = 1,\ldots,m_j) \text{ or}$$

$$Y_u = X_j\beta_0 + \varepsilon_j \qquad (j = 1,\ldots,J) \text{ or}$$

$$Y = X\beta_0 + \varepsilon \qquad \text{where}$$

Y_{ij}, X_{ij}, ε_{ij}, Y_j, X_j, ε_j, Y, X, ε, and β_0 are as described in the previous section.

Assumption I: The data matrix X is nonstochastic.

Assumption II: There exists λ, T, $0 < \lambda \leq T < \infty$ such that for each j and any m_j

$$\lambda \leq \inf_{\substack{Z \varepsilon \mathbb{R}^k \\ \|Z\|=1}} \frac{Z^T X_j^T X}{m_j} Z \leq \sup_{\substack{Z \varepsilon \mathbb{R}^k \\ \|Z\|=1}} \frac{Z^T X_j^T X_j}{m_j} Z \leq T \quad .$$

Assumption III: For all j and all m_j
$\inf_{B \varepsilon \mathbb{R}^k} \|Y_j - X_j B\| \neq 0.$

Assumption IV: The observed values taken by the dependent variable are realizations of an N element random vector Y which can be written $Y = X\beta_0 + \varepsilon$. X is $N \times K$ matrix with real entries satisfying assumptions I and II and β_0 is a $K \times 1$ vector of unknown real numbers. ε is an N element disturbance vector with $E(\varepsilon|X) = E(\varepsilon) = 0$ and with $V(\varepsilon|X) = V(\varepsilon|X) = V(\varepsilon)$ being a diagonal matrix comprised of J blocks, and being of the form:

$$V(\varepsilon) = \left[\begin{array}{c|c|c} \sigma_1^2 I_2 & 0 & 0 \\ \hline 0 & \sigma_2^2 I_2 & 0 \\ \hline 0 & 0 & \sigma_J^2 I_j \end{array}\right]$$

where I_j is the $m_j \times m_j$ identity matrix and σ_j^2 is a positive unknown real number.

Assumption V: 1) For each j, either ε_{ij} are iid random variables, or

2) for all j and all N, $M < \infty$ such that

$$\sum_{i=1}^{m_j(N)} \frac{VAR(\varepsilon_{ij}{}^2)}{i^2} < M$$

and $\{\varepsilon_{ij}\}$ are independently distributed.

<u>DEFINITION</u>: $z_j(m_1,\ldots,j_m\ X,Y,B) = \frac{1}{m_j}(Y-X_jB)^T(Y_j-X_jB)$. When there is no danger of confusion $z_j(m_1,\ldots,m_j\ X,Y,B)$ will simply be denoted by $z_j(B)$ or $z_j^N(B)$, where $N = \Sigma_{j=1}^J m_j$.

<u>DEFINITION</u>: $Z(m_1,\ldots,m_J\ X,Y,B) = (Z_1(B),Z_2(B),\ldots, Z_j(B))$, this will frequently be denoted by $Z(B)$ or $Z^N(B)$. $(Z_j(B) := z_j(B))$.

<u>DEFINITION</u>: $\Sigma(m_1,\ldots,m_j\ X,Y,B) =$

$$\left[\begin{array}{c|c|c} Z_1(B)I_1 & 0 & 0 \\ \hline 0 & Z_2(B)I_2 & 0 \\ \hline 0 & 0 & Z_J(B)I_J \end{array}\right]$$

where I_j is the $m_j \times m_j$ identity matrix.

$\Sigma(m_1,\ldots,m_j\ X, Y, B)$ will also be denoted by $\Sigma(B)$ or $\Sigma^N(B)$.

In the following section, we state and prove our results concerning the iterative estimator. The iterative estimator is constructed by choosing a limit point from the sequence generated by the iterative process. In the following sequence of propositions and theorems, it is proved that the iterative process generate sequences with limit points, that as the sample size increases, the set of limit points degenerate to a singleton almost surely, and the resulting estimator is strongly consistent.

PROPOSITION 3.1. Let $(\sigma^2_{1,.},\ldots,\sigma^2_{1,j}) \in \mathbb{R}^{+J}$, for observed data matrix X and dependent vector Y satisfying assumption I - IV:

1) Let $\Sigma_1 = \begin{bmatrix} \sigma^2_{1,1} I_1 & 0 & 0 \\ 0 & \sigma_{1,2}^2 I_2 & 0 \\ 0 & 0 & \sigma_{1,J}^2 I_j \end{bmatrix}$

2) Let $B_1 = (X^T \Sigma_1^{-1} X)^{-1} X^T \Sigma_1^{-1} Y.$

3) Let $\Sigma_{n+1} = \Sigma(B_n).$

4) Let $B_{n+1} = (X^T \Sigma_{n+1}^{-1} X)^{-1} X^T \Sigma_{n+1}^{-1} Y.$

Then

A) For each $n \geq 1$ B_n exists;

B) The sequence $\{B_n\}$ has at least one limit point;

C) If B^* is a limit point of the sequence $\{B_n\}$ then $B^* = (X^T \Sigma(B^*)^{-1} X)^{-1} X^T \Sigma(B^*)^{-1} Y.$

If the ε_{ij} are normally distributed, then from part c of Proposition 3.1. it is easy to see that the tuple B^*, $Z_1(B^*), \ldots, Z_j(B^*)$ satisfies the first order conditions for maximizing the likelihood function. To see this, we need only observe that part c of Proposition 3.1.implies that $X^T \Sigma(B^*)^{-1} Y - X^T \Sigma(B^*)^{-1} X B^* = 0$ and that (if we consider B to be a column vector) $\frac{dh}{dB} = kH(B)\ [X^T\Sigma(B)^{-1}Y -$ $X^T\Sigma(B)^{-1}XB]$. (K is a constant independent of B). Therefore in the case that ε_{ij} are normally distributed we have that any estimator Θ with the property that there exists $(\sigma_{11}^2,\ldots,\sigma_{1J}^2) \in \mathbb{R}^{+J}$ such that $\Theta(X,Y)$ is a limit point of $B_n(\sigma_{11}^2,\ldots,\sigma_{1J}^2,\ m_1,\ldots,m_J,\ X,\ Y)$ will be consistent. The next result shows that the normality assumption is superfluous.

PROPOSITION 3.2. Let $g:\mathbb{R}^k$ be defined by $g(B) = (X^T\Sigma(B)^{-1}X)^{-1}\ X^T\Sigma(B)^{-1}Y\ [=(\frac{1}{N}\ X^T\Sigma(B)X)^{-1}$ $(\frac{1}{N}\ X^T\Sigma(B)^{-1}Y)]$. Let F be the set of fixed points of g, and let $d = \text{Sup}\{\|B-B_0\| : B \in F\}$. Let $\gamma_0 > 0$, then under assumptions I-V, $d < \gamma_0$ almost surely as $N \uparrow \infty$.

COROLLARY 3.3. Let Θ be an estimator with the property that for each N, there exists positive number $\sigma_{1,1}{}^2,\ldots,\sigma_{1,J}{}^2$ (perhaps depending upon N) such that for each pair (X,Y), $\Theta(X,Y)$ is a limit point of $\{B_n\ \sigma_{1,1}{}^2,\ldots,\sigma_{1,J}{}^2,X,Y)\}$. Then under Assumptions I-V, Θ is strongly consistent.

PROPOSITION 3.4. Let Θ be an estimator with the property for each N, there exists positive numbers $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$ such that for each pair (X,Y), $\Theta(X,Y)$ is a limit point of $\{B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2, X,Y)\}$. Then under Assumptions I-V, the sequence $B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2, X, Y)$ is convergent almost surely as $N \uparrow \infty$.

The next result is a minor improving of the last proposition. It will however pave the way for showing that almost surely as N gets large, our estimator is independent of the $(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2)$ selected.

PROPOSITION 3.5. Let $g:\mathbb{R}^k \to \mathbb{R}^k$ be defined by $g(B) = (\frac{1}{N} X^T\Sigma(B)^{-1}X)^{-1} (\frac{1}{N}X^T\Sigma(B)^{-1}Y)$. Let

$$K(N) := \sup_{\substack{B_1 \in \mathbb{R}^k \\ B_2 \in \mathbb{R}^k \\ \|B_1 - B_0\| \le 1 \\ \|B_2 - B_0\| \le 1 \\ 0 < \|B_1 - B_2\|}} \frac{\|g(B_1) - g(B_2)\|}{\|B_1 - B_2\|} .$$

Then $K(N) < 1/2$ almost surely as $N \uparrow \infty$.

Let $g:\mathbb{R}^k \to \mathbb{R}^k$ be defined as before by $g(B) = (X^T\Sigma(B)^{-1}X)^{-1}\ (X^T\Sigma(B)^{-1}Y)$. Suppose $g(B) = B$ implies $\| B - B_0\| \leq 1$ and that $\| B_1 - B_0\| \leq 1$ and $\| B_2 - B_0\| \leq 1$ implies that $\| g(B_1) - g(B_2)\| \leq 1/2\ \| B_1 - B_2\|$, then it follows that g has a unique fized point. Since, if both B_1 and B_2 are fixed points we have $\| B_1 - B_2\| = \| g(B_1) - g(B_2)\| \leq 1/2\ \| B_1 - B_2\|$ and thus $\| B_1 - B_2\| = 0$.

PROPOSITION 3. 6. Let $g:\mathbb{R}^k \to \mathbb{R}^k$ be defined as in Proposition 3.5. Let $F = \{B \in \mathbb{R}^k | g(B) = B\}$. Let $h(B) = (2\pi e)^{-N/2} \left\{ \Pi_{j=1}^{J} \left[\frac{(Y_j - X_jB)^T(Y_j - X_jB)}{m_j} \right]^{m_j} \right\}^{-1/2}$ then:

1) F is a singleton almost surely as $N \uparrow \infty$.

2) There exists a unique B that maximizes h almost surely as $N \uparrow \infty$.

3) $F = B^*$ where B^* is the unique element of $\mathbb{R}^k$ that maximizes h almost surely as $N \uparrow \infty$.

Proof. Proposition 3.2 gives us that

1) $\displaystyle\sup_{B \in F} \| B - B_0\| < 1$ almost surely as $N \uparrow \infty$.

Proposition 3.5 yields that:

2) $$\sup_{\substack{\|B_1-B_0\|\le 1\\ \|B_2-B_0\|\le 1\\ B_1\ne B_0}} \frac{\|g(B_1)-g(B_2)\|}{\|B_1-B_2\|} \le 1/2$$ almost surely as $N\uparrow\infty$.

Proposition 3.1 shows that F cannot be empty, thus we have;

3) F is a singleton almost surely as $N\uparrow\infty$.

Proposition 3.1 gives us that there exists $B\in\mathbb{R}^k$ that maximize h and furthermore any such B is a fixed point of g. The rest follows immediately.

COROLLARY 3.7. Let $\Theta(X,Y)$ be an estimator with the property that for each N there exist $\sigma_{1,1}^2(N),\ldots,\sigma_{1,J}^2(N)$ positive numbers such that $\Theta(X,Y)$ is a limit point of $B_n(\sigma_{1,1}^2(N),\ldots,{}_{1,J}$ $\{B_n(\sigma_{1,1}^2(N),\ldots,\sigma_{1,J}^2(N),X,Y))\}$ then:

1) $\Theta(X,Y)$ maximizes the function

$$h(B) = (2\pi e)^{-N/2}\left\{\Pi_{j=1}^{J}\left[\frac{(Y_j-X_jB)^T(Y_j-X_jB)}{m_j}\right]^{m_j}\right\}^{-1/2}$$

almost surely as $N\uparrow\infty$.

2) $\Theta(X,Y)$ is independent of the choices for $\sigma_{1,1}^2(N),\ldots,\sigma_{1,J}^2(N)$ almost surely as $N \uparrow \infty$.

Proof. For any choice of $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$, the limit points of $B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2,X,Y)$ are fixed points of g. Part 1 of the corollary now follows.

If B_1 is a fixed point of g, and if $\sigma_{1,j}^2 = Z_j(B_1)$, then we have $B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2,X,Y) = B_1$ for all n. Therefore if $F = \{B^*\}$, where F is the set of fixed points of g, then it follows that for any choice of $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$ the sequence $\{B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2\ X,Y)\}$ must converge to B^*. Therefore whenever F is a singleton, Θ is independent of the choice of $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$.

One might conjecture that the set F of fixed points of g is always a singleton. The following example shows that this is not the case.

EXAMPLE:

Let	Y_{t1}	X_{t1}	Y_{t2}	X_{t2}
$t = 1$	1	0	1	0
$t = 2$	2	1	2	-1

Let $B = 0$, then $Z_1(0) = Z_2(0) = 5/2$ so $g(0) = (5/4)0 = 0$.

$$h(0) = (2\pi e)^{-2} (5/2)^{-2}$$
$$h(2) = (2\pi e)^{-2} (17/4)^{-1} \quad .$$

$\therefore$ $h(2) > h(0)$ so 0 does not maximize h, hence g has at least one other fixed point. Maximizing h, we find that $B = \pm \sqrt{3}$ are also fixed points of g.

In this section we analyze the asymptotic distributional properties of our proposed estimator. Since this estimator equals the maximum likelihood estimator (when the ε_{ij} are normally distributed) almost surely as the sample size grows large, it is not surprising that it has optimal asymptotic properties.

Before proceeding to this analysis we need to introduce the following assumptions:

Assumption VI: For each j, $\lim\limits_{N\to\infty} \frac{m_j(N)}{N}$ exists

Assumption VII: The $\lim\limits_{N\to\infty} \frac{1}{N} X^T \Sigma_N^{-1} X$ exists where $\frac{1}{N} X^T \Sigma_N^{-1} X = \Sigma_{j=1} \frac{m_j(N)}{N} \frac{X_j^T X_j}{\sigma_j^2}$ and Σ_N is the appropriate diagonal matrix.

$$\Sigma_N = \begin{bmatrix} \sigma_1^2 I_{N1} & 0 & 0 \\ 0 & & 0 \\ 0 & 0 & \sigma_J^2 I_{Nj} \end{bmatrix}$$

where I_{Nj} is the $m_j(N) \times M_j(N)$ identity matrix.

Assumption VII (Replaces Assumption V). For all N, $\{\varepsilon_{ij}/\sqrt{\sigma_j^2} \mid j = 1,2,\ldots,J \quad i = 1,2,\ldots,m_j(n)\}$ is a collection of independent and identically distributed random variables.

LEMMA 3.8. Let $\alpha_N = \Sigma_{j=1}^{J} \frac{m_j(N)}{N} \sigma_j^2$. Then:

A) $\lim_{N\to\infty} \alpha_N := \alpha$ exists and $\max_j \sigma_j^2 \geq \alpha_N \quad \min_j \sigma_j^2$ all N

B) trace $\alpha_N^{-1} \Sigma_N = N$ all N

C) $\lim_{N\to\infty} \frac{1}{N} X^T(\alpha_N^{-1} \Sigma_N)^{-1}X$ exists and is positive definite.

PROPOSITION 3.9. Let $\Theta(X,Y)$ be an estimator with the property that $g(\Theta(X,Y) = \Theta(X,Y)$ where $g(B) = (X^T\Sigma(B)^{-1}X)^{-1}(X^T\Sigma(B)^{-1}Y)$, then under the hypothesis above:

A. Θ is asymptotically equivalent to the weighted least squares estimator with known variances in the sense that if W is the latter, $\underset{N\to\infty}{p\lim} N(\Theta-W) = 0$.

B. Θ is asymptotically normally distributed with mean vector B_0 and variance covariance matrix $(X^T\Sigma^{-1}X)^{-1}$ $[= \frac{1}{N} (\frac{1}{N} X^T \Sigma^{-1}X)^{-1}]$.

C. $(\frac{1}{N} X^T \Sigma^{-1} X)^{-1} - (\frac{1}{N} X^T \Sigma(\Theta)^{-1} X)^{-1} \to 0$ almost surely as $N \uparrow \infty$.

It is impossible numerically to identify the limit of a sequence. In paractice, one calls a term of a computed sequence a limit if it differs from the previous term by an amount less in norm than some preassigned tolerance level. The above theorems state that if this procedure does pick a limit of the iterative procedure, then the statistics computed in this last stage have their usual expected asymptotic distribution, that is, one can do asymptotic hypothesis tests and construct asymptotic confidence intervals using this output.

4. Linear Variance Model. Block scalar covariances arise when the variances are structurally related to discrete valued variables. In the case that the variances are structurally related to at least one continuous valued variable, the likely candidate for the variance structure is that it also follows a linear model. Let ε^2 be the vector whose i^{th} component $(\varepsilon^2)_i$ is given by ε_i^2 and let σ^2 be the vector $E(\varepsilon^2)$. The linear variance model for

heteroskedasticity assumes that the variance covariance matrix Ω is diagonal with diagonal elements $\Omega_{ii} = \sigma_i^2$ and furthermore that $\sigma^2 = Z\Gamma$, where Z is an observed matrix and Γ is unknown. The matrices X and Z may share columns or may be unrelated. The usual problems are to:

1) Estimate Γ sufficiently well, that the estimate of Γ yields estimator of σ^2 that can be well used in estimating β, performing hypothesis tests on β, and constructing confidence intervals on β.

2) Test hypothesis on Γ.

3) Construct confidence intervals for Γ.

Glejser [14] and Park [34] have suggested two similar procedures for estimating and performing hypothesis tests on Γ. Both of these procedures are supported by heuristic arguments but both are known to lead to inconsistent hypothesis tests. In this section a new procedure which uses White's theorem is proposed which has the advantages of being easy to construct and understand, and which is shown to yield, consistent test statistics for testing hypothesis on Γ. This procedure yields estimates for Γ and thus for σ^2 that can be used in a weighted least squares procedures to estimate β. If β is estimated by this multiple stage weighted least squares procedures, the final stage estimate for β together with the usual generated statistics have the expected known asymptotic distributions.

In this section we give an heuristic description of Glejser's procedure and describe our analysis. In the following section we present and prove our two main theorems. We end the section with some comments about the validity of our assumption and some comments for extending this work.

Let the model be given by:

(1) $Y_i = X_i \beta_0 + \varepsilon_i \qquad i = 1,2,\ldots,n$

where:

(2) $E(\varepsilon_i) = 0 \qquad E(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & i \neq j \\ \sigma_j^{\,2} & i = j \end{cases}$.

We also assume that:

(3) $\sigma_j^{\,2} = Z_i \Gamma_0 \qquad i = 1,\ldots,n$.

Y_i is an observable real valued variable, X_i and Z_i are vectors of real valued variables, X_i has dimensions $1 \times k$ and Z_i $1 \times m$. It is possible that X_i and Z_i share common components. β_0 and Γ_0 are unknown parameter vectors $\beta_0 \varepsilon \mathbb{R}^k$ $\Gamma_0 \varepsilon \mathbb{R}^m$. $\varepsilon_i (i=1,2,\ldots,n)$ are real valued unobservable random variables. $\sigma_j^{\,2}$ are positive real numbers also unobservable. We assume that the vectors $(\varepsilon_i, X_i, Z_i)$ $i = 1,\ldots,n$ are independently distributed.

Following Glejser's suggestion we let $\hat{\beta}_n = (X^TX)^{-1}X^TY$, the OLS estimate of β_0, and then we let $\hat{\varepsilon}_{in} = Y_i - X_i\hat{\beta}_n$, $\hat{\varepsilon}_{in}$ is the ordinary least squares residual. Now we estimate Γ_0 by $\hat{\Gamma}_n = (Z^TZ)^{-1} Z^T\hat{\varepsilon}_n^{\,2}$, where $\hat{\varepsilon}_n^{\,2}$ is the column vector whose i^{th} component is $(\hat{\varepsilon}_{in})^2$. Glejser then suggests we perform our hypothesis tests on Γ_0 by using $\hat{\Gamma}_n$. If our observed $\hat{\Gamma}_n$ "supports" the model $\sigma_j^{\,2} = Z_j\Gamma_0$, we then use a weighted least square procedure to reestimate β_0 using Z_j $\hat{\Gamma}_n$ as our estimate of the variance $\sigma_j^{\,2}$. In Glejser's paper, he proposes the model

$\varepsilon_j = \nu_j P_g(Z_j) = \nu_j\{m_0 + m_1 f\,(Z_j) + \cdots + m_g f(Z_j)]^g\}$, where ν_j are independent random variables, $E(\nu_j) = 0$ $E(\nu_i\nu_j) = \delta_{ij}\sigma^2$. The Z_j and function f is known, m_j is unknown and each term m_k $f(Z_j)^k$ is assumed positive. Thus he has $\sigma_j^{\,2} = \sigma^2[P_g(Z_j)]^2$, and by taking absolute values and then expectations gets $E|\varepsilon_j| = E(|\nu_j|)\cdot P_g(f(Z_j))$. He then suggests estimating m_i by regressing $|\hat{\varepsilon}_{in}|$ on the values $[f(Z_j)]^i$. R. E. Parks [34] suggests a similar procedure where he assumes $\sigma_j^2 = \sigma^2 Z_j^\gamma e_j^{w_j}$, taking logarithms he gets $\ell n\sigma_j^2 = \ell n\sigma^2 + \gamma\ell nZ_j + w_j$. He then uses $\hat{\varepsilon}_{jn}^2$ to replace σ_j^2 and performs a regression on $\ell n|\hat{\varepsilon}_{jn}^2|$ to estimate $\ell n\sigma^2$ and γ. We find that for our analysis it is more convenient to use the model resulting in (3), that is $\varepsilon_j = \nu_j(Z_j\Gamma_0)^{\frac{1}{2}}$, with $E(\nu_j^2) = \sigma_\nu^2 = 1$. $E(\nu_j) = 0$. We then get (3), $\sigma_j^2 = Z_j\Gamma_0$.

The apparent problem with all of these procedures is, as Glejser observes, that the estimated coefficients are biased. Glejser optimistically states that we should ignore the bias effect in the hope that it will generally be unimportant compared to other contributing terms.

Returning to our model for the variances, we observe that:

(4) $\varepsilon_j^2 = Z_j\Gamma_0 + (\varepsilon_j^2 - \sigma_j^2)$ and that $E(\varepsilon_j^2 - \sigma_j^2) = 0$.

Therefore, if we could only observe ε_j^2, the OLS regression of Z_j on ε_j^2 would yield a consistent estimator for Γ_0. Of course, the error term $\varepsilon_j^2 - \sigma_j^2$, while having zero expectation, need not have constant variance. So it would appear that we have again returned to the problem of hypothesis testing in the presence of heteroskedasticity. Now, however, we are in the position of using White's procedure to generate a consistent estimator of the variance covariance matrix and are able to correctly perform asymptotic χ^2 tests on hypothesis concerning restrictions onthe parameter vector Γ_0. Unfortunately we cannot observe ε_j^2, it is the substance of this paper that we can replace ε_j^2 by $\hat{\varepsilon}_{jn}^2$ and that in so doing we will get estimation and statistics that are asymptotically equivalent to those when we used ε_j^2.

In the next section we formally state our principle result, proofs are given in Chapter IV. Before doing this, however, it is necessary to introduce additional notation. We also state without proof some elementary propositions on convergence in probability and almost sure convergence. In the next section we shall make frequent reference to convergence in norms. Thus for real valued variables x_1, $\|x_i\|$ is the absolute value: if x is a $\ell\times 1$ vector in

$\mathbb{R}^{\ell}$, or a $1\times\ell$ vector in $\mathbb{R}^{\ell}$ the norm is the standard Euclidean norm. Since $(\mathbb{R}^{\ell})^* = \mathbb{R}^{\ell}$, this norm is the same as the dual norm. For X a matrix in $\mathbb{R}^{n\times k}$, we use the norm in $L(\mathbb{R}^k, \mathbb{R}^n)$, that is $\| X \|_{L(\mathbb{R}^k, \mathbb{R}^n)} = \sup_{\substack{y\in\mathbb{R}^k \\ \|Y\|=1}} \|XY\|_{\mathbb{R}^n}$. Since on a finite dimensional Banach Space all norms induce the same topology, if x_n is a sequence of matrices in $\mathbb{R}^{k\times j}$, then $\| x_n - x_0 \|_{L(\mathbb{R}^k, \mathbb{R}^j)} \to 0$, if and only if

$\| X_{n(i,j)} - X_{0(i,j)} \| \to 0$. That is we get convergence in the operator norm of $L(\mathbb{R}^k, \mathbb{R}^j)$ iff we have convergence for each matrix entry and hence if we have convergence in the Euclidean (Hilbert-Schmidt) norm.

We are now in a position to list our assumptions and to prove our results. For the convenience of the reader, we have followed much of the notation of White [59]. We have also borrowed liberally from him on the wording of our assumption.

A1) The model is known to be

$$Y_i = X_i\beta_0 + \varepsilon_i \qquad i = 1,2,\ldots,n$$

$$E(\varepsilon_i) = 0 \qquad i = 1,2,\ldots,n$$

$$E(\varepsilon_i^{\,2}) = \sigma_i^{\,2} = Z_i\Gamma_0 \quad .$$

Where X_i is a $1\times k$ vector of random variables, ε_i and Y_i are real valued random variables, β_0 is a $k\times 1$ vector of real numbers. Y_i and X_i are observable, ε_i is unobservable and β_0 is to be estimated or hypothesis concerning β_0 are to be tested. Z_i is a $1\times m$ vector of real valued random variables which may contain some or all of the variables in the vector X_i. Γ_0 is a $m\times 1$ unknown vector of real numbers which is to be estimated or hypothesis concerning Γ_0 are to be tested.

Let W_i be the vector of length p of random variables whose first entry W_{i1} is the scalar 1 and whose other entries are exactly those random variables that appear in X_i or Z_i. We assume that $E(W_{ij}W_{ir}\varepsilon_i) = 0 \quad 1 \leq j,k \leq p$ and $E(W_i^T(\varepsilon_i^2-\sigma_i^2)) = 0$. We let μ_i denote $\varepsilon_i^2 - \sigma_i^2$. The vectors (W_i,ε_i) are assumed to be a sequence of independent though not necessarily identically distributed random vectors.

A2) I) There exists $0 < \delta \leq 1$ and $\Delta < \infty$ such that for all i:

a) $E(|\varepsilon_i^2 W_{ir}W_{is}W_{it}W_{iv}|^{1+\delta}) \leq \Delta \quad 1 < r,s,t,v \leq p$

b) $E(|\varepsilon_i W_{ik}W_{ir}W_{is}W_{it}W_{iv}|^{1+\delta}) \leq \Delta \quad 1 \leq k,r,s,t,v \leq p$

c) $E(|W_{ij}W_{ik}W_{ir}W_{is}W_{it}W_{iv}|^{1+\delta}) \leq \Delta \quad 1 \leq j,k,r,s,t,v \leq p$

d) $E(|\mu_i^2|^{1+\delta}) \leq \Delta$

e) $E(|\varepsilon_i^3 W_{ir} W_{is} W_{it}|^{1+\delta}) \leq \Delta \qquad 1 \leq r,s,t \leq p$

f) $E(|\varepsilon_i^4 W_{ir} W_{is}|^{1+\delta}) \leq \Delta \qquad 1 \leq r,s, \leq p$

II) Let $M_n^a = n^{-1} \sum_{i=1}^n E(X_i^T X_i)$ and let

$$M_n^b = n^{-1} \sum_{i=1}^n E(Z_i^T Z_i)$$

We assume that there exists $N_0 < \infty$ and $0 < \lambda$ such that for $n \geq N_0$, the minimum eigenvalues of M_n^a exceeds λ and the minimum eigenvalue of M_n^b exceeds λ; (Note by the first part of A2) this is equivalent to the property that for n sufficiently large det M_n^a and det M_n^b is bounded away from zero. Also, observe that we can choose δ and λ so that they are equal.

A3) Let $V_n^a = n^{-1} \Sigma_{i=1}^n E(\varepsilon_i^2 X_i^T X_i)$ and

let $V_n^b = n^{-1} \Sigma_{i=1}^n E(\mu_i^2 Z_i^T Z_i)$.

We assume that there exists $N_0 < \infty$ and $\lambda > 0$ such that for $n \geq N_0$, minimum eigenvalues of V_n^a exceed λ and the minimum eigenvalue of V_n^b exceeds λ. (There is no loss in generality in assuming that N_0, λ is A2 and N_0, λ in A3 are the same). In the presence of A2, A3 is the equivalent of the assumption that for n sufficiently large minimum $(\det V_n^a, \det V_n^b)$ is bounded away from and above zero.

The first theorem is a restatement of a result found in White [ ] and its proof can be found therein. Before stating Theorem 3. we introduce additional notations.

$$\text{Let } \hat{\beta}_n = \begin{cases} (X^TX)^{-1}X^TY & \text{if } (X^TX) \text{ is nonsingular} \\ 0 & \text{if } (X^TX) \text{ is singular} \end{cases} .$$

$$\text{Let } \hat{\alpha}_n = \begin{cases} (Z^TZ)^{-1}Z^T\varepsilon^2 & \text{if } (Z^TZ) \text{ is nonsingular} \\ 0 & \text{if } (Z^TZ) \text{ is singular} \end{cases}$$

ε^2 is the $n\times 1$ column vector whose i^{th} entry is $(\varepsilon_i)^2$).

Let $\hat{\varepsilon}_{in} = Y_i - X_i\hat{\beta}_n$.

Let $\hat{\mu}_{in} = \varepsilon_i{}^2 - Z_i\hat{\alpha}_n$.

Let $\hat{V}_n^a = n^{-1} \Sigma_{i=1}^n \hat{\varepsilon}_{in}^2 X_i^T X_i$.

Let $V_n^b = n^{-1} \Sigma_{i=1}^n \hat{\mu}_{in}^2 Z_i^T Z_i$.

Let R^a be a $q\times k$ matrix of real numbers with full row rank and let r^a be a $q\times 1$ vector of real numbers.

Let R be a $q\times m$ matrix of real numbers of full row rank and let r be a $q\times 1$ vector of real numbers.

THEOREM 4.1. (White) Under Assumptions A1, A2, and A3, we have the following:

i) $\hat{\beta}_n \to \beta_0$ almost everywhere (a.e.).

ii) $\hat{\alpha}_n \to \Gamma_0$ a.e.

iii) $\sqrt{n}\ [(\frac{X^T X}{n})^{-1}\ \hat{V}_n^a\ (\frac{X^T X}{n})^{-1}]^{\frac{1}{2}}\ (\hat{\beta}_n - \beta_0) \overset{A}{\sim} N(0, I_k)$

iv) $\sqrt{n}\ [(\frac{Z^T Z}{n})^{-1}\ \hat{V}_n^b\ (\frac{Z^T Z}{n})^{-1}]^{\frac{1}{2}}\ (\hat{\alpha}_n - \alpha_0) \overset{A}{\sim} N(0, I_m)$

v) under the hypothesis H_0: $R^a \beta_0 = r^a$

$$n(R^a \hat{\beta}_n - r^a)^T [R^a\ (\frac{X^T X}{n})^{-1}\ \hat{V}_n^a\ (\frac{X^T X}{n}) R^{aT}]^{-1} (R^a \beta_n - r^a) \overset{A}{\sim} \chi_q^2$$

vi) under the hypothesis H_0: $R\Gamma_0 = r$

$$n(r\hat{\alpha}_n - r)^T [R(\frac{Z^T Z}{n})^{-1}\ \hat{V}_n^b\ (\frac{Z^T Z}{n})^{-1} R^T] (R\hat{\alpha}_n - r) \overset{A}{\sim} \chi_q^2$$

Since ${\varepsilon_i}^2$ are not observable, $\hat{\alpha}_n$ and the statistics associated with $\hat{\alpha}_n$ are not computable. It is a principle result of this paper, that we can replace ${\varepsilon_1}^2$ by ${\hat{\varepsilon}_{in}}^2$ and obtain asymptotically equivalent results. This notion is more carefully stated and then proved in the next theorem.

Before stating this theorem it is once again necessary to introduce additional notations.

Let $\hat{\varepsilon}_n^{\,2}$ be the $n\times 1$ column vector whose i^{th} entry is $\hat{\varepsilon}_{in}^{\,2}$.

Let $\hat{\Gamma}_n = \begin{cases} (Z^TZ)^{-1} Z^T\hat{\varepsilon}_n^{\,2} & \text{if } Z^TZ \text{ is non-singular.} \\ 0 & \text{otherwise.} \end{cases}$

Let $\hat{W}_{in} = \hat{\varepsilon}_{in}^{\,2} - Z_i\hat{\Gamma}_n$ and

Let $\hat{V}_n^c = n^{-1} \Sigma_{i=1}^n \hat{W}_{in}^{\,2} Z_i^T Z_i$.

Theorem 4.2. Under assumption A1, A2, and A3, the following hold:

i) $\hat{\Gamma}_n \to \Gamma_0$ a.s.

ii) $\sqrt{n}\ [(\frac{Z^TZ}{n})^{-1}\ \hat{V}_n^c(\frac{Z^TZ}{n})^{-1}]^{-\frac{1}{2}}\ (\hat{\Gamma}_n - {}_0) \overset{A}{\sim} N(0,I_m)$

iii) under the hypothesis H_0: $R\Gamma_o = r$ (where R,r are as in Theorem 3.).

$$n(R\hat{\Gamma}_n - r)^T\ [R(\frac{Z^TZ}{n})^{-1}\ \hat{V}_n^c\ (\frac{Z^TZ}{n})^{-1}R^T]^{-1}(R\hat{\Gamma}_n - r) \sim \chi_q^2 \quad .$$

(Note: A part of the statement of this theorem is that matrices whose inverses must be taken well as $n \uparrow \infty$ be nonsingular almost everywhere].

Thus we have shown that we have a valid asymptotic test for testing linear restrictions on the parameters in the variance model. Our next and last result is to show that under additional assumptions, we can use our estimates from the variance model to reestimate the original model and obtain a estimator that is asymptotically equivalent to weighted least squares with variances known.

A4) There exists $\lambda > 0$ such that for all i $\sigma_i^2 > \lambda$.

A5) For all i, $E(\varepsilon_i | W_{is},\ldots,W_{ip}) =$
$E(\varepsilon_i^2 | W_{is},\ldots,W_{ip}) = 0, E(\varepsilon_i^2 \; W_{i1},\ldots,W_{ip} =$
$= Z_i\Gamma_0$.

A6) There exists $M < \infty$, such that for all i $\| Z_i \| < M$.

Theorem 4.3. Let Ω_n denote the $n\times n$ diagonal matrix whose (i,i) entry is σ_i^2 and let $\hat{\Omega}_n$ denote the $n\times n$ matrix whose (i,i) entry is $Z_i\hat{\Gamma}_n$. Let $\hat{B}_n$ denote the Aitken estimator given by

$$B_n = \begin{cases} (X^T\hat{\Omega}_n^{-1}X)^{-1}X^T\Omega_n^{-1}Y & \text{where } X^T\Omega_n^{-1}X \text{ is nonsingular} \\ 0 & \text{if } X^T\Omega_n^{-1}X \text{ is singular.} \end{cases}$$

Let $\hat{B}_n$ be the weighted least squares estimator given by

$$\hat{B}_n = \begin{cases} (X^T\hat{\Omega}_n^{-1})^{-1}(X^T\hat{\Omega}_n^{-1}Y) & \text{if } X^T\hat{\Omega}_n^{-1}X \text{ is nonsingular} \\ 0 & \text{if } X^T\hat{\Omega}_n^{-1}X \text{ is singular.} \end{cases}$$

Under Assumption A1-A6,

i) $\| \hat{B}_n - \beta_0 \| \to 0$ a.s.

ii) $p \lim \sqrt{n}(\hat{B}_n - B_n) = 0$ and

$$\sqrt{n}(n^{-1}X^T\Omega_n^{-1}X)^{\frac{1}{2}} (\hat{B}_n - \beta_0) \overset{A}{\sim} N(0,I_k) \quad .$$

iii) If R is a $q\times k$ matrix of real numbers with full row rank and r is a $q\times 1$ vector of real numbers, then under H_0: $R\beta_0 = r$,

$$n(R\hat{B}_n - r)^T[R(\frac{X^T\hat{\Omega}_n^{-1}X}{n})^{-1} R^T]^{-1} (RB_n-r) \overset{A}{\sim} \chi_q^2 \quad .$$

CHAPTER IV

MATHEMATICAL PROOFS

1. Introduction. This chapter contains the complete statement and proofs of the theorems developed in Chapters II and III. Because of its mathematical nature, the chapter is designed to stand apart from the rest of the dissertation and may be passed over by those with low mathemtical inclination without loosing the content of the rest of the dissertation. The proofs in this chapter, especially those of theorems appearing in Chapter III may be of interest not only because they yield further insight into the contents of the theorems, but also because they are examples of the application of elementary functional analytic and Banach algebraic techniques to statistical analysis. The norms used in showing convergence in the lemmas and theorems in Chapter III are the operator norms. Since any two Hausedorff topological vector spaces over the same scalar field and of the same finite dimension are isomorphic as topological vector spaces, convergence of a sequence in one norm implies convergence in all norms. Definitions and notation, where not explicitly

restated in this chapter, are taken from Chapters II or III where they are first introduced.

2. Market Search.

A1. For each p in $\mathbb{R}^n$, $p > 0$ and for each f in F, there is a choice $y_f(p)$ such that for all y in $Y_f p y_f(p) \geq py$. Furthermore $y_f(p)$ can be choosen such that the max $p \to y_f(p)$ is continuous.

A2. For each p in $\bigoplus_{f\varepsilon F} \mathbb{R}^n$ with $p(f) > 0$ for each f, and for each h in H, there is a choice of $\chi_h^f(p)$ in $\mathbb{R}^N$ such that

$$p(f)\ (\chi_h^f(p) - \overline{x}_h) \leq \sum_{k\varepsilon F} d(h,k)\ p(k)\ y_k(p(k))$$

and $U_h(\chi_h^f(p)) \geq \sup\{U_h(x):$ for all $x \ \varepsilon\ \mathbb{R}^n$ where $p(f)(x-\overline{x}_n) \leq \sum_{k\varepsilon F} d(h,k)\ p(k)\ y_k(p(k)\}$. Furthermore, $x_h^f(p)$ can be choosen such that the map $p \to x_h^f(p)$ is continuous for each h in H and f in F.

Let C be an arbitrary function from H into F. For each firm f and each p in $\bigoplus_f \mathbb{R}^n$ with $p(f) > 0$ for each f, let $Z_f(p) = \sum_{h:c(h)=f} (\chi_h^f(p)-\overline{x}_h) - Y_f(p)$. Let S_n be the unit Simplex in $\mathbb{R}^n$, that is $S^n: = \{p \ \varepsilon\ \mathbb{R}^n: p > 0\ \Sigma_{i=1}^n p(i) = 1\}$. For $p \ \varepsilon \bigoplus_{f\varepsilon F} S_n$, we have:

$\Sigma_f\, p(f)\, Z_f(p) = \Sigma_f\, \Sigma_{h:c(h)=f}\, p(f)(x_h^f(p) - \bar{x}_n) -$
$\Sigma_f\, p(f)\, Y_f(p)$. If $c(h) = f$, then $p(f) = p(c(h))$, so by A2,

$$\Sigma_f\, \Sigma_{h:c(h)=f}\, p(f)\ (x_h^f(p) - \bar{x}_n) \le$$

$$\Sigma_f\, \Sigma_{h:c(h)=f}\, \sum_{k\varepsilon F} d(h,k)\ p(k)\ Y_k(p(k)) =$$

$$+ \sum_{k\varepsilon F} \sum_{h\varepsilon H} d(h,k)\ p(k)\ Y_k(p(k)) = \sum_{k\varepsilon F} p(k)\ Y_k(p(x)) =$$

$$\sum_{f\varepsilon F} p(f)\ Y_f(p(f)).$$

A3. For no p in $\bigoplus_f S_n$ is it the case that $Z_f(p)(i) > 0$ implies that $p(f)(i) = 1$.

A4. For each h in H, $U_h: \mathbb{R}^n \to \mathbb{R}$ is a continuous function.

THEOREM 2.1. Under assumptions A1, A2, A3 and A4, there exists a search equilibrium with $C^* = C$.

Proof. Let $K = \bigoplus_{h\varepsilon H, f\varepsilon F} S_n \bigoplus_{f\varepsilon F} S_n$, then K is a compact convex subset of the finite Cartesian product of copies of $\mathbb{R}$. We identify an element of K by the pair (p,q), where $p(h,f)$ is in S_n and $q(j)$ is in S_n for each h in H and f,j in F. We seek a continuous function $\psi: K \to K$ that has (p^*,q^*) as a fixed point only if (p^*,q^*) is a price profile for a competitive search equilibrium.

Let $\phi: S_n \times \mathbb{R}^n \to \mathbb{R}^n$ be given by $\phi = (\phi^1, \phi^2, \ldots, \phi^n)$ where $\phi^j: S_n \times \mathbb{R}^n \to \mathbb{R}$ is defined as follows:

$$\phi^1(p_1, p_2, \ldots, p_n, a_1, a_2, \ldots, a_n) = (1-p_1)(a_1 \vee 0)$$

$$\phi^{J+1}(p,a) = (-\phi^1 \vee 0), \ldots, (-\phi^j \vee 0)(1-p_{j+1})(a_{j+1} \vee 0) .$$

Here $(p,a) = (p_1, \ldots, p_n, a_1, \ldots, a_n)$, $\phi^j = \phi^j(p,a)$, and $a \vee b = \max(a,b)$. The function ϕ has the following properties:

a) ϕ is continuous.

b) For all (p,a) in $S_n \times \mathbb{R}^n$, $\phi(p,a) \geq 0$.

c) $\phi(f,a) > 0$ if and only if there is an index i with $p_i \neq 1$ and $a_i > 0$.

d) There exists at most one index j with $\phi^j(p,a) \neq 0$.

Let λ be the function from $\mathbb{R}$ into $\mathbb{R}$ defined by $\lambda(a) = (a \vee 0)/(1 + (a \vee 0))$. Then λ is a continuous function with the property that $\lambda(a)$ is in the half open interval $[0,1)$ and $\lambda(a)$ equals 0 if and only if $a \leq 0$. For h in H, f in F and $p \in \bigoplus_{h \in H, f \in F} S_n$, let $\lambda_h^f(p)$ be defined by

$$\lambda_h^f(p) = \lambda\{U_h[X_h^f(p(h,\cdot))] - U_h[X_h^{c(h)}(p(h,\cdot))]\} .$$

The function ψ can now be defined as follows $\psi(p,q) = (\psi^1(p,q), \psi^2(p,q))$ where $\psi^1 : K \to \bigoplus_{h \in H, f \in F} S_n$ and $\psi^2 : K \to \bigoplus_{f \in F} S_n$. The function ψ^1 is defined by

$$\psi^1(p,q)(h,f) = \begin{cases} \frac{1}{2}\, p(h,f) + \frac{1}{2} q(f) & \text{if } f = c(h) \\ \lambda_h^f(p) \cdot p(h,c(h)) + [1 - \lambda_h^f(p)]\, p(h,f) & \\ \qquad \text{if } f \neq c(h) & \end{cases} .$$

$$\psi^2(p,q)(f) = \frac{q(f) + \phi(q(f), Z_f(q))}{1 + \Sigma_{i=1}^n \phi(q(f), Z_f(q))[i]} .$$

Clearly ψ is a continuous function from K into K and so by Brouwer's fixed point for compact convex subsets of $\mathbb{R}^n$, ψ has a fixed point (p^*,q^*).

Let $x_h^* = x_h^{c(h)}(p^*(h,\cdot))$, let $y_f^* = y_f(g^*(f))$, and let $C^* = C$.

Since $\psi^1(p^*,q^*) = p^*$, it follows directly that $p^*(h,c(h)) = q^*(c(h))$. If $f \neq c(h)$, then since $p^*(h,f) = \psi^*(p^*,q^*)(h,f)$, either $\lambda_h^{*f}(p^*) = 0$ or $p^*(h,f) = p^*(h,c(h))$. In either event, $U(x_h^f(p^*)) \leq U(X_h^{c(h)}(p^*))$. Since $\psi^2(p^*,q^*) = q^*$, $Z_f(q^*) \leq 0$ for each f. Otherwise, by assumption A2, we can find a f and an i with $Z_f(q^*)(i) > 0$ and $q(f)(i) \neq 1$. In this case $\phi(g^*(f),Z_f(q^*)) > 0$, so there is an index j with $\phi^j(q^*(f),Z_f(q)) > 0$ and since $q^*(f)(j) \neq 1$,

$$\frac{q^*(f)(j) + \phi^j(q^*(f),Z_f(q^*))}{1 + \phi^j(q^*(f), Z_f(q^*))} > q^*(f)(j) \quad .$$

This contradicts the fact that (p^*,q^*) is a fixed point for ψ, and so $Z_f(q^*) \leq 0$. The price profile (p^*,q^*), consumption allocation vector X^*, defined by $X^*(h) = X_h^*$, the production allocation vector y^* defined by $y^*(f) = y_f^*$ and the choice C^* defined by $C^* = C$ is then a competitive search equilibrium.

COROLLARY 2.2 In a search equilibrium two different firms may post different prices for identical commodities.

PROOF. In a simple general equilibrium model such as that described in Chapter 2 of Arrow and Hahn [ ], we see that the equilibrium price is not independent of initial endowments $\bar{x}_h$. Consider two simple general equilibrium models identical except for initial endowments $\bar{x}_n$ and the resulting equilibrium price vectors. These models are identified by the parameters $(H,F,\bar{X}_n,p^*)$ and $(H^1,F^1,\bar{X}_h,p^{1*})$, where F and F^1 are both singletons. Now consider the search model with $H \cup H^1$ householders, $F \cup F^1$ firms and choice function C defined by $c(h) = f$ if h is in H and $c(h) = f^1$ if h is in H^1. The consumption vectors x_h^* x_h^2, and production vectors y_h^*, y_h^*, from the simple model will also be the desired equilibrium commodities in the search model. We assign ownership of the two firms as follows:

$$d(h,f) = \begin{cases} S(h,f) & \text{if } h \varepsilon H,\ f \varepsilon F \text{ or } h \varepsilon H^1\ f \varepsilon F^1 \\ 0 & \text{elsewhere} \end{cases} .$$

Here $S(h,f)$ is the assignment of ownership in the original model. For each household h in $H \cup H^1$ we have an ordering on S_n defined by $p \geq q$ if and only if

$$\sup_h \{U(X): X \in \mathbb{R}^n \text{ such that } p(X - \overline{X}_h) \leq \leq d(h,f)p^* y_f^* + d(h,f^1)p^{1*} y_{f^1}^*\}$$

is greater that $\sup_n \{U(X): X \in \mathbb{R}^n$ such that

$$q(X - \overline{X}_h) \leq d(h,f)p^* y_f^* + d(h,f^1)p^{1*} y_{f^1}^*\} .$$

For h in H choose p_h in S_n such that $p^* \geq p_h$ and for h in H^1 choose p_h in S_n such that $p^{1*} \geq p_h$.

Now define a price profile by

$$p(h,f) = \begin{cases} p^* & \text{if } h \in H, \quad f \in F \\ p_h & \text{if } h \in H, \quad f \in F^1 \\ p^{1*} & \text{if } h \in H^1, \quad f \in F^1 \\ p_h & \text{if } h \in H^1, f \in F \end{cases}$$

and

$$g(f) = \begin{cases} p^* & \text{if } f \in F \\ p^{1*} & \text{if } f \in F^1 \end{cases} .$$

Then it is easy to see that $(p,q)\ x_h^*,\ y_f^*\ C$ comprise a competitive search equilibrium, but $q^*(f) \neq q^*(f^1)$.

3. The Housing Search Model.

Lemma 3.1.

For $nx>1$ $\quad \ln(\frac{nx-1}{nx}) + \frac{1}{nx-1} > 0$

Proof.

1) Let $g(z) = \ln(\frac{z}{z+1}) + \frac{1}{z}$

2) $g'(z) = \frac{z+1}{z} \frac{1}{(z+1)^2} - \frac{1}{z^2} = \frac{z^2 - z(z+1)}{z^3\ (z+1)} = \frac{-1}{z^2(z+1)}$

$< 0 \quad (z > 0)$

3) $\lim\limits_{z \to \infty} g(z) = 0$ therefore for $z > 0$ $g(z) > 0$
(g decreased down to zero)

4) for $nx>1$ $\quad \ln(\frac{nx-1}{nx}) + \frac{1}{nx-1} = g(nx-1) > 0$

☒

Lemma 3.2

For $nx>2, 1 < nx \ln \frac{nx}{nx-1} < 1.4$

Proof.

1) Let $g(z) = z \ln \frac{z}{z-1}$

2) $g'(z) = \ln(\frac{z}{z-1}) - \frac{1}{z-1}$, therefore

3) $g''(z) = \frac{1}{(z-1)^2} - \frac{1}{(z)(z-1)}$ $\qquad g''(z) > 0$

$g'(z)$ is increasing

4) $\lim_{z\to\infty} g'(z) = 0$, $g'(z) < 0$ for $z > 1$

5) g is a decreasing function for $z > 1$ and for $z > 2$ $\quad g(z) < g(2) = \ln 4 < 1.4$

6) $nx>2 \quad nx \ln \frac{nx}{nx-1} = g(nx) < g(2) < 1.4$

7) $$\lim_{z\to\infty} z \ln \frac{z}{z-1} = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \ln \frac{\frac{1}{\varepsilon}}{\frac{1}{\varepsilon}-1} = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \ln\left(\frac{1}{1-\varepsilon}\right)$$

$$= \lim_{\varepsilon\to 0} -\frac{1}{\varepsilon} \ln(1-\varepsilon) - \lim_{z\to\infty} g(z)$$

8) By L'Hopital's rule, $\lim_{\varepsilon\to 0} -\frac{1}{\varepsilon} \ln(1-\varepsilon)$

$$= \lim_{\varepsilon\to 0} - \frac{-\frac{1}{1-\varepsilon}}{1} = +1$$

9) $g'(z) < 0$ so g is decreasing for $z > 2, g(z) > 1$

☒

Lemma 3.3

Let $Y_1, Y_2, \ldots Y_p$ be a nondecreasing sequence of nonnegative numbers. Let $M_1(\cdot), M_2(\cdot), \ldots M_p(\cdot)$ be a sequence of positive differentiable functions satisfying $\Sigma_{r=1}^{p} M_r(x) \equiv 1$. Let $f(x) = \Sigma_{r=1}^{p} Y_r M_r(x)$.

If $x_1 > x_o$ and $\frac{d \frac{M_r(x)}{M_{r-1}(x)}}{dx} \leq 0 \qquad x_o \leq x \leq x_1 \quad r=2,3,\ldots p$

then $f(x_1) \leq f(x_o)$. The inequality is strict if

$Y_p > Y_1$ and $\dfrac{d\,\dfrac{M_2(x)}{dM_1(x)}}{dx} < 0 \qquad x_o \leq x \leq x_1$

Proof.

1) If $M_r(x_1) < M_r(x_o)$ for $r \leq p-1$, then $M_{r+1}(x_1) < M_{r+1}(x_o)$. If not, then $\dfrac{M_{r+1}(x_1)}{M_r(x_1)} > \dfrac{M_{r+1}(x_o)}{M_r(x_o)}$.

contradicting $\dfrac{d\,\dfrac{M_{r+1}(x)}{M_r(x)}}{dx} \leq 0 \qquad x_o \leq x \leq x_1$

2) $M_1(x_1) \geq M_1(x_o)$; If not $M_r(x_1) < M_r(x_o)$ for $r = 1,2,\ldots p$

$$\Sigma_{r=1}^{p} M_r(x_1) < \Sigma_{r=1}^{p} M_r(x_o) = 1 \text{ contradicts}$$

$$\Sigma_{r=1}^{p} M_r(x_1) = 1.$$

3) If $M_r(x_1) = M_r(x_o)$ $r = 1,2,\ldots p$ then lemma 3 holds trivially. Therefore we may assume $M_r(x_1) \neq M_r(x_o)$ for some r, and for some r $M_r(x_1) < M_r(x_o)$. Let j be the least integer such that $M_j(x_1) < M_j(x_o)$

4) $\Sigma_{r=1}^{p} [M_r(x_1) - M_r(x_o)] = 1-1 = 0$

5) $\Sigma_{r=1}^{j-1} M_r(x_1)-M_r(x_o) + \Sigma_{r=j}^{p} M_r(x_1)-M_r(x_o) = 0$

$$\Sigma_{r=1}^{j-1} M_r(x_1)-M_r(x_o) = \Sigma_{r=j}^{p} M_r(x_o)-M_r(x_1).$$

(all terms in both summands are positive)

6) $\Sigma_{r=1}^{j-1} Y_{j-1} \ [M_r(x_1)-M_r(x_o)]$

$$\leq \Sigma_{r=j}^{p} Y_j \ [M_r(x_o)-M_r(x_1)]$$

7) $\Sigma_{r=1}^{j-1} Y_r \ [M_r(x_1)-M_r(x_o)] \leq \Sigma_{r=j}^{p} Y_r \ [M_r(x_o)-M_r(x_1)]$

since for $r \leq j - 1$ $Y_r \leq Y_{j-1}$ and for $r \geq j$

$Y_r \geq Y_j$; rearrange to get

8) $\Sigma_{r=1}^{p} Y_r M_r(x_1) \leq \Sigma_{r=1}^{p} Y_r M_r(x_o)$.

9) If $\dfrac{d \frac{M_2(x)}{M_1(x)}}{dx} < 0 \quad x_o \leq x \leq x_1$

then $M_1(x_1) > M_1(x_o)$.

(If $M_1(x_1) = M_1(x_o)$ then since $\dfrac{M_2(x_1)}{M_1(x_1)} < \dfrac{M_2(x_o)}{M_1(x_o)}$,

$M_2(x_1) < M_2(x_o)$ and by (1) $M_r(x_1) < M_r(x_o)$

$r=2,3,...p$. This contradicts $1 = \Sigma_{r=1}^{p} M_r(x_1) =$
$\Sigma_{r=1}^{p} M_r(x_o)$)

10) If $\dfrac{d \frac{M_2(x)}{M_1(x)}}{dx} < 0 \quad x_o \leq x \leq x_1$ then

$M_p(x_1) < M_p(x_o)$. Therefore examining (7) we see

if $\dfrac{d \frac{M_2(x)}{M_1(x)}}{dx} < 0 \quad x_o \leq x \leq x_1$ and $Y_p > Y_1$

$\Sigma_{r=1}^{p} Y_r M_r(x_1) < \Sigma_{r=1}^{p} Y_r M_r(x_o)$

☒

PROPOSITION 3.4. (Chapter II, Proposition 4.1.)

1) $\sum_{k=1}^{mB_j(Y_i,j)} P(j,K) =$

$$\sum_{k=1}^{mB_j(Y_i,j)} \left[\left\{ \frac{1}{n} \frac{1}{k} \binom{mB_j(Y_i,j)-1}{k-1} \frac{1}{n}^{k-1} \left(\frac{n-1}{n}\right)^{mB_j(Y_i,j)-k} \right\} \cdot \left\{ \left(\frac{n-1}{n}\right)^{m[H_j(Y_i,j)-B_j(Y_i,j)]} \right\} \right]$$

$$= \left[\left\{ \sum_{k=1}^{mB_j(Y_i,j)} \binom{mB_j(Y_i,j)}{k} \frac{1}{n}^{k} \left(\frac{n-1}{n}\right)^{mB_j(Y_i,j)-k} \right\} \cdot \left\{ \frac{1}{mB_j(Y_i,j)} \mu^{m[H_j(Y_i,j)-B_j(Y_i,j)]} \right\} \right]$$

2) Binomial theorem states $\sum_{k=o}^{N} \binom{N}{k} P^k (1-P)^{N-k} =$

$$[(P+(1-P)]^N = 1$$

3) $\sum_{k=1}^{N} \binom{N}{k} P^k (1-P)^{N-k} = 1 - (1-P)^N$

4) $\sum_{k=1}^{mB_j(Y_i,j)} P(j,k) =$

$$\frac{1}{mB_j(Y_i,j)} [1-\mu^{mB_j(Y_i,j)}]\mu^{m[H_j(Y_i,j)-B_j(Y_i,j)]} =$$

$$\frac{1}{mB_j(Y_i,j)} \left[\mu^{m[H_j(Y_i,j)-B_j(Y_i,j)]} - \mu^{mH_j(Y_i,j)}\right]$$

$$(\mu = \frac{n-1}{n})$$

Propositions above are proved in the body of the text.

Theorem 3.5. (Chapter II, Theorem 4.5.)

1) Let $f(x,\theta) = \mu(x)^{xm\theta}$, then

2) $$P(x,i) = \Sigma_{j\varepsilon A_i} \frac{1}{mB_j(Y_{i,j})} [f(x,H_j(Y_{i,j})-B_j(Y_{i,j})) - f(x,H_j(Y_{i,j})].$$

3) If $A_i \neq \emptyset$, Then $\frac{\partial P}{\partial x}(x,i) > 0$

$$\text{if } \left.\frac{\partial^2 f}{\partial x \partial \theta}\right|_{\substack{(x,\theta) \\ \theta\varepsilon[H_j(Y_{i,j})-B_j(Y_{i,j}),\ H_j(Y_{i,j})]}} < 0$$

$$\text{and } \frac{\partial P}{\partial x}(x,i) < 0 \text{ if } \left.\frac{\partial^2 f}{\partial x \partial \theta}\right|_{\substack{(x,\theta) \\ \theta\varepsilon[H_j(Y_{i,j})-B_j(Y_{i,j}),\\ H_j(Y_{i,j})]}} > 0$$

4) $$\frac{\partial f}{\partial x} = \mu(x)^{xm\theta}\, m\theta[\ln(\frac{nx-1}{nx}) + \frac{1}{nx-1}]$$

$$\frac{\partial^2 f}{\partial x \partial \theta} = [\ln(\frac{nx-1}{nx}) + \frac{1}{nx-1}]m[\mu(x)^{xm\theta} + \theta\mu(x)^{xm\theta}xm \ln \mu(x)].$$

5) $$\frac{\partial^2 f}{\partial x \partial \theta} = [\ln(\frac{nx-1}{nx}) + \frac{1}{nx-1}]m\ \mu(x)^{xm\theta}\ [1 + \theta xm \ln \mu(x)]$$

6) By lemma 1, sign $\frac{\partial^2 f}{\partial x \partial \theta}$ = sign $[1 + \theta xm \ln\mu(x)]$.

7) Sign $\frac{\partial^2 f}{\partial x \partial \theta}$ = sign $[1 + \theta\alpha nx \ln(\frac{nx-1}{nx})]$

8) for $0 \leq \theta \leq 1$ ($\theta = H_j(Y_{i,j}) - B_j(Y_{i,j})$ or $\theta = H_i(Y_{i,j})$)

$$1 + \theta\alpha nx \ln(\frac{nx-1}{nx}) > 1 + \alpha nx \ln(\frac{nx-1}{nx});$$

9) $$1 + \alpha nx \ln(\frac{nx-1}{nx}) > 0 \text{ if } \alpha < -\frac{1}{nx \ln(\frac{nx-1}{nx})} =$$

$$\frac{1}{nx \ln(\frac{nx}{nx-1})}$$

10) By lemma 2, for $nx>2$, $nx \ln(\frac{nx}{nx-1}) < 1.4$, therefore

11) $$\frac{\partial^2 f}{\partial x \partial \theta}\Bigg|_{\substack{nx>2 \\ 0 \leq \theta \leq 1}} > 0 \qquad \text{if } \alpha < 5/7 < -\frac{1}{nx \ln(\frac{nx-1}{nx})} \text{ and}$$

$$\frac{\partial P(x,i)}{\partial x}\Bigg|_{\substack{nx>2 \\ A_i \neq \emptyset}} < 0 \qquad \text{if } \alpha < 5/7 < -\frac{1}{nx \ln(\frac{nx-1}{nx})}$$

(12) $$1 + \theta\alpha nx \ln \frac{nx-1}{nx} = 1 - \theta\alpha nx \ln \frac{nx}{nx-1} < 1 - \theta\alpha$$

(13) If $\frac{1}{\theta} < \alpha$, $1 + \theta\alpha nx \ln \frac{nx-1}{nx} < 0$, there if

(14) If $\sup_{J \in A_i} \{[H_j(Y_{i,j}) - B_j(Y_{i,j})]^{-1}\} < \alpha$,

sign $[1 + \theta\alpha nx \ln \frac{nx-1}{nx}] < 0$

($\theta \in [H_j(Y_{i,j}) - B_j(Y_{i,j}), H_j(Y_{i,j})]$ and

$\frac{\partial P}{\partial x}(x,i) > 0$.

☒

THEOREM 3.6. (Chapter II, Theorem 4.6)

1) $Q(x,j) = 1-\mu(x)^{xmH_j(x_j)}$

2) $\frac{\partial Q}{\partial x}(x,j) = -\mu(x)^{xmH_j(x_j)}\ mH_j(x_j)[\ln(\frac{nx-1}{nx}) + \frac{1}{nx-1}]$

By lemma 1, if $nx>1$, $\ln(\frac{nx-1}{nx}) + \frac{1}{nx-1} > 0$; so

3) If $nx>1$ and $H_j(x_j) \neq 0$ $\frac{\partial Q}{\partial x}(x,j) < 0$.

4) Let $f(x) = E(x,j) =$

$$= \Sigma_{i\varepsilon C_j} \frac{Y_{i,j}}{mB_j(Y_{i,j})} \frac{\left[\mu(x)^{x[mH_j(Y_{i,j})-mB_j(Y_{i,j})]} - \mu(x)^{xmH_j(Y_{i,j})}\right]}{1-\mu(x)^{xmH_j(x_j)}}$$

5) Partition C_j into k disjoint subsets $C_j^1, C_j^2, \ldots C_j^k$ such that:

a) $i\ \varepsilon\ C_j^s \quad t\ \varepsilon\ C_j^s \quad Y_{i,j} = Y_{t,j}$

b) $i\ \varepsilon\ C_j^s \quad t\ \varepsilon\ C_j^{s+1} \quad Y_{i,j} < Y_{t,j}$

6) Let i_s be chosen such that $i_s\ \varepsilon\ C_j^s$ $s=1,2,\ldots k$, then $\|C_j^s\| = mB_j(Y_{i_s,j})$

7) $E(x,j) = f(x) = \Sigma_{s=1}^{k} Y_{i_s,j}$

$$\frac{\left[\mu(x)^{x[mH_j(Y_{i_s,j})-mB_j(Y_{i_s,j})]} - \mu(x)^{xmH_j(Y_{i_s,j})}\right]}{1-\mu(x)^{xmH_j(x_j)}}$$

8) $$\sum_{\theta = mH_j(x_j)-mH_j(Y_{i_s},j)+1}^{mH_j(x_j)-mH_j(Y_{i_s},j)+mB_j(Y_{i_s},j)} \frac{\left[\mu(x)^{xmH_j(x_j)-\theta} - \mu(x)^{xmH_j(x_j)-\theta+1}\right]}{1-\mu(x)^{xmH_j(x_j)}} =$$

$$\frac{\mu(x)^{x[mH_j(Y_{i_s},j)-mB_j(Y_{i_s},j)]} - \mu(x)^{xmH_j(Y_{i_s},j)}}{(1-\mu(x)^{xmH_j(x_j)})}$$

9) $mH_j(x_j)-mH_j(Y_{i_s},j)+mB_j(Y_{i_s},j)=mH_j(x_j)-mH_j(Y_{i_{s+1}},j)$

$(s=1,2,\ldots k-1)$

$mH_j(x_j) = mH_j(Y_{i_1},j)$

$mH_j(Y_{i_k},j) = mB_j(Y_{i_k},j)$

10) $$E(x,j) = f(x) = \Sigma_{r=1}^{mH_j(x_j)}\left[Y_r \cdot \frac{\left[\mu(x)^{x[mH_j(x_j)-r} - \mu(x)^{x[mH_j(x_j)-r+1]}\right]}{1-\mu(x)^{xmH_j(x_j)}}\right]$$

where $Y_r = Y_{i_s,j}$

if $mH_j(x_j)-mH_j(Y_{i_s},j)+1 \le r \le mH_j(x_j)-mH_j(Y_{i_s},j)+mB_j(Y_{i_s},j)$

11) $$f(x) = \Sigma_{r=1}^{mH_j(x_j)} Y_r\ M_r(x)$$

where $$M_r(x) = \frac{\mu(x)^{x[mH_j(x_j)-r]} - \mu(x)^{x[mH_j(x_j)-r+1}}{1-\mu(x)^{xmH_j(x_j)}}$$

12) $$\frac{M_r(x)}{M_{r-1}(x)} = \frac{\mu(x)^{x[mH_j(x_j)-r]} - \mu(x)^{x[mH_j(x_j)-r+1]}}{\mu(x)^{x[mH_j(x_j)-r+1]} - \mu(x)^{x[mH_j(x_j)-r+2]}}$$

$$= \frac{1}{\mu(x)^x} \qquad [r=2,\ldots mH_j(x_j)]$$

13) $$\frac{d\,\frac{M_r(x)}{M_{r-1}(x)}}{dx} = -\frac{1}{\mu(x)^{2x}}\,\mu(x)^x[\ln(\frac{nx-1}{nx}) + \frac{1}{nx-1}]$$

$$\frac{d\,\frac{M_r(x)}{M_{r-1}(x)}}{dx} < 0 \quad \text{for } 2<nx \text{ and } r=2,3\ldots mH_j(x_j)$$

14) $\Sigma_{r=1}^{mH_j(x_j)} M_r(x) \equiv 1$, so apply lemma 3.3 to get

if $x_1 > x_o$ $(x_o > 2/n)$ then $f(x_1 \leq f(x_o)$ if $x_o > 2/n$

and then exist i and i' with $Y_{i,j} > Y_{i,j} \geq x_j$ then

$f(x_1) < f(x_o)$.

☒ Theorem 2

4. <u>Block Scalar Variance Covariance Matrix.</u>

PROPOSITION 4.1. Let $(\sigma^2_{1,1},\ldots,\sigma^2_{1,J}) \in \mathbb{R}^{+J}$, for observed data matrix X and dependent vector Y satisfying Assumptions I-IV:

1) Let $$\Sigma_1 = \begin{bmatrix} \sigma^2_{1,1}I_1 & 0 & 0 \\ 0 & \sigma^2_{1,2}I_2 & 0 \\ 0 & 0 & \sigma^2_{1,J}I_J \end{bmatrix}$$

2) Let $B_1 = (X^T\Sigma_1^{-1}X)^{-1} X^T\Sigma_1^{-1}Y$.

3) Let $\Sigma_{n+1} = \Sigma(B_n)$.

4) Let $B_{n+1} = (X^T\Sigma_{n+1}^{-1}X)^{-1} X^T\Sigma_{n+1}^{-1} Y$.

Then

A) For each $n \geq 1$ B_n exists;

B) The sequence $\{B_n\}$ has at least one limit point;

C) If B^* is a limit point of the sequence $\{B_n\}$ then $B^* = (X^T \Sigma(B^*)^{-1}X)^{-1} X^T\Sigma(B^*)^{-1}Y$.

PROOF. Our proof is similar to that done by Oberhofer and Kmenta.

1) Let $f:\mathbb{R}^k \to \mathbb{R}^{+J}$ be defined by

$$f(B, Z^2,\ldots,Z_j^2) = (2\pi)^{-N/2} [\Pi_{j=1}^J (Z_j^2)^{m_j}]^{-1/2} .$$

$$\exp - 1/2 \,\Sigma_{j=1}^J \frac{(Y_j-X_jB)^T (Y_j-X_jB)}{Z_j^2} .$$

$$(N = \Sigma_{j=1}^J m_j)$$

f is, of course, the likelihood function should the ε_{ij} be normally distributed.

2) The concentrated likelihood function is defined by

$$h(N) = (2\pi e)^{-N/2} \left\{ \Pi_{j=1}^J \left[\frac{(Y_j-X_jB) (Y_j-X_jB)}{m_j}\right]^{m_j} \right\}^{-1/2}$$

3) Since $\lim_{\|B\|\to\infty} \sup_{(Z^2,\ldots,Z_j^2)\varepsilon\mathbb{R}^{=J}} f(B,Z^2,\ldots,Z_j^2) =$

$\lim_{\|B\|\to\infty} h(B) = 0$. It follows that for any $\delta > 0$

$\{B \in \mathbb{R}^k \mid \exists \ (Z^2,\ldots,Z_j{}^2) \in \mathbb{R}^{+J} \ \ f(B, Z^2,\ldots,Z_j{}^2) > \delta$ is either bounded or empty and hence has compact closure.

4) It also follows from 3 that the function f is bounded.

5) It is immediate that for any $B \in \mathbb{R}^k$,

$f(B,Z_1(B),\ldots,Z_j(B)) \geq f(B,Z^2,\ldots,Z_j^2)$ for all $(Z^2,\ldots,Z_j^2) \in \mathbb{R}^{+J}$ and that $f(B,Z_1(B),\ldots,Z_j(B)) = f(B,Z^2,\ldots,Z_j^2) \Rightarrow Z_j(B) = Z_j^2$ all j.

6) It is also immediate that the unique B that maximizes $f(\_,Z_1{}^2,\ldots,Z_J{}^2)$ where $Z_1{}^2,\ldots,Z_j{}^2$ are fixed positive numbers is given by $B = (X^T\Sigma^{-1}X)^{-1}(X^T\Sigma^{-1}Y)$ where

$$\Sigma = \begin{bmatrix} Z_1^2 I_1 & 0 & \\ 0 & Z_2^2 I_2 & 0 \\ 0 & 0 & Z_j^2 I_j \end{bmatrix}$$

I_j is $m_j \times m_j$ identity matrix.

We show that the sequence $\{B_n\}$ exists by induction on n.

7) Let $\sigma_{1,1}{}^2,\ldots,\sigma_{1,J}{}^2$ be any J positive numbers.

The matrix $X^T\Sigma_1^{-1}X = \Sigma_{j=1}^J \dfrac{X_j{}^TX_j}{\sigma_{1,j}{}^2} = \Sigma_{j=1}^J \dfrac{m_j}{\sigma_{1,j}{}^2} \dfrac{X_j{}^TX_j}{m_j}$.

For each j, $\dfrac{X_j{}^TX_j}{m_j}$ is a positive definite symmetric matrix and $\dfrac{m_j}{\sigma_{1,j}{}^2}$ is a positive number. Therefore, $X^T\Sigma^{-1}X$ is a positive definite symmetric matrix and, in particular, is nonsingular. Therefore, B_1 which equals $(X^T\Sigma^{-1}X)^{-1}$ $X^T\Sigma^{-1}Y$ exists.

Assume $B_1, \ldots, B_n$ exist. By assumption $\inf_{B \in \mathbb{R}^k} \| Y_j - X_j B \| > 0$, so that for each j, $Z_j(B_n) > 0$.

In the above argument which shows B_1 exists, replace $\sigma_{1,j}{}^2$ by $Z_j(B_n)$ to see that B_{n+1} exists.

[X] A.

8) It follows from 5) and 6) that

$f(B_n, Z_1(B_n), \ldots, Z_j(B_n)) \leq f(B_{n+1}, Z_1(B_n), \ldots, Z_j(B_n)) \leq$ $f(B_{n+1}, Z_1(B_{n+1}), \ldots, Z_j(B_{n+1}))$ hence from 3) $\{B_n\}$ is bounded and therefore has a convergent subsequence.

[X] B.

9) Let B^* be one limit point of $\{B_n\}$ and let

$$B_{n_k} \to B^* \quad .$$

$$f(B_{n_k}, Z_1(B_{n_k}), \ldots, Z_J(B_{n_k})) \leq f(B_{n_k+1}, Z_1(B_{n_k}), \ldots, Z_J(B_{n_k}))$$

$$f(B_{n_{k+1}}, Z_1(B_{n_{k+1}}), \ldots, Z_J(B_{n_{k+1}})) \leq$$

$$\leq f(B_{n_{k+1}}, Z_1(B_{n_{k+1}}), \ldots, Z_J(B_{n_{k+1}})) \quad .$$

Since:

i) f is a bounded function.

ii) $B_{n_k+1} = (X^T \Sigma(B_{n_k})^{-1} X)^{-1} X^T \Sigma(B_{n_k})^{-1} Y$ so

B_{n_k+1} converges to $(X^T \Sigma(B^*)^{-1} X^T \Sigma(B^*)^{-1} Y$

which we denote by B^*.

Letting k increase to ∞ and using the fact that the function f and Z_j are continuous, we have

$f(B^*, Z_1(B^*), \ldots, Z_J(B^*)) \leq f(B^*, Z_1(B^*), \ldots, Z_J(B^*))$ B^* is the unique element of $\mathbb{R}^k$ that maximizes $f(\_, Z_1(B^*), \ldots, Z_J(B^*))$, therefore $\overline{B}^* = B^*$.

[X] C.

PROPOSITION 4.2. Let $g: \mathbb{R}^k \to \mathbb{R}^k$ be defined by $g(B) = (X^T\Sigma(B)^{-1}X)^{-1}X^T\Sigma(B)^{-1}Y$ $[=(\frac{1}{N}X^T\Sigma(B)X)^{-1}(\frac{1}{N}X^T\Sigma(B)^{-1}Y)]$. Let F be the set of fixed points of g, and let $d = \text{Sup}\{\|B-B_0\| : B \in F\}$. Let $\gamma_0 > 0$, then under Assumptions I-V, $d < \gamma_0$ almost surely as $N \uparrow \infty$.

Proof. It should be understood that X, Y, and N all appear explicitly in the definition of g, and that, therefore, for each N, d is a random variable. It should be remembered that $N = \Sigma_{j=1}^{J} m_j$ and that, in effect, $m_j = m_j(N)$.

For a fixed pair (X,Y) let $\hat{B}$ be a fixed pointof g. Recall that X is an $N \times K$ real matrix and Y is an $N \times 1$ vector. Where no confusion is likely to arise, we drop the superscript. Thus:

1) $$g(\hat{B}) = (\frac{1}{N}X^T\Sigma(B)^{-1}X)^{-1}(\frac{1}{N}X^T\Sigma(B)^{-1}Y) = \hat{B} \quad .$$

Recall that:

2) $$Y = XB_0 + \varepsilon$$

and rearrange 1) to get:

3) $$\frac{1}{N}[(X^T\Sigma(\hat{B})^{-1}X)(B_0-\hat{B}) + X^T\Sigma(\hat{B})^{-1}\varepsilon] = 0 \quad .$$

Now premultiply both sides of 3) by $(B_0-\hat{B})^T$ and expand to get:

4) $$\Sigma_{j=1}^{J} \frac{m_j}{N} \left[\frac{(B_o-\hat{B})^T \frac{X_j^T X_j}{m_j} (B_o-\hat{B}) + (B_o-\hat{B})^T \frac{X_j^T \varepsilon_j}{m_j}}{Z_j(\hat{B})} \right] = 0$$

Let $S_1 = \{j \mid m_j(N)$ is unbounded as $N \uparrow \infty\}$ and let $S_2 = \{j \mid m_j(N)$ is bounded $\}$. S_1 cannot be empty S_2, however, may be empty. By convention, we take $\sum_{j \varepsilon \phi} X_j = 0$ where ϕ is the empty set. We therefore get from equation 4):

5) $$\Sigma_{j \varepsilon S_1} \frac{m_j}{N} \left[\frac{(B_o-\hat{B})^T \frac{X_j^T X_j}{m_j} (B_o-\hat{B}) + (B_o-\hat{B})^T \frac{X_j^T \varepsilon_j}{m_j}}{Z_j(\hat{B})} \right] =$$

$$-\sum_{j \varepsilon S_2} \frac{m_j}{N} \left[\frac{(B_o-\hat{B})^T \frac{X_j^T X_j}{m_j} (B_o-\hat{B}) + (B_o-\hat{B}) \frac{X_j^T \varepsilon_j}{m_j}}{Z_j(\hat{B})} \right]$$

We now analyze the RHS of equation 5) and examine each term in the summand to get:

6) $$\frac{m_j}{n} \left[\frac{[(B_o-\hat{B})^T \frac{X_j^T X_j}{m_j} (B_o-\hat{B}) + (B_o-\hat{B})^T \frac{X_j^T \varepsilon_j}{m_j}]}{Z_j(\hat{B})} \right] \geq$$

$$\frac{m_j}{N}\left[\frac{[\lambda\|B_0-\hat{B}\|^2 - |\frac{X_j^T\varepsilon_j}{m_j}|\ \|B_0-\hat{B}\|\,]}{Z_j(\hat{B})}\right]$$

$$\frac{m_j}{N}\left[\frac{-|\frac{X_j^T\varepsilon_j}{m_j}|^2}{4\ \lambda\ Z_j(\hat{B})}\right]$$

(From elementary calculus if $a>0$ $b>0$ then $ax^2-bx \geq -\frac{b^2}{4a}$)

For $j \in S_2$, and for N sufficiently large, $m_j(N)$, X_j and ε_j are independent of N. Hence, $\inf_{B\in\mathbb{R}^K} Z_j(B)$ is also independent of N for N sufficiently large, and so $Z_j(\hat{B}^N)$ is bounded away from zero.

For $j \in S_2$, $m_j(N)$ is bounded, therefore, for $j\in S_2$ $\lim_{N\to\infty} \frac{m_j(N)}{N} = 0$.

Let $\gamma > 0$, it follows immediately from the above that: that:

$$7)\quad \Sigma_{j\in S_2} \frac{m_j(N)}{N}\left[\frac{(B_0-\hat{B}^N)^T\frac{X_j^TX_j}{m_j(N)}(B_0-\hat{B}^N) + (B_0-\hat{B}^N)^T\frac{X_j^T\varepsilon_j}{m_j(N)}}{z_j(\hat{B}^N)}\right] > -\gamma$$

almost surely as $N \uparrow \infty$

By 5) and 7) we now have that for arbitrary $\gamma > 0$

$$8)\quad \Sigma_{j\in S_1} \frac{m_j(N)}{N}\ \frac{(B_0-\hat{B}^N)^T\frac{X_j^TX_j}{m_j(N)}(B_0-\hat{B}^N)^T + (B_0-\hat{B}^N)^T\frac{X_j^T\varepsilon_j}{m_j(N)}}{z_j(B^N)} < \gamma$$

almost surely as $N \to \infty$.

Before proceeding with our proof of consistency, we need to establish two elementary numerical lemmas.

9) **Lemma** 4.3. Let a, b, c, be positive real numbers, then on $\{x \mid x \geq 0\}$, the function $f(x) = \frac{ax}{bx+c}$ is an increasing function.

(Proof. $f'(x) = \frac{ac}{(bx+c)^2} > 0$)

10) **Lemma** 4.4. Let a,b,c,d, be positive real numbers. Let $f(x) = \frac{ax^2-bx}{cx^2+2bx+d}$. Then for $x \geq \frac{2b}{a}$; $f(x) \geq \frac{2ab^2}{4(c+a)b^2+a^2d}$

[Proof. $x \geq \frac{2b}{a} \to \frac{ax^2-bx}{cx^2+2bx+d} \geq \frac{\frac{a}{2}x^2}{(c+a)x^2+d}$

Now use lemma 4.3 to get:

$$x \geq \frac{2b}{a} \to \frac{ax^2-bx}{cx^2+2bx+d} \geq \frac{2\frac{b^2}{a}}{(c+a)\frac{4b^2}{a^2}+d} = \frac{2ab^2}{4(c+a)b^2+a^2d}$$]

We now analyse the LHS of equation 5) For $||B_o-\hat{B}^N|| \geq \dfrac{2\,||\frac{X_j^T \varepsilon_j}{m_j}||}{\lambda}$

11) $$\frac{(B_o-\hat{B}^N)\ \frac{X_j^T X_j}{m_j}\ (B_o-\hat{B}^N) + (B_o-\hat{B}^N)^T\ \frac{X_j^T \varepsilon_j}{m_j}}{Z_j(\hat{B}^N)} \geq$$

$$\frac{\lambda||B_o-\hat{B}^N||^2 - ||B_o-\hat{B}^N||\ ||\frac{X_j^T \varepsilon_j}{m_j}||}{T||B_o-\hat{B}^N||^2 + 2||B_o-\hat{B}^N||\ ||\frac{X_j^T \varepsilon_j}{m_j}|| + \frac{\varepsilon_j^T \varepsilon_j}{m_j}}$$

(Recall that $Z_j(\hat{B}^N) = (B_o-\hat{B}^N)^T\ \frac{X_j^T X_j}{m_j}\ (B_o-\hat{B}^N)\ +$
$+\ 2(B_o-\hat{B}^N)\ \frac{X_j^T \varepsilon_j}{m_j} + \frac{\varepsilon_j^T \varepsilon_j}{m_j}$ use 10) to get:

For $||B_o-\hat{B}^N|| \geq \frac{2}{\lambda}\ ||\frac{X_j^T \varepsilon_j}{m_j}||$

12) $$\frac{(B_o-\hat{B}^N)^T\ \frac{X_j^T X_j}{m_j}\ (B_o-\hat{B}^N) + (B_o-\hat{B}^N)^T\ \frac{X_j^T \varepsilon_j}{m_j}}{Z_j(\hat{B}^N)} \geq$$

$$\frac{2\lambda\ ||\frac{X_j^T \varepsilon_j}{m_j}||^2}{4(T+\lambda)\ ||\frac{X_j^T \varepsilon_j}{m_j}||^2 + \lambda^2\ \frac{\varepsilon_j^T \varepsilon_j}{m_j}}$$

If $\left|\left|\frac{X_j^T \varepsilon_j}{m_j}\right|\right| < \frac{\delta\lambda}{2}$, and if $\frac{\varepsilon_j^T \varepsilon_j}{m_j} < V$ then using 10), 12)

becomes:

For $||B_0-\hat{B}^N|| > \delta$

13)

$$\frac{(B_0-\hat{B}^N)^T \frac{X_j^T X_j}{m_j} (B_0-\hat{B}^N) + (B_0-\hat{B}^N)^T \frac{X_j^T \varepsilon_j}{m_j}}{Z_j(\hat{B}^N)} \qquad \frac{\lambda\delta^2}{8(T+\lambda)\ \delta^2 + 2V}$$

For all N, $\sum_{j=1}^{T} \frac{m_j(N)}{N} \equiv 1$, therefore for N sufficiently large $\sum_{j\varepsilon S_1} \frac{m_j(N)}{N} > 1/2$. Let $V = 1 + \sum_{j=1} \sigma_j^2$, then it follows $V > \sigma_j^2$ all j. In equation 8) above, let

$\gamma = 1/2 \dfrac{\lambda\delta^2}{8(T+\lambda)\delta^2+2V}$.

By the Strong Law of Large Numbers we have

$$\max_{j\varepsilon S_1} \left|\left|\frac{X_j^T \varepsilon_j}{m_j(N)}\right|\right| < \frac{\delta\lambda}{2} \tag{14}$$

almost surely $N \uparrow \infty$ and

$$\max_{j\varepsilon S_1} \left|\left|\frac{\varepsilon_j^T \varepsilon_j}{M_j(N)}\right|\right| < V \tag{15}$$

almost surely $N \uparrow \infty$.

8) gives us that

$$\Sigma_{j\varepsilon S_1} \frac{m_j(N)}{N} \quad \frac{(B_o-\hat{B}^N) \frac{X_j^T X_j}{M_j(N)} (B_o-\hat{B}^N) + (B_o-\hat{B}^N)^T \frac{X_j^T \varepsilon_j}{M_j(N)}}{Z_j(\hat{B}^N)} <$$

$$\frac{1/2\ \lambda\delta^2}{8(T+\lambda)\delta^2 + 2V}$$

almost surely $N \uparrow \infty$.

Now observe that if:

$$\Sigma_{j\varepsilon S_1} \frac{M_j(N)}{N} > 1/2, \quad \max_{j\varepsilon S_1} \left|\left|\frac{X_j^T \varepsilon_j}{M_j(N)}\right|\right| < \frac{\delta\lambda}{2}, \quad \max_{j\varepsilon S_1} \left|\left|\frac{\varepsilon_j^T \varepsilon_j}{M_j(N)}\right|\right| < V$$

and $$\Sigma_{j\varepsilon S_1} \frac{M_j(N)}{N} \quad \frac{(B_o-\hat{B}^N)^T \frac{X_j^T X_j}{M_j(N)} (B_o-\hat{B}^N) + (B_o-\hat{B}^N)^T \frac{X_j^T \varepsilon_j}{M_j(N)}}{Z_j(\hat{B}^N)} <$$

$$\frac{1/2\ \lambda\delta^2}{8(T+\lambda)\delta^2 + 2V}$$

Then $||B_o-\hat{B}^N|| < \delta$.

We can now conclude that for $\delta > 0$ $||B_o-\hat{B}^N|| < \delta$ almost surely as $N \uparrow \infty$.

COROLLARY 4.5. Let Θ be an estimator with the property that for each N, there exists positive number $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$ (perhaps depending upon N) such that for each pair (X,Y), $\Theta(X,Y)$ is a limit point of $\{B_n\ \sigma_{1,1}^2,\ldots,\sigma_{1,J}^2,X,Y)\}$. Then under Assumptions I-V, Θ is strongly consistent.

PROPOSITION 4.6. Let Θ be an estiator with the property for each N, there exists positive numbers $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$ such that for each pair (X,Y), $\Theta(X,Y)$ is a limit point of $\{B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2, X,Y)\}$. Then under Assumptions I-V, the sequence $\{B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2, X,Y)\}$ is convergent almost surely as $N \uparrow \infty$.

Proof. Consider the mapping $g\colon \mathbb{R}^k \to \mathbb{R}^k$ defined by:

$$1) \qquad g(B) = (X^T\Sigma(B)^{-1}X)^{-1}\, X^T\Sigma(B)^{-1}Y \quad .$$

It follows immediately from the definition of g and the proof of Proposition 4.2. that $B_{n+1} = g(B_n)$ and that $\Theta(X,Y)$ is a fixed point of g. As in the proof of Proposition 4.2, let $\hat{B} = \Theta(X,Y)$. Since there exists a subsequence $\hat{B}_{n_k}$ converging to B, the original sequence $\{B_n(\sigma_{1,1}{}^2,\ldots,\sigma_{1,J}{}^2, X,Y)\}$ will converge to $\hat{B}$ if g is a contractionmapping near $\hat{B}$. Therefore to show convergence of $\{B_n\}$ we need only show that $\| B - \hat{B}\| < 1/2$ implies that

2) $$\| g(B) - g(\hat{B})\| \leq 1/2 \| B - \hat{B}\| \quad .$$

[Here as elsewhere in this proof, all norms refer to the operate norm. To see that this suffices, observe that $B_{n_k} \to \hat{B}$ implies that there exists p such that $\| B_{n_p} - B\| < 1/2$. Since $\hat{B} = g(\hat{B}) = g(g(\hat{B})) : = g^2(\hat{B}) = g^r(\hat{B})$, and since $G_{n_p+r} = g^r(B_{n_p})$, we have that:

3) $$\| B_{n_p+r} - \hat{B}\| = \| g^r(B_{n_p}) - g^r(\hat{B})\|$$
$$\leq (\tfrac{1}{2})^2 \| B_{n_p} - \hat{B}\| \leq (\tfrac{1}{2})^{r+1}$$

4) $$g(B) - g(\hat{B}) = (\frac{1}{N} X^T\Sigma(B)^{-1}X)^{-1} (\frac{1}{N} X^T\Sigma(B)^{-1}Y)$$
$$- (\frac{1}{N} X^T\Sigma(\hat{B})^{-1}X)^{-1} (\frac{1}{N} X^T\Sigma(\hat{B})^{-1}Y) \quad .$$

Use the following facts:

5) $$Y = XB_0 + \varepsilon$$

6) $$\| AB - CD\| \leq \| A\| \ \| B - D\| + \| A - C\| \ \| D\|$$

7) $\|A^{-1} - B^{-1}\| \le \|A^{-1}\| \; \|B^{-1}\| \; \|A - B\|$

to get from 4):

8)

$$\|g(B) - g(\hat{B})\| \le \|(\tfrac{1}{N} X^T\Sigma(B)^{-1}X^{-1})\| \; \|\tfrac{1}{N} X^T\Sigma(B)^{-1}\varepsilon - \tfrac{1}{N} X^T\Sigma(B)^{-1}\varepsilon\| + \|\tfrac{1}{N} X^T\Sigma(\hat{B})^{-1}\varepsilon\| \; \|(\tfrac{1}{N} X^T\Sigma(B)^{-1}X)^{-1}\| \; \|(\tfrac{1}{N} X^T\Sigma(\hat{B})^{-1}X)^{-1}\| \; \|(\tfrac{1}{N} X^T\Sigma(B)^{-1}X - \tfrac{1}{N} X^T\Sigma(\hat{B})^{-1}X\| .$$

We wish to show that for $\|B - \hat{B}\| < 1/2$, $\|g(B) - g(\hat{B})\| < 1/2 \; \|B - \hat{B}\|$ almost surely as $N \uparrow \infty$. We can confine our analysis to the unit ball about B_0, since we have shown $\|\hat{B} - B_0\| < 1/2$ almost surely as $N \uparrow \infty$.

Let $S_1 = \{j \,|\, M_j(N)$ is unbounded as $N \uparrow \infty\}$.

Let $S_2 = \{j \,|\, M_j(N)$ is bounded as $N \uparrow \infty\}$.

$$Z_j(B) = (B_0-B)^T \frac{X_j^T X_j}{m_j}(B_0-B) + 2(B_0-B)^T \frac{X_j^T \varepsilon_j}{m_j} + \frac{\varepsilon_j^T \varepsilon_j}{m_j} ,$$

therefore for $\|B_0 - B\| \le 1$, we have

9)

$$\frac{\varepsilon_j^T \varepsilon_j}{m_j} - \frac{\dfrac{\|X_j^T \varepsilon_j\|^2}{m_j}}{\lambda} \le Z_j(B) \le T + 2 \left|\frac{X_j^T \varepsilon_j}{m_j}\right| + \frac{\varepsilon_j^T \varepsilon_j}{m_j} .$$

Let $m = \frac{1}{2} \min_j (\sigma_j^2)$ and let $M = T + 2 + \Sigma_{j=1}^J \sigma_j^2$; then $0 < m \le M < \infty$ and for $j \in S_1$ and $\|B-B_0\| \le 1$, we have:

10) $\quad m \leq Z_j(B) \leq M \quad$ almost surely as $N \uparrow \infty$.

For B such that $\| B - B_0 \| \leq 1$, We wish to find $\| (\frac{1}{N} X^T \Sigma(B)^{-1} X)^{-1} \|$. Since $\frac{1}{N} X^T \Sigma(B)^{-1} X$ is positive definite symmetric, to find a upper bound on the norm of its inverse, we need only take an inverse of a positive lower bound of its minimum eigenvalue. The minimum eigenvalue of $(\frac{1}{N} X^T \Sigma(B)^{-1} X)$ is given by

$$11) \qquad \inf_{\substack{Z \varepsilon \mathbb{R}^k \\ \| Z \| = 1}} Z^T (\frac{1}{N} X^T \Sigma(B)^{-1} X) Z \quad .$$

For $Z \varepsilon \mathbb{R}^k$, $\| Z \| = 1$,

$$12) \quad Z^T (\frac{1}{N} X^T \Sigma(B)^{-1} X) Z = \sum_{j=1} \frac{m_j}{N} \left[\frac{\frac{Z^T X_j^T X_j Z}{m_j}}{Z_j(B)} \right] \geq \sum_{j \varepsilon S_1} \frac{m_j}{N} \frac{\lambda}{Z_j(B)}$$

Therefore by 10) and for $\| B - B_0 \| \leq 1$, the minimum eigenvalue of $\frac{1}{N}(X^T \Sigma(B)^{-1} X) \leq \frac{\lambda}{2M}$ almost surely as $N \uparrow \infty$. Hence, for $\| B - B_0 \| \leq 1$, $\| (\frac{1}{N} X^T \Sigma(B)^{-1} X)^{-1} \| \leq \frac{2M}{\lambda}$ almost surely as $N \uparrow \infty$. In particular, since $\| \hat{B} - B_0 \| < 1/2$ almost surely as $N \uparrow \infty$, we have:

13) $$\left\| \frac{1}{N} \left(X^T \Sigma(\hat{B})^{-1} X\right)^{-1} \right\| \qquad \frac{2M}{\lambda}$$

almost surely as $N \uparrow \infty$.

14) $$\frac{1}{N} X^T \Sigma(B)^{-1} \varepsilon - \frac{1}{N} X^T \Sigma(\hat{B})^{-1} \varepsilon = \sum_{j \varepsilon S} 2 \frac{m_j}{N} \frac{1}{Z_j(B)} \frac{1}{Z_j(B)} \frac{X_{j \varepsilon j}^T}{m_j} (Z_j(\hat{B})$$

$$- Z_j(B)) + \sum_{j \varepsilon S_2} \frac{m_j}{N} \frac{1}{Z_j(B)} \frac{1}{Z_j(B)} \frac{X_j^T \varepsilon_j}{m_j} (Z_j(\hat{B}) - Z_j(B)).$$

Now

$$Z_j(\hat{B}) - Z_j(B) = (B_0 - \hat{B})^T \frac{X_j^T X_j}{m_j} (B_0 - \hat{B}) + 2(B_0 - B)^T \frac{X_j^T \varepsilon_j}{m_j}$$

$$- (B_0 - B)^T \frac{X_j^T X_j}{m_j} (B_0 - B) - 2(B_0 - B)^T \frac{X_j^T \varepsilon_j}{m_j},$$

15)

$$|Z_j(\hat{B}) - Z_j(B)| \leq T\|B_0 - \hat{B}\| \ \|\hat{B} - B\| + T\|B_0 - B\| \ \|\hat{B} - B\|$$

$$+ 2\left\|\frac{X_j^T \varepsilon_j}{m_j}\right\| \|\hat{B} - B\| \leq \left[T \|B_0 - \hat{B}\| + T\|B_0 - B\|\right.$$

$$\left. + 2 \ \frac{X_j^T \varepsilon_j}{m_j}\right\|\Bigg] \|\hat{B} - B\| \quad .$$

Let $\gamma_1 > 0$, then for $j \varepsilon S_1$ and $\|B - B\| \leq 1$, it follows that:

16) $$\left\| \frac{1}{Z_j(B)} \frac{1}{Z_j(B)} \frac{X_j^T \varepsilon_j}{m_j} (Z_j(\hat{B}) - Z_j(B)) \right\| \leq \frac{\gamma_1}{2} \|B - B\|$$

almost surely as $N \uparrow \infty$.

For $j \varepsilon S_2$, X_j, $m_j(N)$, ε_j are all eventually independent of N. By assumption $Z_j(B) > 0$ all

$B \varepsilon \mathbb{R}^k$ and for $j \varepsilon S_2$, $\lim_{N \uparrow \infty} \frac{m_j(N)}{N} = 0$, therefore we have for $\| B - B_0 \| \leq 1$:

17) $$\Sigma_{j \varepsilon S_2} \| \frac{m_j}{N} \frac{1}{Z_j(B)} \frac{1}{Z_j(\hat{B})} \frac{X_j^T \varepsilon_j}{m_j} (Z_j(\hat{B}) - Z_j(B)) \|$$

$$\frac{\gamma_1}{2} \| B - \hat{B} \|$$ almost surely as $N \uparrow \infty$.

Combining 15) and 16), we get:

18) $$\sup_{\substack{B \varepsilon \mathbb{R}^k \\ 0 < \| B - B_0 \| \leq 1}} \left\{ \frac{\| \frac{1}{N} X^T \Sigma(B)^{-1} \varepsilon - \frac{1}{N} X^T \Sigma(B)^{-1} \|}{\| B - B_0 \|} \right\} \leq \gamma_1$$

almost surely as $N \uparrow \infty$.

$$\frac{1}{N} X^T \Sigma(B)^{-1} \varepsilon = \sum_{j \varepsilon S_1} \frac{m_j}{N} \frac{\frac{X_j^T \varepsilon_j}{m_j}}{Z_j(B)} + \sum_{j \varepsilon S_2} \frac{m_j}{N} \frac{\frac{X_j^T \varepsilon_j}{m_j}}{Z_j(B)}$$ and

therefore for $\| B - B_0 \| \leq 1$ and $\gamma_2 > 0$

$\| \frac{1}{N} X^T \Sigma(\hat{B})^{-1} \varepsilon \| \leq \gamma_2$ almost surely as $N \uparrow \infty$. Hence,

19) $$\| \frac{1}{N} X^T \Sigma(B)^{-1} \varepsilon \| \leq \gamma_2$$ almost surely as $N \uparrow \infty$.

20) $$\frac{1}{N} X^T \Sigma(B)^{-1} X - \frac{1}{N} X^T \Sigma(\hat{B})^{-1} X = \sum_{j \varepsilon S_1} \frac{m_j}{N} \frac{X_j^T X_j}{m_j} \frac{1}{Z_j(B)} \frac{1}{Z_j(\hat{B})}$$

$$(Z_j(\hat{B}) - Z_j(B)) + \sum_{j \varepsilon S_2} \frac{m_j}{N} \frac{X_j^T X_j}{m_j} \frac{1}{Z_j(B)} \frac{1}{Z_j(\hat{B})} (Z_j(\hat{B}) - Z_j(B))$$

By 15 and 20) and for $\| B - B_0 \| \leq 1$ we get

21) $\| \frac{1}{N} X^T \Sigma(B)^{-1} X - \frac{1}{N} X^T \Sigma(\hat{B})^{-1} X \| \leq \frac{T}{m^2} [2T+3] \| B - \hat{B} \|$

almost surely as $N \uparrow \infty$.

Combining 8), 13), 18), 19), 21) and by the proper choices of γ_1 and γ_2 we get:

22) $$\sup_{\substack{B \in \mathbb{R}^k \\ 0 < \| B - \hat{B} \| < 1/2}} \frac{\| g(B) - g(\hat{B}) \|}{\| B - \hat{B} \|} \leq 1/2$$

almost surely as $N \uparrow \infty$.

[X]

The next result is a minor improving of the last proposition. It will however pave the way for showing that almost surely as N gets large, our estimator is independent of the $(\sigma_{1,1}^2, \ldots, \sigma_{1,J}^2)$ selected.

PROPOSITION 4.7. Let $g: \mathbb{R}^k \to \mathbb{R}^k$ be defined by $g(B) = (\frac{1}{N} X^T \Sigma(B)^{-1} X)^{-1} (\frac{1}{N} X^T \Sigma(B)^{-1} Y)$.

Let $$K(N) := \sup_{\substack{B_1 \in \mathbb{R}^k \\ B_2 \in \mathbb{R}^k \\ \| B_1 - B_0 \| \leq 1 \\ \| B_2 - B_0 \| \leq 1 \\ 0 < \| B_1 - B_2 \|}} \frac{\| g(B_1) - g(B_2) \|}{\| B_1 - B_2 \|}.$$

Then $K(N) < 1/2$ almost surely as $N \uparrow \infty$.

Proof. As in the previous proposition all norms refer to the operator norm. We use the same notation as in the proof of the last proposition.

$$1)\ \| g(B_1)-g(B_2)\| \le \| (\tfrac{1}{N} X^T\Sigma(B_2)^{-1}X)^{-1}\| \ \| \tfrac{1}{N}X^T\Sigma(B_2)^{-1}\varepsilon - \tfrac{1}{N} X^T\Sigma(B_1)^{-1}\varepsilon\| + \| \tfrac{1}{N} X^T\Sigma(B_1)^{-1}\varepsilon\| \ \| (\tfrac{1}{N}X^T\Sigma(B_2)^{-1}X)^{-1}\| \ \| \tfrac{1}{N} X^T\Sigma(B_2)^{-1}X - \tfrac{1}{N} X^T\Sigma(B_1)^{-1}X\| \ \| (\tfrac{1}{N} X^T\Sigma(B_1)^{-1}X)^{-1}\| .$$

From the proof of the last proposition, we have already shown that:

$$2)\qquad \sup_{\| B - B_0\| \le 1} \| (\tfrac{1}{N} X^T\Sigma(B)^{-1}X)^{-1}\| \le \frac{2M}{\gamma}$$

almost surely as $N \uparrow \infty$.

Here as elsewhere in this proof, let B_1, B_2 be elements of the unit ball about B_0.

$$Z_j(B_1) - Z_j(B_2) = (B_0 - B_1)^T \frac{X_j^T X_j}{m_j} (B_0 - B_1) + 2(B_0 - B_1)^T \frac{X_j^T \varepsilon_j}{m_j} - (B_0 - B_2)^T \frac{X_j^T X_j}{m_j} (B_0 - B_2) - 2(B_0 - B_2)^T \frac{X_j^T \varepsilon_j}{m_j}$$

and so:

3) $|Z_j(B_1) - Z_j(B_2)| \leq 2\| B_2 - B_1 \| \ T +$

$$+ 2 \| \frac{X_j{}^T \varepsilon_j}{m_j} \| \ \| B_2 - B_1 \| \leq [2T + 2 \| \frac{X_j{}^T \varepsilon_j}{m_j}] \| B_2 - B_1 \| .$$

Let $\gamma_1 > 0$, then for $j \in S_1$, it follows that

4) $$\| \frac{1}{Z_j(B_1)} \frac{1}{Z_j(B_2)} \frac{X_j{}^T \varepsilon_j}{m_j} (Z_j(B_1) - Z_j(B_2)) \| \leq \gamma_1 \| B_1 - B_2 \|$$

almost surely as $N \uparrow \infty$.

For $j \in S_2$, it immediately follows from the facts that $m_j(N)$, X_j, and ε_j are all eventually independent of N that:

5) $$\sum_{j \, S_2} \| \frac{m_j}{N} \frac{1}{Z_j(B_1)} \frac{1}{Z_j(B_2)} \frac{X_j{}^T \varepsilon_j}{m_j} (Z_j(B_1) - Z_j(B_2)) \|$$

$$\leq \frac{\gamma_1}{2} \| B_1 - B_2 \|$$

almost surely as $N \uparrow \infty$.

$$\frac{1}{N} X^T \Sigma(B_2)^{-1} \varepsilon - \frac{1}{N} X^T \Sigma(B_1)^{-1} \varepsilon =$$

$$= \sum_{j=1}^{J} \frac{m_j}{N} \frac{1}{Z_j(B_2)} \frac{1}{Z_j(B_1)} \frac{X_j{}^T \varepsilon_j}{m_j} (Z_j(B_1) - Z_j(B_2))$$

and therefore:

6) $$\sup_{\substack{||B_1-B_o||\leq 1\\ ||B_2-B_o||\leq 1\\ B_1\neq B_2}} \left\{ \frac{||\frac{1}{N} X^T\Sigma(B_1)^{-1}\varepsilon - \frac{1}{N} X^T\Sigma(B_2)^{-1}\varepsilon||}{||B_1 - B_2||} \right\} < \gamma_1$$

almost surely as $N \uparrow \infty$.

Let $\gamma_2 > 0$, we showed in the proof of the last prop that:

7) $$\sup_{||B-B_o||\leq 1} ||\tfrac{1}{N} X^T\Sigma(B)^{-1}\varepsilon|| < \gamma_2 \quad \text{almost surely as } N \uparrow \infty.$$

$$\frac{1}{N} X^T\Sigma(B_2)^{-1}X - \frac{1}{N} X^T\Sigma(B_1)^{-1}X = \Sigma_{j\varepsilon S_2} \frac{m_j}{N} \frac{X_j^T X_j}{m_j} \frac{1}{Z_j(B_2)}\frac{1}{Z_j(B_1)}$$

$$(Z_j(B_1)-Z_j(B_2)) + \sum_{j\varepsilon S_2} \frac{m_j}{N} \frac{X_j^T X_j}{m_j} \frac{1}{Z_j(B_2)} \frac{1}{Z_j(B_1)} (Z_j(B_1) - Z_j(B_2))$$

and so by 3) we get:

8) $$\sup_{\substack{||B_1 - B_o|| \leq 1\\ ||B_2 - B_o|| \leq 1\\ B_1\neq B_2}} \left\{ \frac{|| \frac{1}{N} X^T\Sigma(B)^{-1}X - \frac{1}{n} X^T\Sigma(B_1)^{-1}X||}{|| B_2 - B_1||} \right\} \leq 2T + 3$$

Now by the appropriate choices of γ_1 and γ_2 we get our desired result that:

9) $$\sup_{\substack{||B_1-B_o|| \leq 1\\ ||B_2-B_o|| \leq 1\\ B_1\neq B_2}} \frac{||g(B_1) - g(B_2)||}{||B_1 - B_2||} < \frac{1}{2} \quad \text{almost surely as } N \uparrow \infty.$$

Let $g:\mathbb{R}^k \to \mathbb{R}^k$ be defined as before by $g(B) = (X^T\Sigma(B)^{-1}X)^{-1}(X^T\Sigma(B)^{-1}Y)$. Suppose $g(B) = B$ implies $\| B - B_0\| < 1$ and that $\| B_1 - B_0\| \leq 1$ and $\| B_2 - B_0\| \leq 1$ implies that $\| g(B_1) - g(B_2)\| \leq 1/2$ $\| B_1 - B_2\|$, then it follows that g has a unique fixed point. Since, if both B_1 and B_2 are fixed points we have $\| B_1 - B_2\| = \| g(B_1) - g(B_2)\| \leq 1/2$ $\| B_1 - B_2\|$ and thus $\| B_1 - B_2\| = 0$.

PROPOSITION 4.8. Let $g:\mathbb{R}^k \to \mathbb{R}^k$ be defined as in Proposition 4.7. Let $F = \{B \in \mathbb{R}^k \,|\, g(B) = B\}$.

Let $$h(B) = (2\pi e)^{-N/2} \left\{ \prod_{j=1}^{J} \left[\frac{(Y_j - X_jB)^T(Y_j - X_jB)}{m_j} \right]^{m_j} \right\}^{-1/2}.$$

Then:

1) F is a singleton almost surely as $N \uparrow \infty$.

2) There exists a unique B that maximizes h almost surely as $N \uparrow \infty$.

3) $F = \{B^*\}$ where B^* is the unique element of $\mathbb{R}^k$ that maximizes h almost surely as $N \uparrow \infty$.

Proof. Proposition 4.2. gives us that:

1) $$\sup_{B\varepsilon f} \| B - B_0\| < 1 \text{ almost surely as } N \uparrow \infty.$$

Proposition 4.7. yields that:

2) $$\sup_{\substack{\|B_1-B_0\|\leq 1\\ \|B_2-B_0\|\leq 1\\ B_1\neq B_0}} \frac{\|g(B_1)-g(B_2)\|}{\|B_1-B_2\|} \leq 1/2$$

almost surely as $N \uparrow \infty$.

Proposition 4.1. shows that F cannot be empty, thus we have:

3) F is a singleton almost surely as $N \uparrow \infty$.

Proposition 4.1. gives us that there exists $B \in \mathbb{R}^k$ that maximize h and furthermore any such B is a fixed point of g. The rest follows immediately.

COROLLARY 4.9. Let $\Theta(X,Y)$ be an estimator with the property that for each N there exist $\sigma_{1,1}{}^2(N),\ldots,\sigma_{1,J}{}^2(N)$ positive numbers such that $\Theta(X,Y)$ is a limit point of $\{B_n(\sigma_{1,1}{}^2(N),\ldots,\sigma_{1,J}{}^2(N),X,Y))\}$ then:

1) $\Theta(X,Y)$ maximizes the function

$$h(B) = (2\pi e)^{-N/2}\left\{\prod_{n=1}^{J}\left[\frac{(Y_j-X_jB)^T(Y_j-X_jB)}{m_j}\right]^{m_j}\right\}^{-1/2}$$

almost surely as $N \uparrow \infty$.

almost surely as $N \uparrow \infty$.

2) $\Theta(X,Y)$ is independent of the choices for $\sigma_{1,1}^2(N),\ldots,\sigma_{1,J}^2(N)$ almost surely as $N \uparrow \infty$.

Proof. For any choice of $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$, the limit points of $\{B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2,X,Y)\}$ are fixed points of g. Part 1 of the corollary now follows.

If B_1 is a fixed point of g, and if $\sigma_{1,j}^2 = Z_j(B_1)$, then we have $B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2,X,Y) = B_1$ for all n. Therefore if $G = \{B^*\}$, where F is the set of fixed points of g, then it follows that for any choice of $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$ the sequence $\{B_n(\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2,X,Y)\}$ must converge to B^*. Therefore whenever F is a singleton, Θ is independent of the choice of $\sigma_{1,1}^2,\ldots,\sigma_{1,J}^2$.

LEMMA 4.10.

Let $\alpha_N = \Sigma_{j=1}^{J} \frac{m_j(N)}{N} \sigma_j^2$ Then:

A) $\lim_{N \to \infty} \alpha_N := \alpha$ exists and $\max_j \sigma_j^2 \geq \alpha_N \geq \min_j \sigma_j^2$ all N

B) trace $\alpha_N^{-1}\Sigma_N = N$ all N

C) $\lim_{N \to \infty} \frac{1}{N} X^T(\alpha_N^{-1}\Sigma_N)^{-1}X$ exists and is positive definite.

1) Part A follows immediately from assumption 6 and from the fact that α_N is a convex combination of the σ_j^2.

2) Trace $\alpha_N^{-1}\Sigma_N = \alpha_N^{-1}$ trace $\Sigma_N = \frac{1}{\alpha_N} \Sigma_{j=1}^{J} m_j(N)\sigma_j^2 =$

$$\frac{1}{\alpha_N} N \Sigma_{j=1}^{J} \frac{m_j(N)}{N} \sigma_j^2 = \frac{1}{\alpha_N} N \alpha_N = N$$

3) $\frac{1}{N} (X^T(\alpha_N^{-1}\Sigma_N)^{-1}X = \alpha_N \frac{1}{N} X^T\Sigma_N^{-1}X$

By part A) $\lim_{N \to \infty} \alpha_N$ exists and by Assumption VII $\lim_{N\to\infty} \frac{1}{N} X^T\Sigma_N^{-1}X$ exists; thus

4) $\lim_{N \to \infty} \frac{1}{N} X^T(\alpha_N^{-1}\Sigma_N)^{-1}X$ exists.

5) $\frac{1}{N} (X^T(\alpha_N^{-1}\Sigma_N)^{-1}X) = \alpha_N \Sigma_{j=1}^{J} \frac{m_j(N)}{N} \frac{X_j^TX_j}{m_j(N)} \frac{1}{\sigma_j^2}$

Thus

$$|| [\frac{1}{N} (X^T(\alpha_N^{-1}\Sigma_N)X)]^{-1}|| \leq \frac{\max_j \sigma_j^2}{\min_j \sigma_j^2} \cdot \frac{1}{\lambda}$$

and hence: $\lim_{N \to \infty} \frac{1}{N} X^T(\alpha_N^{-1}\Sigma_N)^{-1}X$ is invertible with the norm of the inverse bounded by $\frac{\max_j \sigma_j^2}{\min_j \sigma_j^2} \cdot \frac{1}{\lambda}$.

LEMMA 4.11.

$$E\left\{\left\|\frac{X_j^T \varepsilon_j}{\sqrt{m_j}}\right\|^2\right\} \leq KT\sigma_j^2$$

Proof

$$\left\|\frac{X_j^T \varepsilon_j}{\sqrt{m_j}}\right\|^2 = \Sigma_{k=1}^{k} \frac{1}{\sqrt{m_j}} \Sigma_{i=1} x_{jik}\varepsilon_{ji} \frac{1}{\sqrt{m_j}} \Sigma_{s=1} x_{jsk}\varepsilon_{js}$$

Since $E\,(x_{jik}\varepsilon_{ji}x_{jsk}\varepsilon_{js}) = \begin{cases} 0 & i \neq S \\ x_{jik}^2\ \sigma_j^2 & i = S \end{cases}$

we have:

1) $$E\left\|\frac{X_j^T \Sigma_j}{\sqrt{m_j}}\right\|^2 = \Sigma_{k=1}^{k} \sigma_j^2 \frac{1}{m_j} \Sigma_{i=1}^{m_j} x_{jik}^2$$

(where x_{jik} is the ik element of the matrix X_j).

By assumption $\left\|\frac{X_j^T X_j}{m_j}\right\| \leq T$, and since $\frac{X_j^T X_j}{m_j}$ is positive semi definite symmetric, each element on the diagonal must have values less than or equal to T. Therefore:

2) $$\Sigma_{j=1}^{m_j} \frac{1}{m_j} x_{jik}^2 \leq T$$

Since $E(\varepsilon_{ji}^2) = \sigma_j^2$ and since x_{jik} are non stochastic, we have from 1):

3) $$E\left\|\frac{X_j^T \varepsilon_j}{\sqrt{m_j}}\right\|^2 = \Sigma_{k=1} \sigma_j^2 \Sigma_{i=1}^{m_j} \frac{1}{m_j} x_{jik}^2 \leq K\,\sigma_j^2 T$$

The next lemma is an immediate consequence of Chebychev's inequality.

LEMMA 4.12. Let $X_n \to 0$ in probability and suppose $EY_n^2 \leq M < \infty$, then $X_N Y_n \to 0$ is probability.

Proof. For any $C > 0$, $\delta > 0$, we have:

1) $\text{Prob}\{|X_n Y_n < \delta\} \geq \text{Prob}\{|X_n| < \frac{\delta}{C+1}$ and $|Y_n| \leq C\}$.

By Chebychev's inequality we have:

2) $$\text{Prob}\{|Y_n| \leq C\} \geq |- \frac{E\, Y_n^2}{C^2} \geq | - \frac{M}{C^2} \quad .$$

Let $\gamma > 0$ and let $\delta > 0$; by 2) we can choose C sufficiently large that $\text{Prob}\{|Y_n \geq C\} > | -\frac{\gamma}{2}$. By hypothesis we can find N_o such that $\text{Prob}\{|X_n| < \frac{\delta}{C+1}\} > |- \frac{\gamma}{2}$ for $n \geq No$ Hence for $n \geq No$, we have $\text{Prob}\{|X_n| \geq \frac{\delta}{C+1}$ or $|Y_n| > C\}$ $< \gamma/2 + \gamma/2 = \gamma$ and so $\text{Prob}\{|X_n| < \frac{\delta}{C+1}$ and $|Y_n| \leq C\} \geq |-\gamma$.

[X] Lemma 4.

PROPOSITION 4.13. Let $\theta(X,Y)$ be an estimator with the property that $g(\Theta(X,Y)) = \Theta(X,Y)$ where $g(B) = (X^T\Sigma(B)^{-1}X)^{-1}$ $g(B) = (X^T\Sigma(B)^{-1}X)^{-1}(X^T\Sigma(B)^{-1}Y)$, then under the hypothesis a above:

A: θ is asymptotically equivalent to the weighted least squares estimator with known variances in the sense that if W is the latter, $\underset{N \to \infty}{\text{p lim}} \sqrt{N}(\theta - W) = 0$.

B: θ is asymptotically normally distributed with mean vector B_0 and variance covariance matrix $(X^T\Sigma^{-1}X)^{-1}$ $[= \frac{1}{N}(\frac{1}{N}X^T\Sigma^{-1}X)^{-1}]$.

C: $(\frac{1}{N}X^T\Sigma^{-1}X)^{-1} - (\frac{1}{N}X^T\Sigma(\theta)^{-1}X)^{-1} \to 0$ almost surely as $N \uparrow \infty$.

Proof. Let $V_N = \alpha_N^{-1}\Sigma_N$, then $W = (X^T\Sigma_N^{-1}X)^{-1}X^T\Sigma_N^{-1}Y =$ $= (X^T(\alpha_N^{-1}\Sigma_N)^{-1}X)^{-1}(X^T(\alpha_N^{-1}\Sigma_N)^{-1}Y) = (X^TV_N^{-1}X)^{-1}X^TV_N^{-1}Y$ where X is a $N \times K$ matrix.

Let $\hat{\alpha}_N = \Sigma_{j=1}^{J} \frac{m_j(N)}{N} Z_j(\theta)$ and let $\hat{V}_N = \hat{\alpha}_N^{-1}\Sigma(\theta)$, then $\theta(X,Y) = (X^T \hat{V}_N^{-1}X)^{-1}X^T\hat{V}_N^{-1}Y$.

By theorem 8.4 of Thiel's Principles of Econometrics, to prove A and B it suffices to show

i) $\underset{N \to \infty}{\text{plim}} \frac{1}{N} X^T(\hat{V}_N^{-1} - V_N^{-1})X = 0$ and

ii) $\underset{\to \infty}{\text{plim}} \frac{1}{\sqrt{N}} X^T(\hat{V}_N^{-1} - \hat{V}_N^{-1})\varepsilon = 0$

$\frac{1}{N} X^T(\hat{V}_N^{-1} - V_N^{-1})X = \frac{1}{N} X^T(\hat{\alpha}_N\Sigma(\theta)^{-1} - \alpha_N\Sigma^{-1})X$

Therefore:

1) $\frac{1}{N} X^T(\hat{V}_N^{-1} - V_N^{-1})X = \Sigma_{j=1}^{N} \frac{m_j(N)}{N} \frac{X_j^TX_j}{m_j} (\frac{\hat{\alpha}_N}{Z_j(\theta)} - \frac{\alpha_N}{\sigma_j^2}$

Recall that $\hat{\alpha}_N = \Sigma \frac{m_j(N)}{N} Z_j(\theta)$ and that for any j, $\frac{m_j(N)}{N} Z_j(\theta) - \frac{m_j(N)}{N} \sigma_j^2 \to 0$ almost surely $N \uparrow \infty$.

Hence:

2) $\hat{\alpha}_N - \alpha_N \to 0$ almost surely as $N \uparrow \infty$

and we have $\frac{1}{N} X^T(\hat{V}_N^{-1} - V_N^{-1})X = 0$ almost surely as $N \uparrow \infty$.

$(\frac{1}{\sqrt{N}}) X^T(\hat{V}_N^{-1} - V_N^{-1})\varepsilon = (\frac{1}{\sqrt{N}}) X^T(\hat{\alpha}_N \Sigma(\theta)^{-1} - \alpha_N\Sigma^{-1})\mathcal{E}$ and thus we have:

3) $(\frac{1}{\sqrt{N}})X^T(\hat{V}_N^{-1} - V_N^{-1})\varepsilon = \Sigma_{j=1}\sqrt{\frac{m_j(N)}{N}} \frac{X_j^T\varepsilon_j}{\sqrt{m_j(N)}} (\frac{\hat{\alpha}_N}{Z_j(\theta)} - \frac{\hat{\alpha}_N}{\sigma_j^2})$

By lemma 2 $E\,\left|\left|\frac{X_j{}^T\varepsilon_j}{\sqrt{m_j(N)}}\right|\right|^2 \leq KT\sigma_j{}^2$ independent of N.

For $j \in S_1$, $\frac{\hat{\alpha}_N}{Z_j(\theta)} - \frac{\alpha_N}{\sigma_j{}^2} \to 0$ almost surely and since $0 \leq \sqrt{\frac{m_j(N)}{N}} \leq 1$, it follows that $\sqrt{\frac{m_j(N)}{N}}\left(\frac{\hat{\alpha}_N}{Z_j(\theta)} - \frac{\alpha_N}{\sigma_j{}^2}\right) \to 0$ almost surely for $N \uparrow \infty$.

If $j \in S_2$, then since $\exists\, M < \infty$ such that

$$\left|\frac{\hat{\alpha}_N}{Z_j(\theta)} - \frac{\alpha_N}{\sigma_j{}^2}\right| \leq M \quad \text{almost surely as } N \uparrow \infty.$$

and since $\sqrt{\frac{m_j(N)}{N}} \to 0$ as $N \uparrow \infty$, we have

$\sqrt{\frac{m_j(N)}{N}}\left(\frac{\hat{\alpha}_N}{Z_j(\theta)} - \frac{\alpha_N}{\sigma_j{}^2}\right) \to 0$ almost surely as $N \uparrow \infty$.

We are now in position to use lemma 3, to conclude that,

$$\Sigma_{j=1}^{J}\ \left|\left|\frac{X_j{}^T\varepsilon_j}{m_j(N)}\right|\right|\sqrt{\frac{m_j(N)}{N}}\left(\frac{\hat{\alpha}_N}{Z_j(\theta)} - \frac{\alpha_N}{\sigma_j{}^2}\right),$$

converger to zero in probability and thus:

3) $$\operatorname*{plim}_{N \uparrow \infty} \frac{1}{\sqrt{N}} X^T(\hat{V}_N{}^{-1} - V_N{}^{-1})\varepsilon = 0.$$

$$\frac{1}{N} X^T\Sigma^{-1}X - \frac{1}{N} X^T\Sigma(\theta)^{-1}X = \Sigma_{j=1}^{J} \frac{m_j(N)}{N} \frac{X_j{}^T X_j}{m_j(N)} \left[\frac{1}{\sigma_j{}^2} - \frac{1}{Z_j(\theta)}\right]$$

so part C follows immediately.

5. Linear Variance Model

PROPOSITION 5.1. (Strong Law of large numbers) Let ϕ be a positive even and continuous function on $\mathbb{R}^1$ such that as $|x|$ increases $\frac{\phi(x)}{x} \uparrow$ and $\frac{\phi(x)}{x^2} \downarrow$. Let $\{x_n\}$ be a sequence of independent random variables with $E(x_n) = 0$ for each n and let $0 < a_n \uparrow \infty$. If $\Sigma_n \frac{E(\phi(X_n))}{\phi(a_n)} < \infty$ then $\Sigma_n \frac{x_n}{a_n}$ converges almost every and $\frac{1}{a_n} \Sigma_{j=1}^{n} X_j \to 0$ a.e.

Proof. Chung [9].

PROPOSITION 5.2. Let $\{X_n\}_{n \geq 1}$ be a sequence of independent random variables and suppose for some P, $1 < p \leq 2$ there exists $m < \infty$ with $E|X_n|^p \leq m < \infty$ for all n. Then $\limsup_{n\to\infty} |\frac{1}{n}\Sigma_{n=1}^{n} X_j| < m+2$ almost everywhere.

Proof. 1) Let $Y_n = X_n - E(X_n)$,

$|E(X_n)| \leq E|X_n| \leq 1 + m$

2) $E|Y_n|^p = \int |X_n - E(X_n)|^p$, By Minkowski

$$(E|Y_n|^p)^{1/p} = (\int |X_n - E(X_n)|^p)^{1/p} \leq (\int |X_n|^p)^{1/p} +$$

$$+ (\int |E(X_n)|^p)^{1/2}$$

$$(E|Y_n|^p)^{1/2} \leq m + m + 1 = 2m + 1 \quad .$$

3) In Proposition 5.1. let $\phi(x) = |x|^p$ and let $a_n = n$. We then conclude that since

$$\Sigma_n \frac{E|Y_n|^p}{n^p} \leq \Sigma_n \frac{2m+1}{n^p} \leq 2m+1 \Sigma_n \frac{1}{n^p} < \infty$$

that $\lim_{n\to\infty} \frac{1}{n} \Sigma_{n=1}^{n} Y_j \to 0.$ a.e.

4) $\limsup_{n\to\infty} |\frac{1}{n} \Sigma_{n=1}^{n} Y_j| < |$ a.e. and

5) $\limsup_{n\to\infty} |\frac{1}{n} \Sigma_{n=1}^{n} X_j - \frac{1}{j} \Sigma_{j=1}^{n} E(X_j)| < |$ a.e.

6) $\limsup_{n\to\infty} |\frac{1}{n} \Sigma_{j=1}^{n} X_j| < 1 + m + 1 = m + 2$ a.e.

PROPOSITION 5.3. Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables taking values in the same fixed space E (where $E = \mathbb{R}^1$, $\mathbb{R}^j$ or $\mathbb{R}^{j\times k}$). Let $\{X_n\}_{n\geq 1}$ have a limiting asymptotic distribution D. Let $\{Y_n\}_{n\geq 1}$ be another sequence of random variables taking values in E and if $\underset{n\to\infty}{p\lim} \| X_n - Y_n \| = 0$, then $\{Y_n\}_{n\geq 1}$ has the limiting asymptotic distribution D.

Thiel [48]

Note: Convergence almost everywhere implies convergence in probability.

PROPOSITION 5.4. Let H denote an arbitrary Hilbert Space and $P(H) \subset L(H)$ be the set of bound-d positive linear operators. Then the positive square root function that maps P onto P $(A \mapsto A^{\frac{1}{2}})$ is uniformly continuous.

[$A \in L(H)$ is positive if $A=A^*$ and $(Ax,x) \geq 0$ for all x in H.] ($A=A^*$ is implied by (Ax,x) real valued for all x in H).

Proof. (Sketch)

1) The positive square root function is a monotone function when restricted to P(H).

2) The square root function defined on $(0,\infty)$ is uniformly continuous.

3) Therefore if $\{A_n\}$ and $\{B_n\}$ are sequences of positive operator such that $\|A_n - B_n\| \to 0$ and if $\varepsilon > 0$, then there is a $N_0 < \infty$ such that for $n \geq N_0 A_n \leq B_n + \varepsilon I$ and $B_n \leq A_n + \varepsilon I$.

4) $\sqrt{A_n} \quad \sqrt{B_n + \varepsilon I} = \sqrt{B_n} + \eta_{B_n}(\varepsilon) \sqrt{B_n} \leq \sqrt{A_n + \varepsilon I} =$
$\sqrt{A_n} + \eta_{A_n}(\varepsilon)$ where $\lim_{\varepsilon \to 0} \|\eta_{B_n}(\varepsilon)\| \to 0$ independent of B_n.

We are now in a position to list our assumptions and to prove our results. For the convenience of the reader, we have followed much of the notation of White [59]. We have also borrowed liberally from him on the wording of our assumption.

A1) The model is known to be

(6) $Y_i = X_i \beta_0 + \varepsilon_i \qquad i = 1,2,\ldots,n$

(7) $E(\varepsilon_i) = 0 \qquad i = 1,2,\ldots,n$

(8) $E(\varepsilon_i^{\,2}) = \sigma_i^{\,2} = Z_i \Gamma_0$

Where X_i is a 1×k vector of random variables, ε_i and Y_i are real valued random variables, β_0 is a k×1 vector of real numbers. Y_i and X_i are observable, ε_i is unobservable and β_0 is to be estimated or hypothesis concerning β_0 are to be tested. Z_i is a 1×m vector of real valued random variables which may contain some or all of the variables in the vector X_i. Γ_0 is a m×1 unknown vector of real numbers which is to be estimated or hypothesis concerning Γ_0 are to be tested.

Let W_i be the vector of length p of random variables whose first entry W_{i1} is the scalar 1 and whose other entries are exactly those random variables that appear in X_i or Z_i. We assume that $E(W_{ij}W_{ir}\varepsilon_i) = 0 \;\; 1 \leq j,k \leq p$ and $E(W_i^T(\varepsilon_i^2 - \sigma_i^2))=0$. We let μ_i denote $\varepsilon_i^2 - \sigma_i^2$. The vectors (W_i, ε_i) are assumed to be a sequence of independent though not necessarily identically distributed random vectors.

A2) I) There exists $0 < \delta \leqslant 1$ and $\Delta < \infty$ such that for all i:

a) $E(|\varepsilon_i^2 W_{ir}W_{is}W_{it}W_{iv}|^{1+\delta}) \leqslant \Delta \qquad 1 \leqslant r,s,t,v \leqslant p$

b) $E(|\varepsilon_i W_{ik}W_{ir}W_{is}W_{it}W_{iv}|^{1+\delta}) \leqslant \Delta \qquad 1 \leqslant k,r,s,t,v \leqslant p$

c) $E(|W_{ij}W_{ik}W_{ir}W_{is}W_{it}W_{iv}|^{1+\delta}) \leqslant \Delta \qquad 1 \leqslant j,k,r,s,t,v \leqslant p$

d) $E(|\mu_i^2|^{1+\delta}) \leqslant \Delta$

e) $E(|\varepsilon_i^3 W_{ir}W_{is}W_{it}|^{1+\delta}) \leqslant \Delta \qquad 1 \leqslant r,s,t \leqslant p$

f) $E(|\varepsilon_i^4 W_{ir}W_{is}|^{1+\delta}) \leqslant \Delta \qquad 1 \leqslant r,s \leqslant p$

II) Let $\bar{M}_n^a = n^{-1} \Sigma_{i=1}^n E(X_i^T X_i)$ and let

$$\bar{M}_n^b = n^{-1} \Sigma_{i=1}^n E(Z_i^T Z_i)$$

We assume that there exists $N_0 < \infty$ and $0 < \lambda$ such that for $n \geq N_0$ minimum eigenvalue of $\overline{M}_n^a \geq \lambda$ and minimum eigenvalue of $\overline{M}_n^b \geq \lambda > 0$ (Note by the first part of A2) this is equivalent to the property that for n sufficiently large det $\overline{M}_n^a$ and det $\overline{M}_n^b$ is bounded away from zero. Also, observe that we can choose δ and λ so that they are equal.

A3) Let $\overline{V}_n^a = n^{-1} \Sigma_{i=1}^n E(\varepsilon_i{}^2 X_i{}^T X_i)$ and

let $\overline{V}_n^b = n^{-1} \Sigma_{i=1}^n E(\mu_i{}^2 Z_i{}^T Z_i)$.

We assume that there exists $N_0 < \infty$ and $\lambda > 0$ such that for $n \geq N_0$ minimum eigenvalues of $\overline{V}_n^a \geq \lambda > 0$ and minimum eigenvalue of $\overline{V}_n^b \geq \lambda \geq 0$. (There is no loss in generality in assuming that N_0, λ is A2 and N_0, λ in A3 are the same. In the presence of A2, A3 is the equivalent of the assumption that for n sufficiently large minimum (det $\overline{V}_n^a$, det $\overline{V}_n^b$) is bounded away from and above zero.

The first theorem is a restatement of a result found in White [ ] and its proof can be found therein. Before stating Theorem 1 we introduce additional notation.

$$\text{Let } \hat{\beta}_n = \begin{cases} (X^TX)^{-1}X^TY & \text{if } (X^TX) \text{ is nonsingular} \\ 0 & \text{if } (X^TX) \text{ is singular} \end{cases}$$

$$\text{Let } \hat{\alpha}_n \begin{cases} (Z^TZ)^{-1}Z^T\varepsilon^2 & \text{if } (Z^TZ) \text{ is nonsingular} \\ 0 & \text{if } (Z^TZ) \text{ is singular} \end{cases}$$

ε^2 is the n×1 column vector whose i^{th} entry is $(\varepsilon_i)^2)$

Let $\hat{\varepsilon}_{in} = Y_i - X_i\hat{\beta}_n$

Let $\hat{\mu}_{in} = \varepsilon_i^2 - Z_i\hat{\alpha}_n$

Let $\hat{V}_n^a = n^{-1} \Sigma_{i=1}^{n} \hat{\varepsilon}_{in}^2 X_i^T X_i$

Let $V_n^b = n^{-1} \Sigma_{i=1}^{n} \hat{\mu}_{in}^2 Z_i^T Z_i$

Let R^a be a $q \times k$ matrix of real numbers with full row rank and let r^a be a $q \times 1$ vector of real numbers.

Let R be a $q \times m$ matrix of real numbers of full row rank and let r be a $q \times 1$ vector of real numbers.

THEOREM 5.5. (White)

Under assumptions A1, A2, and A3, we have the following:

i) $\hat{\beta}_n \to \beta_o$ almost everywhere (a.e.)

ii) $\hat{\alpha}_n \to \Gamma_o$ a.e.

iii) $\sqrt{n}\ [(\frac{X^T X}{n})^{-1}\ \hat{V}_n^a\ (\frac{X^T X}{n})^{-1}]^{\frac{1}{2}}\ (\hat{\beta}_n - \beta_o) \overset{A}{\sim} N(0, I_k)$

iv) $\sqrt{n}\ [(\frac{Z^T Z}{n})^{-1}\ \hat{V}_n^b\ (\frac{Z^T Z}{n})^{-1}]^{\frac{1}{2}}\ (\hat{\alpha}_n - \alpha_o) \overset{A}{\sim} N(0, I_m)$

v) under the hypothesis H_o: $R^a \beta_o = r^a$

$$n(R^a \hat{\beta}_n - r^a)^T [R^a\ (\frac{X^T X}{n})^{-1}\ \hat{V}_n^a\ (\frac{X^T X}{n}) R^{aT}]^{-1} (R^a \beta_n - r^a) \overset{A}{\sim} \chi_q^2$$

vi) under the hypothesis H_o: $R\Gamma_o = r$

$$n(R\hat{\alpha}_n - r)^T [R(\frac{Z^T Z}{n})^{-1}\ \hat{V}_n^b\ (\frac{Z^T Z}{n})^{-1} R^T]^{-1} (R\hat{\alpha}_n - r) \overset{A}{\sim} \chi_q^2$$

Before stating this theorem it is once again necessary to introduce additional notation.

Let $\hat{\varepsilon}_n^2$ be the n×1 column vector whose i^{th} entry is $\hat{\varepsilon}_{in}^2$.

$$\text{Let } \hat{\Gamma}_n = \begin{cases} (Z^TZ)^{-1}Z^T\hat{\varepsilon}_n^2 & \text{if } Z^TZ \text{ is nonsingular} \\ 0 & \text{otherwise} \end{cases}$$

Let $\hat{w}_{in} = \hat{\varepsilon}_{in}^2 - z_i\hat{\Gamma}_n$ and

Let $\hat{V}_n^c = n^{-1} \Sigma_{i=1}^n \hat{w}_{in}^2 z_i^T z_i$

THEOREM 5.6.

Under assumption A1, A2, and A3, the following hold:

i) $\hat{\Gamma}_n \to \Gamma_o$ a.s.

ii) $\sqrt{n}\ [(\frac{Z^TZ}{n})^{-1}\ \hat{V}_n^c(\frac{Z^TZ}{n})^{-1}]^{-\frac{1}{2}}\ (\hat{\Gamma}_n - \Gamma_o) \overset{A}{\sim} N(0, I_m)$

iii) under the hypothesis H_o: $R\Gamma_o = r$ (where R,r are as in theorem 1

$$n(R\hat{\Gamma}_n - r)^T\ [R(\frac{Z^TZ}{n})^{-1}\ \hat{V}_n^c\ (\frac{Z^TZ}{n})^{-1}R^T]^{-1}(R\hat{\Gamma}_n - r) \sim \chi_q^2$$

(Note: A part of the statement of this theorem is that matrices whose inverses must be taken ~~with~~ are, as $n \uparrow \infty$, ~~be~~ nonsingular almost everywhere).

Proof. See Chapter IV.

1) The idea of the proof is quite simple. We show that the difference is norm between the statistics stated in Theorem 2 and the associated (noncomputable) statistics in Theorem 1 converges to zero as $n \uparrow \infty$ almost everywhere. Hence by

proposition 3 and theorem 1, the desired results hold. Unfortunately, the only proof I know involves a large amount of computation.

2) $\| \hat{\Gamma}_n - \Gamma_0 \| \le \| \hat{\Gamma}_n - \hat{\alpha}_n \| + \| \hat{\alpha}_n - \Gamma_0 \|$ and so by theorem I, it suffices to show $\| \hat{\Gamma}_n - \hat{\alpha}_n \| \to 0$ a.e.

3) $$\hat{\Gamma}_n - \hat{\alpha}_n = \left(\frac{Z^T Z}{n}\right)^{-1} \left(\frac{Z^T(\hat{\varepsilon}_n^{\,2} - \varepsilon^2)}{n}\right)$$

$$\| \hat{\Gamma}_n - \hat{\alpha}_n \| \le \left\| \left(\frac{Z^T Z}{n}\right)^{-1} \right\| \left\| \frac{Z^T(\hat{\varepsilon}_n^{\,2} - \varepsilon^2)}{n} \right\|$$

By SLLN (Prop. 1) $\left\| \frac{Z^T Z}{n} - n^{-1} \Sigma_{i=1}^n E(Z_i^T Z_i) \right\| \to 0$ a.e.

By A2 for $n \ge N_0$. $\| [n^{-1} \Sigma_{i=1}^n E(Z_I^T Z_i)]^{-1} \| \le \frac{1}{\lambda}$.

It follows from standard Banach Algebra techniques that if $\left\| \left(\frac{Z^T Z}{n}\right) - n^{-1} \Sigma_{i=1}^n E(Z_i^T Z_i) \right\| < \frac{1}{2}\lambda$, $\frac{Z^T Z}{n}$ is invertible and $\left\| \left(\frac{Z^T Z}{n}\right)^{-1} \right\| < \frac{2}{\lambda}$. Therefore, to show i) it is enough to show

$$\left\| \frac{Z^T(\hat{\varepsilon}_n^{\,2} - Z^T \varepsilon^2}{n} \right\| \to 0 \text{ a.e.}$$

$n^{-1} Z^T(\hat{\varepsilon}_n^{\,2} - \varepsilon^2) = n^{-1} \Sigma_{i=1}^n Z_i^{\,T} [(Y_i - X_i \hat{\beta}_n)^2 - \varepsilon_i^{\,2}] =$

$n^{-1} \Sigma_{i=1}^n Z_i^{\,T} [X_i \beta_0 + \varepsilon_i - X_i \hat{\beta}_n)^2 - \varepsilon_i^{\,2}] =$

$n^{-1} \Sigma_{i=1}^n Z_i^{\,T} [X_i (\beta_0 - \hat{\beta}_n) X_i (\beta_0 - \hat{\beta}_n) + 2\varepsilon_i X_i (\beta_0 - \hat{\beta}_n)] =$

$n^{-1} \Sigma_{i=1}^n Z_i^T \Sigma_{r=1}^m \Sigma_{s=1}^m X_{ir} (\beta_0 - \hat{\beta}_n)_r X_{is} (\beta_0 - \hat{\beta}_n)_s + 2\varepsilon_i X_{ir} (\beta_0 - \hat{\beta}_n)_r$.

Now $n^{-1} Z^r(\hat{\varepsilon}_n^2 - \varepsilon^2)$ is a m×1 vector and m is a fixed finite number; thus to prove convergence to zero, it suffices to prove component wise convergence to zero. Furthermore, since the sums indexed by r and s are of a fixed finite length, we are reduced to showing.

4) $n^{-1} (\Sigma_{i=1}^n Z_{ir} X_{ir} X_{is}) (\beta_0 - \hat{\beta}_n)_r (\beta_0 - \hat{\beta}_n)_2 + 2(n^{-1} \Sigma_{i=1}^n Z_{ir} \varepsilon_i X_{ir}) (\beta_0 - \hat{\beta}_n)_r$

converge to zero a.e. ($1 \leq k_1 r, s \leq m$).

This is an immediate consequence of A2, proposition

5) If $\| XY \| \leq \| X \| \, \| Y \|$, then to prove $\text{plim} \, \| A_n B_n - C_n D_n \| = 0$, it suffices to show:

a) $\text{plim} \, \| A_n - C_n \| = 0$

b) $\text{plim} \, \| B_n - D_n \| = 0$

c) for any $\varepsilon > 0$, there exists $M(\varepsilon) < \infty$ and $N(\varepsilon) < \infty$ such that for $n \geq N(\varepsilon)$.

$\text{prob} \, \{\| A_n \| \leq M(\varepsilon)\} \geq 1-\varepsilon$ and

$\text{prob} \, \{\| B_n \| \leq M(e)\} \geq 1-\varepsilon$

This is true since $\| A_n B_n - C_n D_n \| = \| A_n B_n - A_n D_n + A_n D_n - C_n D_n \|$

$$\leq \| A_n \| \, \| B_n - D_n \| + \| A_n - C_n \| \, \| D_n \|$$

$$\leq \| A_n \| \, \| B_n - D_n \| + \| A_n - C_n \| (\| D_n - B_n \| + \| B_n \|)$$

6) $\sqrt{n} \, (\hat{\Gamma}_n - \hat{\alpha}_n) = (\frac{\hat{Z}^T \hat{Z}}{n})^{-1} \, [\sqrt{n} \, n^{-1} \, Z^T (\hat{\varepsilon}_n{}^2 - \varepsilon^2)]$ as we saw in 3) above, to show $\text{plim} \sqrt{n} \, (\Gamma_n - \alpha_n) = 0$ we are reduced to showing for $1 \leq k, r, s \leq m$

a) $\text{plim}(n^{-1} \, \Sigma_{i=1}^{n} Z_{ir} X_{ir} X_{is}) (\sqrt{n} (\beta_o - \hat{\beta}_n)_r) \, (\beta_o - \hat{\beta}_n)_s = 0$

b) $\text{plim} \, (n^{-1} \, \Sigma_{i=1}^{n} \varepsilon_i Z_{ik} X_{ir}) (\sqrt{n} (\beta_o - \hat{\beta}_n)_r) = 0$

$$\sqrt{n} \, \bar{V}_n^{a^{-\frac{1}{2}}} \, \bar{M}_n^a \, (\hat{\beta}_{on} - \beta_o) \sim N(0, I).$$

$$\| (\bar{V}_n^{a^{-\frac{1}{2}}} \bar{M}_n^a)^{-1} \| \leq \| (\bar{M}_n^a)^{-1} \| \cdot \| (\bar{V}_n^a)^{\frac{1}{2}} \| \leq \frac{1}{\lambda} \| \bar{V}_n^{a \, \frac{1}{2}} \| \leq \frac{1}{\lambda} \, C_n^{\frac{1}{2}}$$

Assumption A2, I.a. ensures us that C_n is uniformly bounded.

for any $\varepsilon < 0$, there is a $M(\varepsilon) < \infty$ and an $N(\varepsilon) < \infty$ such that $n \geq N(\varepsilon)$ implies probability $\{\|\sqrt{n}(\hat{\beta}_n - \beta_0)\| \leq M(\varepsilon)\} \geq 1-\varepsilon$. Since $(\hat{\beta}_n - \beta_0) \to 0$ a.e., $n^{-1}\Sigma_{i=1}^{n}\varepsilon_i Z_{ik} X_{ir} \to 0$ a.e. and since there exists $M < \infty$ such that $\| n^{-1}\Sigma_{i=1}^{n} Z_{ik} X_{ir} X_{is} \| < M$ a.e. $(n \to \infty)$, 6a and 6b hold, and plim $\sqrt{n}\ (\hat{\Gamma}_n - \hat{\alpha}_n) = 0$.

$$7)\quad \left(\frac{Z^TZ}{n}\right)^{-1} \hat{V}_n^c \left(\frac{Z^TZ}{n}\right)^{-1} - \left(\frac{Z^TZ}{n}\right)^{-1} \hat{V}_n^b \left(\frac{Z^TZ}{n}\right)^{-1} =$$

$$\left(\frac{Z^TZ}{n}\right)^{-1} [\hat{V}_n^c - \hat{V}_n^b] \left(\frac{Z^TZ}{n}\right)^{-1}$$

$$8)\quad \hat{V}_n^c - \hat{V}_n^b = n^{-1} \Sigma_{i=1}^{n} (\hat{w}_{in}{}^2 - \hat{\mu}_{in}{}^2) Z_i^T Z_i$$

It is necessary to show that plim $\| \hat{V}_n^c - \hat{V}_n^b \| = 0$. Unfortunately this requires the following long computation.

Since $\hat{V}_n^c - \hat{V}_n^b$ is a $m \times m$ matrix and m is a fixed finite number show for $1 \leqslant j,\ell \leqslant m$ plim $n^{-1} \Sigma_{i=1}^{n} Z_{ij} Z_{i\ell} (\hat{w}_{in}{}^2 - \hat{\mu}_{in}{}^2) = 0$

$$\hat{w}_{in} = \hat{\varepsilon}_{in}{}^2 - Z_i \hat{\Gamma}_n = \varepsilon_i^2 - Z_i \hat{\alpha}_n + Z_i(\hat{\alpha}_n - \hat{\Gamma}_n) + \hat{\varepsilon}_{in}{}^2 - \varepsilon_i{}^2$$

$$\hat{\mu}_{in} = \varepsilon_i^2 - Z_i \hat{\alpha}_n,$$

$$9)\quad n^{-1} \Sigma_{i=1}^{n} Z_{ij} Z_{i\ell} (\hat{w}_{in}{}^2 - \hat{\mu}_{in}{}^2) =$$

$$n^{-1} \Sigma_{i-1}^{n} Z_{ij} Z_{i\ell} \{[Z_i(\hat{\alpha}_n - \hat{\Gamma}_n)]^2 + [\hat{\varepsilon}_{in}{}^2 - \varepsilon_i^2]^2$$

$$2(\varepsilon_i^2 - Z_i\hat{\alpha}_n) Z_i (\hat{\alpha}_n - \hat{\Gamma}_n) + 2(\varepsilon_i^2 - Z_i \hat{\alpha}_n)(\hat{\varepsilon}_{in}{}^2 - \varepsilon_i^2) + 2Z_i(\hat{\alpha}_n - \hat{\Gamma}_n)(\hat{\varepsilon}_{in}{}^2 - \varepsilon_i^2)\}.$$

It is now convenient to proceed term by term

$$10)\quad n^{-1} \Sigma_{i=1}^{n} Z_{ij} Z_{i\ell} [Z_i(\hat{\alpha}_n - \hat{\Gamma}_n)]^2 =$$

$$\Sigma_{r=1}^{m} \Sigma_{s=1}^{m} (n^{-1} \Sigma_{i=1}^{m} Z_{ij} Z_{i\ell} Z_{ir} Z_{is}) (\hat{\alpha}_n - \hat{\Gamma}_n)_s (\hat{\alpha}_n - \hat{\Gamma}_n)_r$$

Since $\|\hat{\alpha}_n - \hat{\Gamma}_n\| \to 0$ a.e. $(n \to \infty)$, it follows that
$n^{-1}\Sigma_{i=1}^{n} Z_{ij} Z_{i\ell}[Z_i(\hat{\alpha}_n - \hat{\Gamma}_n)]^2 \to 0.$ a.e. $(n \to \infty)$

(11)

$$\hat{\varepsilon}_{in} = Y_i - X_i\hat{\beta}_n = X_i\beta+\varepsilon_i - X_i\hat{\beta}_n = X_i(\beta-\hat{\beta}_n)+\varepsilon_i$$

$$\hat{\varepsilon}_{in}{}^2 = \varepsilon_i^2 + 2\varepsilon_i X_i(\beta-\hat{\beta}_n) + [X_i(\beta-\hat{\beta}_n)]^2$$

$$(\hat{\varepsilon}_{in}{}^2 - \varepsilon_i^2)^2 = 4\varepsilon_i^2[X_i(\beta-\hat{\beta}_n)]^2 + 4\varepsilon_i[X_i(\beta-\hat{\beta}_n)]^3 + [X_i(\beta-\hat{\beta}_n)]^4$$

$$n^{-1}\Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}[\hat{\varepsilon}_{in}{}^2-\varepsilon_i^2]^2 =$$

$$\Sigma_{r=1}^{k}\Sigma_{s=1}^{k} 4(n^{-1}\Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}\varepsilon_i^2 X_{ir}X_{is})(\beta-\hat{\beta}_n)_s(\beta-\hat{\beta}_n)_r +$$

$$\Sigma_{r=1}^{k}\Sigma_{s=1}^{k}\Sigma_{t=1}^{k}(4\ n^{-1}\ \Sigma_{i=1}^{n}\varepsilon_i Z_{ij}Z_{i\ell}X_{ir}X_{is}X_{it})(\beta-\hat{\beta}_n)_t(\beta-\hat{\beta}_n)_s(\beta-\hat{\beta})_r$$

$$\Sigma_{r=1}^{k}\Sigma_{s=1}^{k}\Sigma_{t=1}^{k}\Sigma_{v=1}^{k}(n^{-1}\ \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}X_{ir}X_{is}X_{it}X_{iv})(\beta-\hat{\beta}_n)_v (\beta-\hat{\beta}_n)_t(\beta-\hat{\beta}_n)_s(\beta-\hat{\beta}_n)_r$$

$\|\beta-\hat{\beta}_n\| \to 0$ a.e. $(n \to \infty)$, so again by
assumption A2, $n^{-1}\ \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}[\hat{\varepsilon}_{in}{}^2-\varepsilon_i^2] \to 0$ a.e.

(12)

$$n^{-1}\ \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}\ 2(\varepsilon_i^2 - Z_i\hat{\alpha}_n)Z_i(\hat{\alpha}_n - \hat{\Gamma}_n) =$$

$$2\Sigma_{r=1}^{m}(n^{-1}\ \Sigma_{i=1}^{n}\varepsilon_i^2 Z_{ij}Z_{i\ell}Z_{ir})(\hat{\alpha}_n-\hat{\Gamma}_n)_r +$$

$$2\Sigma_{r=1}^{m}\Sigma_{s=1}^{m}(n^{-1}\ \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}Z_{ir}Z_{is})\hat{\alpha}_{ns}\ (\hat{\alpha}_n-\hat{\Gamma}_n)_r$$

Now $\hat{\alpha}_{ns} = \Gamma_{os} + (\hat{\alpha}_n - \Gamma_o)_s$

so $\|\hat{\alpha}_n\| < \|\Gamma_o\| + 1$ a.e. $(n \to \infty)$.

Thus, $n^{-1}\Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}\, 2(\varepsilon_i^2 - Z_i\hat{\alpha}_n)Z_i(\hat{\alpha}_n - \hat{\Gamma}_n) \to 0$ a.e.

(13)

$$n^{-1}\, \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell} 2(\varepsilon_i^2 - Z_i\hat{\alpha}_n)(\hat{\varepsilon}_{in}{}^2 - \varepsilon_i^2) =$$

$$n^{-1}\, \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}\, 2(\varepsilon_i^2 - Z_i\hat{\alpha}_n)(2\varepsilon_i X_i(\beta-\hat{\beta}_n) + [X_i(\beta-\hat{\beta}_n)]^2) =$$

$$\Sigma_{r=1}^{k}(4n^{-1}\, \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}X_{ir}\varepsilon_i^3)(\beta-\hat{\beta}_n)_r \;+$$

$$\Sigma_{r=1}^{k}\Sigma_{s=1}^{k}\, 2(n^{-1}\, \Sigma_{i=1}^{n}\varepsilon_i^2 Z_{ij}Z_{i\ell}X_{ir}X_{is})(\beta-\hat{\beta}_n)_s(\beta-\hat{\beta}_n)_r \;+$$

$$-\Sigma_{r=1}^{m}\Sigma_{s=1}^{k}\, 4(n^{-1}\, \Sigma_{i=1}^{n}\varepsilon_i Z_{ij}Z_{i\ell}Z_{ir}X_{is})\hat{\alpha}_{nr}(\beta-\hat{\beta}_n)_s \;+$$

$$-\Sigma_{r=1}^{m}\Sigma_{s=1}^{k}\Sigma_{t=1}^{k}\, 2(n^{-1}\Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}Z_{ir}X_{is}X_{it})\hat{\alpha}_{nr}(\beta-\hat{\beta}_n)_s(\beta-\hat{\beta}_n)_t$$

So we conclude $n^{-1}\Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}2(\varepsilon_i^2 - Z_i\hat{\alpha}_n)(\hat{\varepsilon}_{in}{}^2 - \varepsilon_i^2) \to 0$ a.e.

(14)

$$n^{-1}\, \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}\, 2Z_i(\hat{\alpha}_n - \hat{\Gamma}_n)(\hat{\varepsilon}_{in}{}^2 - \varepsilon_i^2) =$$

$$n^{-1}\Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}\, 2Z_i(\hat{\alpha}_n - \hat{\Gamma}_n)(2\varepsilon_i X_i(\beta-\hat{\beta}_n) + [X_i(\beta-\hat{\beta}_n)]^2) =$$

$$\Sigma_{r=1}^{m}\Sigma_{s=1}^{k}\, 4(n^{-1}\Sigma_{i=1}^{n}\varepsilon_i Z_{ij}Z_{i\ell}Z_{ir}X_{is})(\hat{\alpha}_n - \hat{\Gamma}_n)_r(\beta-\hat{\beta}_n)_s \;+$$

$$\Sigma_{r=1}^{m}\Sigma_{s=1}^{k}\Sigma_{t=1}^{k}\, 2(n^{-1}\, \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}Z_{ir}X_{is}X_{it})(\hat{\alpha}_n - \hat{\Gamma}_n)_r(\beta-\hat{\beta}_n)_s(\beta-\hat{\beta}_n)_t$$

so $n^{-1}\, \Sigma_{i=1}^{n} Z_{ij}Z_{i\ell}\, 2Z_i(\hat{\alpha}_n - \hat{\Gamma}_n)(\hat{\varepsilon}_{in}{}^2 - \varepsilon_i^2) \to 0$ a.e. $(n \to \infty)$.

and thus

15) $\| \hat{V}_n^b - \hat{V}_n^c \| \to 0$ a.e. $(n \to \infty)$.

16) $\| \hat{V}_n^b - \overline{V}_n^b \| = \| \hat{V}_n^b - n^{-1} \Sigma_{i=1}^n E(\mu_i^2 Z_i^T Z_i) \| \to 0$ a.e. $(n \to \infty)$.

hence $\| \hat{V}_n^c - \overline{V}_n^b \| \to 0$ a.e. $(n \to \infty)$.

For n sufficiently large, $(\overline{V}_n^b)^{-1}$ exists and $\| (\overline{V}_n^b)^{-1} \| \le \frac{1}{\lambda}$

Therefore $\| (\hat{V}_n^c)^{-1} \| \leq \frac{2}{\lambda}$ a.e. $(n \to \infty)$, which of course implies the existence of $(\hat{V}_n^c)^{-1}$.

17) $\frac{Z^T Z}{n} = n^{-1} \Sigma_{i=1}^n Z_i^T Z_i$, By Proposition 1,

$\| (\frac{Z^T Z}{n}) - \overline{M}_n^b \| \to 0$ a.e. $(n \to \infty)$.

By assumption A2.c) there exists $M < \infty$ such that $\| \overline{M}_n^b \| < M$ all n, hence $\| (\frac{Z^T Z}{n}) \| < M+1$ a.e. as $n \to \infty$.

18) Furthermore, by assumption, for n sufficiently large, $\| (\overline{M}_n^b)^{-1} \| \leq \frac{1}{\lambda}$. We conclude then that

19) $\| [(\frac{Z^T Z}{n})^{-1} \hat{V}_n^b (\frac{Z^T Z}{n})^{-1}]^{-1} \| \leq (M+1)^2 \cdot \frac{2}{\lambda}$ a.e. $(n \to \infty)$

and that

20) $\| [(\frac{Z^T Z}{n})^{-1} \hat{V}_n^b (\frac{Z^T Z}{n})^{-1}]^{-1} - [(\frac{Z^T Z}{n})^{-1} \hat{V}_n^c (\frac{Z^T Z}{n})^{-1}]^{-1} \| \to 0$ a.e.

21) By 20) and Proposition 4).

$\| [(\frac{Z^T Z}{n})^{-1} \hat{V}_n^b (\frac{Z^T Z}{n})^{-1}]^{-\frac{1}{2}} - [(\frac{Z^T Z}{n})^{-1} \hat{V}_n^c (\frac{Z^T Z}{n})^{-1}]^{-\frac{1}{2}}] \| \to 0$ a.e.

22) It follows from White [60], that $\sqrt{n}\,\bar{V}_n^{b-\frac{1}{2}}\,\bar{M}_n^{b}(\hat{\alpha}_n-\Gamma_o) \overset{A}{\sim}$ $N(0,I_m)$ and hence for any $\varepsilon > 0$, there exists $M(\varepsilon) < \infty$, $N(\varepsilon) < \infty$ such that for $n \geq N(\varepsilon)$

$$\text{prob}\ \{\sqrt{n}\,\|\hat{\alpha}_n-\Gamma_o\| \leqslant M(\varepsilon)\} \geqslant 1-\varepsilon$$

24) Recalling 5) we see that

$$\text{plim}\ \|\,[(\tfrac{Z^TZ}{n})^{-1}\ \hat{V}_n^b(\tfrac{Z^TZ}{n})^{-1}]^{-\frac{1}{2}}\ \sqrt{n}(\hat{\alpha}_n-\Gamma_o) -$$

$$- [(\tfrac{Z^TZ}{n})^{-1}\ \hat{V}_n^c(\tfrac{Z^TZ}{n})^{-1}]^{-\frac{1}{2}}\ \sqrt{n}(\hat{\Gamma}_n-\hat{\Gamma}_o)\,\| = 0$$

where we need only note $\|\sqrt{n}(\hat{\alpha}_n-\Gamma_o) - \sqrt{n}(\hat{\Gamma}_n-\Gamma_0)\| = \|\sqrt{n}(\hat{\alpha}_n-\hat{\Gamma}_n)\|$
Thus part ii follows.

27) $$\|R(\tfrac{Z^TZ}{n})^{-1}\ \hat{V}_n^c(\tfrac{Z^TZ}{n})^{-1}R^T - R(\tfrac{Z^TZ}{n})^{-1}\ \hat{V}_n^b(\tfrac{Z^TZ}{n})^{-1}R^T\| =$$

$$\|R(\tfrac{Z^TZ}{n})^{-1}(\hat{V}_n^c - \hat{V}_n^b)(\tfrac{Z^TZ}{n})^{-1}R^T\| \leqslant \|R\|^2\,\|(\tfrac{Z^TZ}{n})^{-1}\|^2\,\|\hat{V}_n^c - \hat{V}_n^b\|$$

hence

$$\|R(\tfrac{Z^TZ}{n})^{-1}\ \hat{V}_n^c(\tfrac{Z^TZ}{n})^{-1}R^T - R(\tfrac{Z^TZ}{n})^{-1}\ \hat{V}_n^b(\tfrac{Z^TZ}{n})^{-1}R^T\| \to 0 \quad \text{a.e.}$$

$$(n \to \infty)$$

28) Let P be a positive $m\times m$ matrix and $\|P^{-1}\| < s$
Now R is a $q\times m$ matrix of full row rank; hence there is a $\delta > 0$ $\inf\limits_{\substack{X\varepsilon\mathbb{R}^q \\ \|X\|=1}} \|R^TX\| > \delta$.

$$\langle RPR^TX,X\rangle = \langle PR^TX, R^TX\rangle = \frac{1}{m}\|R^TX\|^2 \geqslant \frac{\delta^2}{s}$$

hence RPR^T is invertible and $\|(RPR^T)^{-1}\| \leqslant \dfrac{s}{\delta^2} = \dfrac{\|P^{-1}\|}{\delta^2}$

29) $(\frac{Z^TZ}{n})^{-1}\ \hat{V}_n^c(\frac{Z^TZ}{n})^{-1}$ is a positive operator, such that

$$\| [(\frac{Z^TZ}{n})^{-1}\ \hat{V}_n^c(\frac{Z^TZ}{n})^{-1}]^{-1} \| < \frac{(M+1)^2 2}{\lambda} \text{ a.e. } (n \to \infty).$$

where $\| (\frac{Z^TZ}{n}) \| < M+1$ a.e. $(n \to \infty)$.

Hence we conclude

$$\| [R(\frac{Z^TZ}{n})^{-1}\ \hat{V}_n^c(\frac{Z^TZ}{n})^{-1}R^T]^{-1} - [R(\frac{Z^TZ}{n})^{-1}\ \hat{V}_n^b(\frac{Z^TZ}{n})^{-1}R^T]^{-1} \| \to 0$$

a.e.

30) $\| \sqrt{n}(R\hat{\alpha}_n - r) - \sqrt{n}(R\hat{\Gamma}_n - r) \| = \| \sqrt{n}(R\hat{\alpha}_n - r)^T - \sqrt{n}(R\hat{\Gamma}_n - r)^T \|$

$$\leqslant \| R \| \, \| \sqrt{n}(\hat{\alpha}_n - \hat{\Gamma}_n) \|;$$

$$\text{plim} \| \sqrt{n}(R\hat{\alpha}_n - r) - \sqrt{n}(R\Gamma_n - r) \| =$$

$$\text{plim} \| \sqrt{n}(R\hat{\alpha}_n - r)^T - \sqrt{n}(R\hat{\Gamma}_n - r)^T \| = 0$$

31) Under H_o; $\sqrt{n}(R\hat{\alpha}_n - r) = \sqrt{n}(R\hat{\alpha}_n - R\Gamma_o) = R\sqrt{n}(\hat{\alpha}_n - \Gamma_o)$;

thus under H_o, for $\varepsilon > 0$, there is a $M(\varepsilon) < \infty$ and $N(\varepsilon) < \infty$ such that $n \geq N(\varepsilon)$ implies prob $\{\| \sqrt{n}(R\hat{\alpha}_n - r) \| \leqslant M(\varepsilon)\} \geqslant 1-\varepsilon$.

Observing that our norm structures are such that $\| XY \| \leq \| X \| \, \| Y \|$ and hence that $\| ABC-DEF \| \leq \| A \| \, \| B \| \, \| CF \| + \| A \| \, \| F \| \, \| B-E \| + \| A - D \| \, \| E \| \, \| F \|$, we have just shown that:

32) Under H_o, $\| \sqrt{n}(R\hat{\alpha}_n - r)[R(\frac{Z^TZ}{n})^{-1} \hat{V}_n^b(\frac{Z^TZ}{n})^{-1}R^T]^{-1}\sqrt{n}(R\hat{\alpha}_n - r)^T$

$$- \sqrt{n}(R\hat{\Gamma}_n - r)[R(\frac{Z^TZ}{n}) \hat{V}_n^c(\frac{Z^TZ}{n})^{-1}R^T]^{-1} \sqrt{n}(R\hat{\Gamma}_n - r)^T \| = 0$$

and so part iii follows.

[X]

Thus we have shown that we have a valid asymptotic test for testing linear restrictions on the parameters in the variance model. Our next and last result is to show that under additional assumptions, we can use our estimates from the variance model to reestimate the original model and obtain an estimator that is asymptotically equivalent to weighted least squares with variances known.

A4) There exists $\lambda > 0$ such that for all i $\sigma_i^2 > \lambda$.

A5) For all i, $E(\varepsilon_i | W_{is}, \ldots W_{ip}) = 0, E(\varepsilon_i^2 | W_{i1}, \ldots W_{ip}) = Z_i\Gamma_o$

A6) There exists $M < \infty$, such that for all i $\| Z_i \| < M$.

THEOREM 5.7.

Let Ω_n denote the n×n diagonal matrix whose (i,i) entry is σ_i^2 and let $\hat{\Omega}_n$ denote the n n matrix whose (i,i) entry is $Z_i\hat{\Gamma}_n$. Let $\hat{B}_n$ denote the Aitken estimator given by

$$B_n = \begin{cases} (X^T\Omega_n^{-1}X)^{-1}X^T\Omega_n^{-1}Y & \text{where } X^T\Omega_n^{-1}X \text{ is nonsingular} \\ 0 & \text{if } X^T\Omega_n^{-1}X \text{ is singular.} \end{cases}$$

Let $\hat{B}_n$ be the weighted least squares estimator given by

$$\hat{B}_n = \begin{cases} (X^T\hat{\Omega}_n^{-1})^{-1}(X^T\hat{\Omega}_n^{-1}Y) & \text{if } X^T\hat{\Omega}_n^{-1}X \text{ is nonsingular} \\ 0 & \text{if } X^T\hat{\Omega}_n^{-1}X \text{ is singular.} \end{cases}$$

Under assumption A1-A6.

i) $\|\hat{B}_n - \beta_o\| \to 0$ a.e.

ii) plim $\sqrt{n}(\hat{B}_n - B_n) = 0$ and

$$\sqrt{n}(n^{-1}X^T\hat{\Omega}_n^{-1}X)^{\frac{1}{2}}(\hat{B}_n-\beta_o) \overset{A}{\sim} N(0,I_k)$$

iii) If R is a q×k matrix of real numbers with full row rank and r is a q×1 vector of real numbers, then under $H_0: R\beta_0 = r$, $n(R\hat{B}_n-r)^T[R(\frac{X^T\hat{\Omega}_n^{-1}X}{n})^{-1}R^T]^{-1}(RB_n-r) \overset{A}{\sim} \chi_q^2$.

Proof:

1) Consider the transformed model

$$\frac{Y_i}{\sigma_i} = \frac{X_i}{\sigma_i}\beta_o + \frac{\varepsilon_i}{\sigma_i} \qquad i=1,\ldots n \text{ and } \sigma_i = \sqrt{\sigma_i^2}$$

a) $$E(\frac{X_{ij}\varepsilon_i}{\sigma_i^2}) = \int \frac{X_{ij}\varepsilon_i}{\sigma_i^2}\, dP = \int \frac{X_{ij}\varepsilon_i}{Z_i\Gamma_o}\, dP$$

$$\int E(\varepsilon_i|w_{i1}\ldots w_{ip})\frac{X_{ij}}{Z_i\Gamma_o}\, dP = 0 \qquad 1 \leq j \leq k$$

b) $$E(\frac{\varepsilon_i}{\sigma_i}) = E(\frac{\varepsilon_i}{Z_i\Gamma_o}) = \int \frac{1}{Z_i\Gamma_o} E(\varepsilon_i|w_{i1}\ldots w_{ip})dP = 0$$

c) $$E(\frac{\varepsilon_i^2}{\sigma_i^2}) = \int \frac{1}{Z_i\Gamma_o} E(\varepsilon_i^2|w_{i1}\ldots w_{ip})dP = \int \frac{1}{Z_i\Gamma_o} Z_i\Gamma_o\, dP$$

2) The OLS estimator on the transformed model is B_n. Since assumptions A4 and A6 insure that $0 < \lambda \leq \sigma_i^{\,2} \leq M\| \Gamma_0 \| < \infty$ for all i, so all the assumption of White's theorem are satisfied and we conclude

a) $\| B_n - \beta_o \| \to 0$ a.e.

b) $\sqrt{n}(n^{-1}X^T\Omega_n^{\,-1}X)^{\frac{1}{2}}(B_n-\beta_o) \overset{A}{\sim} N(0, I_k)$.

c) under the hypothesis H_o:

$$n(RB_n-r)^T[R(X^T\Omega_n^{\,-1}X)^{-1}R^T]^{-1}(RB_n-r) \overset{A}{\sim} \chi^2_q$$

3) In the case $X^T\Omega_n^{\,-1}X$ and $X^T\hat{\Omega}_n^{\,-1}X$ are both nonsingular,

$$B_n-\hat{B}_n = (X^T\Omega_n^{\,-1}X)^{-1}X^T\Omega_n^{\,-1}(X\beta_o+\varepsilon) - (X^T\hat{\Omega}_n^{\,-1}X)^{-1}X^T\hat{\Omega}_n^{\,-1}$$

$$(X\beta_0+\varepsilon) = (n^{-1}X^T\Omega_n^{-1}X)^{-1}(n^{-1}X^T\Omega_n^{-1}\varepsilon) - (n^{-1}X^T\hat{\Omega}_n^{-1}X)^{-1}(n^{-1}X^T\hat{\Omega}_n^{-1}\varepsilon)$$

4) $n^{-1}X^T\Omega_n^{\,-1}X - n^{-1}X^T\hat{\Omega}_n^{\,-1}X =$

$$n^{-1}\Sigma_{i=1}^{n} \; x_i^{\,T}x_i\left(\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2}\right) \quad \text{where } \hat{\sigma}_{in}^{\;2} = z_i\hat{\Gamma}_n$$

$$= n^{-1} \Sigma_{i=1}^{n} \; x_i^{\,T}x_i \; \frac{1}{\sigma_i^2}\,\frac{1}{\hat{\sigma}_{in}^{\;2}}\,(\hat{\sigma}_{in}^{\;2} - \hat{\sigma}_i^2).$$

5) $X^T\Omega_n^{\,-1}X - X^T\Omega_n^{\,-1}X$ is a k×k matrix where k is a fixed finite number. To show $\| X^T\Omega_n^{\,-1}X - X^T\hat{\Omega}_n^{\,-1} \| \to 0$ a.e., it suffices to show convergence for each matrix entry.

6) $\| n^{-1} \Sigma X_{ij} X_{i\ell} (\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2}) \| \leq$

$\max\limits_{1\leq i\leq n} | \frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2} | n^{-1} \Sigma_{i=1}^{n} |X_{in} X_{i\ell}|$

7) $| \frac{1}{\sigma_i^2} \quad \frac{1}{\hat{\sigma}_{in}^2} | = | \frac{1}{(Z_i \Gamma_o)} \quad \frac{1}{Z_i \hat{\Gamma}_n} \quad Z_i (\Gamma_n - \hat{\Gamma}_o) |$

$\| Z_i \hat{\Gamma}_n \| + \| Z_i (\Gamma_o - \hat{\Gamma}_n) \| \geq \| Z_i \Gamma_o \|$ so

$| Z_i \hat{\Gamma}_n | = \| Z_i \hat{\Gamma}_n \| \geq \| Z_i \Gamma_o \| - \| Z_i (\Gamma_o - \hat{\Gamma}_n) \|$ and

$| Z_i \hat{\Gamma}_n | \geq \lambda - M \| \Gamma_o - \hat{\Gamma}_n \|$ and hence

$| Z_i \hat{\Gamma}_n | > \frac{\lambda}{2}$ a.e. $(n \to \infty)$ and

$\| Z_i (\hat{\Gamma}_n - \Gamma_o) \| \leq M \| \hat{\Gamma}_n - \Gamma_o \| \to 0$ a.e. $(n \to \infty)$.

Thus $\sup\limits_{1\leq i\leq n} | \frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2} | \to 0$ a.e. $(n \to \infty)$.

Therefore by proposition 2 and assumption A2', for $1 \leq j$, $\ell \leq m$,

$\| n^{-1} \Sigma_{1=1}^{n} X_{ij} X_{i\ell} (\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2}) \| \to 0$ a.e. $(n \to \infty)$ and hence

$\| n^{-1} (X^T \Omega_n^{-1} X) - n^{-1} (X^T \hat{\Omega}_n^{-1} X) \| \to 0$ a.e. $(n \to \infty)$.

(8)

Let $Z \in \mathbb{R}^k$; $\langle n^{-1}(X^T\Omega_n^{-1}X)Z, Z\rangle = n^{-1}\langle\Omega_n^{-1}XZ, XZ\rangle$

$$\geq n^{-1} \frac{1}{M\|\Gamma_o\|} \|XZ\|^2 \geq n^{-1} \frac{1}{M\|\Gamma_o\|} \langle XZ, XZ\rangle$$

$$\geq \frac{1}{M\|\Gamma_o\|} \langle n^{-1}X^TXZ, Z\rangle$$

$$\frac{1}{M\|\Gamma_o\|} \langle n^{-1}X^TXZ, Z\rangle \geq \frac{1}{M\|\Gamma_o\|} \frac{1}{\|[n^{-1}(X^TX]^{-1}\|} \|Z\|^2$$

(providing these exist).

For n sufficiently large $\|[E(n^{-1}X^TX)]^{-1}\| \leq \frac{1}{\lambda}$

$$\langle n^{-1}(X^T\Omega_n^{-1}X)Z, Z\rangle \geq \frac{1}{\lambda}\frac{1}{2}\frac{1}{M\|\Gamma_o\|} \|Z\|^2 \quad \text{a.e. } (n \to \infty).$$

and thus

$$\|[n^{-1}(X^T\Omega_n^{-1}X)]^{-1}\| \leq \frac{2\lambda}{M\|\Gamma_o\|} \quad \text{a.e.} \quad (n \to \infty)$$

and $\|[n^{-1}(X^T\Omega_n^{-1}X)]^{-1} - [n^{-1}(X^T\hat{\Omega}_n^{-1}X)]^{-1}\| \to 0$ a.e. $(n \to \infty)$

(9)

$$n^{-1}[(X^T\Omega_n^{-1}\varepsilon - X^T\hat{\Omega}_n^{-1}\varepsilon)] = n^{-1} \Sigma\, X_i^T\varepsilon_i\left(\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2}\right)$$

$n^{-1}\ [X^T\Omega_n^{-1}\varepsilon - X^T\hat{\Omega}_n^{-1}\ \varepsilon]$ is a k×1 vector so again we need only show coordinate convergence to zero.

$$|n^{-1}\ \Sigma_{i=1}^{n}X_{ij}\varepsilon_i\left(\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2}\right)| \leq \underset{1\leq i\leq n}{\text{MAX}}\ \left|\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2}\right|\ |n^{-1}\ \Sigma_{i=1}^{n}|X_{ij}\varepsilon_i|.$$

Therefore we conclude $\| n^{-1}(X^T\Omega_n^{-1}\varepsilon - X^T\hat{\Omega}_n^{-1}\varepsilon\| \to 0$ a.e. $(n\to\infty)$.

$| n^{-1} \Sigma_{i=1}^{n} x_{ij}\varepsilon_i \frac{1}{\sigma_i^2}| \leq \frac{1}{\lambda} n^{-1}\Sigma_{i=1}^{n} | x_{ij}\varepsilon_i |$, therefore

there exists $C < \infty$ such that

$\| n^{-1}X^T\Omega_n^{-1}\varepsilon \| < C$ a.e. $(n \to \infty)$.

10) Hence $\| \hat{B}_n - B_n \| \to 0$ a.e. $(n \to \infty)$

and $\| \hat{B}_n - \beta_o \| \to 0$ a.e. $(n \to \infty)$ (part i)

{Recall $\| UV-XY \| \leq \| U \| \| V-Y \| + \| U-X \| (\| V \| + \| Y-V \|)$ }.

11) $\sqrt{n}[n^{-1}X^T(\Omega_n^{-1} - \hat{\Omega}_n^{-1})\varepsilon]$ is a $k\times 1$ vector, to show

plim $\sqrt{n}\ [n^{-1}X^T(\Omega_n^{-1} - \hat{\Omega}_n^{-1})\varepsilon] = 0$, it suffices to show

that this holds by coordinate.

$$\sqrt{n}\ [n^{-1} \Sigma_{i=1}^{n} x_{ij}\varepsilon_i (\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2})] =$$

$$n^{-1} \Sigma_i \frac{x_{ij}\varepsilon_i}{\sigma_i^2\sigma_{in}^2} \sqrt{n}\ z_i[\hat{\Gamma}_n - \Gamma_o] =$$

$$\Sigma_{r=1}^{m}(n^{-1} \Sigma_{i=1}^{n} \frac{x_{ij}\varepsilon_i z_{ir}}{\sigma_i^2\hat{\sigma}_{in}^2}) \sqrt{n}[\hat{\Gamma}_n - \Gamma_o]_r$$

12) We have already seen in the proof of theorem 2, that for any $\varepsilon < 0$, there exists $M(\varepsilon) < \infty$ and $N(\varepsilon) < \infty$ such that $n \geq N(\varepsilon)$ implies prob. $\{\sqrt{n}\| \hat{\Gamma}_n - \Gamma_0 \| \leq M(\varepsilon)\} \geq 1-\varepsilon$.

We now must show that

$$n^{-1}\Sigma_{i=1}^{n} \frac{X_{ij}\varepsilon_i Z_{ir}}{\sigma_i^2\ \hat{\sigma}_{in}^2} \to 0 \quad \text{a.e.} \ 1 \leqslant j \leqslant k \quad 1 \leqslant r \leqslant m \quad (n \to \infty)$$

13) $E(|\frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^4}|^{1+\delta}) < \Delta \cdot \frac{1}{\lambda^4}$ (where δ, Δ are as in assumption

A2. Therefore by proposition 1,

$$|n^{-1}\ \Sigma_{i=1}^{n} \frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^4} - n^{-1}\ \Sigma_{i=1}^{n} E(\frac{\varepsilon_i X_{ij} Z_{ir}}{(Z_i \Gamma_o)^4})| \to 0 \quad \text{a.e.}$$

$(n \to \infty)$ and hence

$$|n^{-1}\ \Sigma_{i=1}^{n} \frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^2 \sigma_i^2}| \to 0 \quad \text{a.e.}\ (n \to \infty).$$

14) $|n^{-1}\ \Sigma_{i=1}^{n} \frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^2 \hat{\sigma}_{in}^2}| \leqslant |n^{-1}\ \Sigma_{i=1}^{n} \frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^2 \hat{\sigma}_i^2}| +$

$$+ |n^{-1}\ \Sigma_{i=1}^{n} \frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^2} (\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2})|$$

$$\leqslant |n^{-1}\Sigma_{i=1}^{n} \frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^4}| + \underset{1 \leqslant i \leqslant n}{\text{MAX}} |\frac{1}{\sigma_i^2} - \frac{1}{\hat{\sigma}_{in}^2}|\ |n^{-1}\Sigma_{i=1}^{n}| \frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^2}|$$

So $\quad n^{-1}\Sigma_{i=1}^{n} \frac{\varepsilon_i X_{ij} Z_{ir}}{\sigma_i^2 \hat{\sigma}_{in}^2} \to 0$ a.e. $(n \to \infty)$

and $\underset{n \to \infty}{\text{plim}} \ \| \sqrt{n}[n^{-1}(X^T \Omega_n^{-1} \varepsilon) - n^{-1}(X^T \hat{\Omega}_n^{-1} \varepsilon)] \| = 0.$

15) $\| \sqrt{n}(n^{-1}X^T\Omega_n^{-1}X)^{\frac{1}{2}}\ (B_n - \beta_o) - \sqrt{n}(n^{-1}X^T\hat{\Omega}_n^{-1}X)^{\frac{1}{2}}(B_n - \beta_o) \| \leqslant$

$$\| \sqrt{n}(B_n - \beta_o)\| \cdot \|(n^{-1}X^T\Omega_n^{-1}X)^{\frac{1}{2}} - (n^{-1}X^T\hat{\Omega}_n^{-1}X)^{\frac{1}{2}}\| +$$

$$\| \sqrt{n}(B_n - \hat{B}_n) \| \quad (\| (n^{-1}X^T\Omega_n^{-1}X)^{\frac{1}{2}} - (n^{-1}X^T\hat{\Omega}_n^{-1}X)^{\frac{1}{2}} \|$$

$$+ \| (n^{-1}X^T\Omega_n^{-1}X)^{\frac{1}{2}} \|)$$

therefore $\operatorname*{plim}_{n\to\infty} \| \sqrt{n}(n^{-1}X^T\Omega_n^{-1}X)^{\frac{1}{2}}(B_n-\beta_o) -$
$\sqrt{n}(n^{-1}X^T\Omega_n^{-1}X)^{\frac{1}{2}}(\hat{B}_n-\beta_0) \| = 0$ and part ii follows.

16) $\| \sqrt{n}(R\hat{B}_n - r) - \sqrt{n}(RB_n - r) \| \leq \| R \| \, \| \sqrt{n}(\hat{B}_n - B_n) \|$

so $\operatorname*{plim}_{n\to\infty} \| \sqrt{n}(R\hat{B}_n - r) - \sqrt{n}(RB_n - r) \| = 0$ and

$\operatorname*{plim}_{n\to\infty} \| \sqrt{n}(R\hat{B}_n - r)^T - \sqrt{n}(RB_n - r)^T \| = 0$

Under the hypothesis H_o: $R\beta_o = r$, for an $\varepsilon > 0$ there exist $M(\varepsilon) < \infty$ $N(\varepsilon) < \infty$ such that $n \geq N(\varepsilon)$ implies

$$\text{prob } \{\| \sqrt{n}(RB_n - r) \| \leq M(\varepsilon)\} \geq 1-\varepsilon$$

17) $\| R(n^{-1}X^T\Omega_n^{-1}X)^{-1}R^T - R(n^{-1}X^T\hat{\Omega}_n^{-1}X)^{-1}R^T \| \leq$

$\| R \|^2 \, \| (n^{-1}X^T\Omega_n^{-1}X)^{-1} - (n^{-1}X^T\hat{\Omega}_n^{-1}X)^{-1} \|.$

$\| (n^{-1}X^T\Omega_n^{-1}X)^{-1} - (n^{-1}X^T\hat{\Omega}_n^{-1}X)^{-1} \| \to 0$ a.e. $(n \to \infty)$

(from 8), Therefore

$\| [R(n^{-1}X^T\Omega_n^{-1}X)^{-1}R^T] - [R(n^{-1}X^T\hat{\Omega}_n^{-1}X)^{-1}R^T] \| \to 0$ a.e.
$(n \to \infty)$

18) R^T is a m×q matrix of full column rank, therefore there

exist $\delta > 0$ such that $\| R^T Z \| > \delta \| Z \|$ for all $Z \in \mathbb{R}^m$

$\langle R(n^{-1}X^T\Omega_n^{-1}X)R^T Z, Z\rangle = \langle (n^{-1}X^T\Omega_n^{-1}X)^{-1}R^T Z, R^T Z\rangle$

$$\geq \frac{1}{\| n^{-1}X^T\Omega_n^{-1}X \|} \delta^2 \|Z\|^2;$$

The (s,t) component of $n^{-1}X^T\Omega_n^{-1}X = n^{-1}\Sigma_{i=1}^{n} \dfrac{X_{is}X_{it}}{\sigma_n^2}$

so there exists $C < \infty$ such that

$\| n^{-1}X^T\Omega_n^{-1}X \| < C$ a.e. $(n \to \infty)$.

Therefore $\| [R(n^{-1}X^T\Omega_n^{-1}X)^{-1}R^T]^{-1} \| < \dfrac{2C}{\delta^2}$ a.e. $(n \to \infty)$

and

$$\| [R(n^{-1}X^T\hat{\Omega}_n^{-1}X)^{-1}R^T]^{-1} - [R(n^{-1}X^T\hat{\Omega}_n^{-1}X)^{-1}R^T] \| \to 0 \quad \text{a.e.} \quad (n \to \infty).$$

We can therefore conclude that under H_o:

19) $$\operatorname*{plim}_{n \to \infty} \| n(RB_n-r)^T[R(n^{-1}X^T\Omega_n^{-1}X)^{-1}R^T]^{-1}(RB_n-r) - n(R\hat{B}_n-r)[R(n^{-1}X^T\hat{\Omega}_n^{-1}X)^{-1}R^T](R\hat{B}_n-r) \| = 0$$

and so by proposition 3, part iii holds.

[X]

CHAPTER V

SUMMARY AND CONCLUSIONS

The state of the nation's housing program, particularly those for the urban poor, should convince even the skeptic that we do not have a firm understanding of the housing market. This market is unique among economic markets. Patterns of residential development have a profound effect upon the social and economic development of the family, the municipality, and the nation as a whole. Urban economics and planning professions must seek to develop new theoretical and empirical procedures to help us analyze the housing market.

In the preceding chapters, I have presented new theoretical procedures that should be of interest to the urban economist and policy developer. It is not meant to be nor is it a finished tome. Rather it points to new directions for future work. In this last chapter, I'll briefly discuss directions that future research might take.

The search model presented in the first essay is not completely analyzed. It would be of great interest to build into it a mechanism for replacing buyers and sellers and then examining the implications over time for an individual buyer and seller resulting from market aggregation. Ultimately one desires to have a search model where bid structure, sellers behavior, and search strategy are all endogenous. One would then be able to examine the time paths of the buyers and sellers behavior as well as that of the market as a whole. It would be advantageous to be able to determine the effects on this model from altering the distribution of incomes of buyers and from adding more buyers and sellers from different income groups without having to replicate each agent. If we had this type of model, we would be in a much better position for understanding effects of discrimination on market prices of housing as well as in neighborhood residential patterns.

While White's work and that in the third essay may seem to answer the questions of estimation and hypothesis testing in the presence of heteroskedasticity, this is far from the fact. The results that we have derived, and those in the second essay also, yield asymptotic properties. What is clearly needed are small sample properties of the various estimates and statistics that are designed to deal with heteroskedasticity. As computers become more powerful and more available, maximum likelihood procedures become more accessible. It would be of great value to know the circumstances under which each of these estimators dominates in the small (finite) sample case.

It is clear that urban policies and programs have little chance of success until their designers gain a better grasp on the behavior of the urban housing market. While much theoretical and empirical research remains to be done, there have been large gains in developing the theoretical and technical tools necessary for effective analysis of the housing market. However, the gains made in the development of these theories and techniques will be insubstantial unless the policy and program planners develop their technical skills and mathematical maturity sufficient to understand and utilize the theories.

BIBLIOGRAPHY

1. Alonzo, W. Location and Land Use. Harvard University Press, Cambirdge, Mass., 1970.

2. Arrow, Kenneth, and F. H. Hahn, General Competitive Analysis, Holden Day, Inc., San Francisco, 1971.

3. Ash, Robert B. Real Analysis and Probability. Academic Press, New York, N.Y., 1972.

4. Box, G.E.P., "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance," Annals of Mathematical Statistics, 1954.

5. Brown, A., and Pearcy, C. Introduction to Operator Theory, I: Elements of Functional Analysis. New York, Springer-Verlag, 1977.

6. __________. Introduction to Operator Theory II. (In preparation).

7. Butters, G. "Equilibrium Distribution of Sales and Advertising Prices." Review Economic Studies. October 1977, vol. 44, 465-491.

8. Chow, Y.S., Robbins, H., and Siegmund, D. Great Expectations: The Theory of Optimal Stopping. Houghton Mifflin Co., Boston, Mass., 1971.

9. Chung, Kai Lai. A Course in Probability Theory. Academic Press, New York, N.Y., 1974.

10. Diamond, P. "Wage Determination in Search Equilibrium," Working Paper 253, Department of Economics, M.I.T., Cambridge, Mass., 1980.

11. Diamond, P.A. and Stiglitz. "Increases in Risk and Risk Aversion." Journal of Economic Theory, 1974, vol. 8, 337-360.

12. Downs, George W., and Rocke, David M., "Interpreting Heteroskedasticity," American Journal of Political Science, vol. 23, no. 4, November 1979 pp. 816-828.

13. Feller, William. An Introduction to Probability Theory and Its Applications. Volumes I and II, John Wiley and Sons, New York, N.Y., 1957.

14. Glejser, H. "A New Test for Heteroskedasticity." Journal of the American Statistical Association, 64 (1969), 316-323.

15. Goldfeld, S.M., and Quandt, R.E. "Some Tests for Homoskedasticity." Journal of the American Statisical Association, 60 (1965), 539-559.

16. Halmos, P.R. A Hilbert Space Problem Book. New York, Van Nostrand, 1967.

17. Ito, K. and Schull, W.J. "On the Robustness of the T^2 Test in Multivariate Analysis of Variance When Variance-Covariance Matrices are not Equal." Biometrika, vol. 51, June 1964.

18. Kmenta, Jon. Elements of Econometrics. 1971, Macmillian Publishing Co., Inc., New York, N.Y.

19. Kohn, Meir G. and Shavell, Steven. "The Theory of Search." Journal of Economic Theory, 1974, vol. 9, 93-123.

20. Kosters, Marvin and Welch, Finis. "The Effects of Minimum Wages by Race, Age, and Sex." in Anthony H. Pascal, ed. Racial Discrimination in Economic Life, D.C. Heath, Lexington, Mass., 1972.

21. Lippman, S.A., and McCall, J.J. "The Economics of Job Search: A Survey, Part I: Optimal Job Search Policies." Economic Inquiry, June 1976, 155-189.

22. Lippman, S.A., and McCall, J.J. "The Economics of Job Search: A Survey, Part II: Empirical and Policy Implications of Job Search." Economic Inquiry, September 1976, pp. 347-368.

23. Lucas, R.E. and Prescott. "Equilibrium Search and Unemployment." Journal of Economic Theory, 1974, vol. 7, 188-209.

24. Malinvaud, E., Statistical Methods of Econometrics, 2nd. Edition, English translation by Mrs. A. Silvey, 1970, North-Holland Publishing Co., Amsterdam.

25. McCall, John J. Income Mobility, Racial Discrimination and Economic Growth. D. C. Heath and Co., Lexington, Mass., 1973.

26. Mieszkowski, Peter, and Straszheim, Mahlon, Current Issues in Urban Economics, 1979, Johns

27. Mills, Edwin S., Studies in the Structure of the Urban Economy, Resources for the Future, Inc., Johns Hopkins Press, Baltimore, MD., 1972.

28. Mirman, L.J. and Porter, W.R. "A Microeconomic Model of the Labor Market Under Uncertainty." Economic Inquiry, June 1974, vol. 12, 135-145.

29. Montgomery, Roger, and Marshall, Dale Rogers, Housing Policy for the 1980's, Lexington Books, D.C. Heath and Company, Lexington, Mass., 1980.

30. Mortensen, Dale T. "Job Search, the Duration of Unemployment and the Phillips Curve." American Economic Review, December 1970, 847-862.

31. Muth, Richard, Cities and Housing, U. of Chicago Press, Chicago, IL., 1969.

32. Norse, Hugh O., The Effect of Public Housing on Housing Markets, Lexington Books, D.C. Heath and Company, Lexington, MA., 1973.

33. Oberhofer, W. and Kmenta, J., "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models." Econometrica, vol. 42, no. 3, May 1974.

34. Park, R.E. "Estimation with Heteroskedastic Error Terms." Econometrica, vol. 34, no. 4 (1966), p. 888.

35. Polinsky, A. Mitchell, and Elwood, David T., "An Empirical Reconciliation of Micro and Grouped Estimates of the Demand for Housing," The Review of Economics and Statistics, LXI, #2, May, 1979.

36. Prais, S.J., and Houthakker, H.S. The Analysis of Family Budgets, The University Press, 1955, Cambridge, England.

37. Rao, C.R. Linear Statistical Inference and Its Applications. New York, John Wiley and Sons, 1973.

38. Reid, Margaret, Housing and Income, University of Chicago Press, Chicago, IL., 1962.

39. Rothchild, M. and Stiglitz, J. Increasing Risk: I. A Definition." Journal of Economic Theory, 1970, pp. 225-243.

40. Scheffe, H. The Analysis of Variance, 1959 John Wiley, New York, N.Y.

41. Smith, Barton, and Campbell, J.M., Jr., "Aggregation Bias and the Demand for Housing," International Economic Review, 19 June 1978, 495-505.

42. Smith, Wallace F., Housing, University of California Press, Berkeley, CA., 1971.

43. Solomon, Arthur P., Housing the Urban Poor, A publication of the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, 1974.

44. Stigler, G.J. "The Economics of Information." Journal of Political Economy, June 1961, 213-225.

45. __________. Information in the Labor Market." Journal of Political Economy, October 1962, 94-104.

46. Straszheim, Mahlon R., An Econometric Analysis of the Urban Housing Market. National Bureau of Economic Research, Columbia University Press, New York, N.Y., 1975.

47. Telser, Lester G. "Searching for the Lowest Price." American Economic Review, May 1973, 40-49.

48. Thiel, Henri. Principles of Econometrics, 1971 John Wiley and Sons, Inc., New York, N.Y.

49. Tiebout, Charles M., "A Pure Theory of Local Expenditures," Journal of Political Economy, 64, October, 1956, pp. 416-424.

50. U.S. Bureau of the Census, Statistical Abstract of the United States: 1979, 100th edition, Washington, D.C., 1979.

51. United States Commission on Civil Rights, Twenty Years After Brown: Equal Opportunities in Housing, U.S. Commission on Civil Rights, December, 1975.

52. Varian, Hal R. "A Model of Sales." The American Economic Review, September 1980, vol. 70, #4.

53. Weicher, John C. Housing, American Enterprise Institute for Public Policy Research, Washington, D.C., 1980.

54. Welch, B.L. "The Significance of the Difference Between Two Means when the Population Variances are Unequal." Biometrika, 1937.

55. Wendt, Paul F. "The Role of the Federal Government in Housing," No. 460 in the series "National Economic Problems," American Enterprise Association, Inc., Washington, D.C., 1956.

56. Wheaton, W.C. "A Bid Rent Approach to Housing Demand." Journal of Urban Economics, 1977.

57. __________. "Income and Urban Residence: An Analysis of Consumer Demand for Location." The American Economic Review, vol. 67, no. 4, September 1977, pp. 620-631.

58. __________, Milgram, Grace, and Meyerson, Margy Ellin, Urban Housing, The Free Press, MacMillan Company, New York, N.Y., 1966.

59. White, Halbert. "A Heteroskedasticity Consistent Convariance Matrix Estimator and a Direct Test for Heteroskedasticity." Econometrica, vol. 48, 1980, pp. 817-838.

60. Wohlstetter, Albert and Coleman, Sinclair. "Race Difference in Income." in Anthony Pascal, ed. Racial Discrimination in Economic Life, D.C. Heath, Lexington, Mass., 1972.

61. Yinger, John. "A Search Model of Real Estate Broker Behavior." The American Economic Review, September 1981, pp. 591-605.

62. Anas, Alex, "A Model of Residential Change and Neighborhood Typing" JUE 7, 1980, pp. 358-370.

63. __________, "A Probabilitic Approach to the Structure of Rental Housing Markets" JUE 7, 1980, pp. 225-247.

64. Bailey, M.J., "Effects of Race and Other Demographic Factors on the Values of Single Family Homes," Land Economics 42, 1966, pp. 215-220.

65. Burstein, Nancy, "Voluntary Income Clustering and the Demand Housing and Local Public Goods" JUE 7, 1980, pp. 175-185.

66. Carliner, Geoffrey, "Income Elasticity of Housing Demand" RES 4, 1973, pp. 528-532.

67. Chinloy, P., "The Estimation of Net Depreciation Rates on Housing" JUE 6, 1979, pp. 432-443.

68. Courant, P.N. and J. Yinger, "On Models of Racial Prejudice and Urban Residential Structure," JUE 4, 1977, pp. 272-291.

69. Dusansky, R., M. Ingber, and N. Karatjas, "The Impact of Property Taxation on Housing Value and Rents, JUE 10, 1981, pp. 240-255.

70. Dusenberry, J.S. and H. Kristin, "The Role of Demand in the Economic Structure" Studies in the American Economy, edited by W. Leontief, Oxford U. Press, N.Y., 1953, pp. 451-482.

71. Erekson, O. H., and A. D. Witte, "The Demand for Housing Comment" Southern Economic Journal, 46, 1979, pp. 641-648.

72. Fare, R. and J. Y. Bong, "Variable Elasticity of Substitution in Urban Housing Production" JUE 10, 1981, pp. 369-374.

73. Farley, R. S. Bianchi, and D. Colasanto, "Barriers to the Racial Integration of Neighborhood: The Detroit Case." The Annals of the American Academy of Political and Social Science 44, 1979, pp. 97-113.

74. Follain, J. R. and S. Malprezzi, "Another Look at Racial Differences in Housing Prices," Urban Studies 18, 1981, pp. 195-203.

75. Grieson, R. and J. White, The Effects of Zoning on Structure and Land Markets," JUE 10, 1981, pp. 271-285.

76. Herbert J. H., " Model Estimation and Verification--Some Recent Approaches." Empirical Economies 4, 1979, pp. 87-99.

77. Hirsch, W. Z., and Cheung-Kwork Law, "Habitability Laws and the Shrinkage of Substandard Rental Housing Stock." Urban Studies 76, 1979, pp. 19-28.

78. Kain, J. F., and J. M. Quigley, "Housing Markets and Racial Discrimination," National Bureau of Economic Research, New York, 1975.

79. Kaluzny, R. L., "Changes in the Consumption of Housing Services: The Gary Experiment" The Journal of Human Resources, 14, 1979, pp. 496-506.

80. Kau, J. B. and Sirmans, C. F., "Urban Land Value Functions and the Price Elasticity of Demand for Housing," JUE 6, 1979, pp. 112-121.

81. Kern, Clifford R., "Racial Prejudices and Residential Segregation: The Ginger-Model Revisited." JUE 10, 1981, pp. 164-172.

82. King, A. T., "The Demand for Housing: A Lancastrian Approach" Southern Economic Journal 43, 1976, pp. 1077-1087.

83. Kraft, J., and A. Kraft, "Benefit and Cost of Low Rent Public Housing" Journal of Regional Science, 19, 1979, pp. 309-317.

84. Lee, T. H., "Housing and Permanent Income: Tests Based on a Three Year Reinterview Survey," RES 50, 1968, pp. 480-490.

85. deLeeuw, Frank, "The Demand for Housing: A Review of Cross-Section Evidence" Econometrics 53, 1971, pp. 1-7.

86. MacMinn, R., "Search and Market Equilibrium," Journal of Political Economy, 88(2), 1980, pp. 308-327.

87. MacRae, C. D., and M. A. Turner, "Estimating Demand for Owner Occupied Housing Subject to the Income Tax" JUE 10, 1981, pp. 338-356.

88. Maisel, S. J., Burnham, and John Austin, "The Demand for Housing: A Comment" RES 53, 1971, pp. 410-413.

89. Miesykowski, P., and R. Syron, "Economic Explanations for Housing Segregation," New England Economic Review, Federal Reserve Bank, Boston, Mass., Nov/Dec 1979, pp. 33-39.

90. Mills, D., "Segregation, Rationing and Zoning," Southern Economic Journal 45(4), 1979, pp. 1195-1207.

91. Murray, M., "Tenant Benefits in Alternative Federal Housing Programmes" Urban Studies 17, 1980, pp. 25-34.

92. Myers, S. J. and K. E. Phillips, "Housing Segregation and Black Employment: Another Look at the Ghetto Dispersal Strategy," AER 69, 1979, pp. 298-302..

93. Phillips, R. S. "A Note on the Determinants of Residential Succession," JUE 9, 1981, pp. 49-55.

94. Polinsky, M., "The Demand for Housing: A Study in Specification and Grouping." Econometrica 45, 1977, pp. 447-461.

95. Power, T., "Urban Size (Dis)amenities Revisited," JUE 9, 1981, pp. 85-89.

96. Reschovsky, A., "Residential Choice and the Local Public Sector: An Alternative Test of the Tiebout Hypothesis" JUE 6, 1979, pp. 501-520.

97. Rose-Ackerman, S. "Racism and Urban Structure," JUE 2, 1975, pp. 85-103.

98. Rosen, H. S., "Estimation Inter-City Differences in the Price of Housing Services." Urban Studies, 15, 1978, pp. 351-355.

99. Rydell, C. P., "Supply Response to the Housing Allowance Program." International Regional Science Review 5, 1980, pp. 119-138.

100. Schelling, T. C., "Neighborhood Typing," in Racial Discrimination in Economic Life, Heath, Lexington, Mass., 1972.

101. Sirmans, C. F., J. Kau, and C. F. Lee, "The Elasticity of Substitution in Urban Housing Production: A YES Approach." JUE 6, 1979, pp. 407-415.

102. Stegman, Michael A. and H. J. Sumka, "Income Elasticities of Demand for Rental Housing in Small Cities" Urban Studies 15, 1978, pp. 51-61.

103. Stevens, Barbara, "Employment Permanent Income, and the Demand for Housing," JUE 6, 1979, pp. 480-500.

104. Vickerman, R. W., "The Evaluation of Urban Change: Equilibrium and Adaptive Approaches," Urban Studies 16, 1979, pp. 81-93.

105. Walden, M. L. "A Note on Benefit and Cost Estimates in Public Assisted Housing" _Journal of Regional Sciences_, 21, 1981, pp. 421-423.

106. Westkoff, F., "Policy Inference in Community Choice Models: A Caution" JUE 6, 1979, pp. 535-549.

107. Winger, A. R., "Housing and Income," _Western Economic Journal_ 6, 1968, pp. 226-232.

108. Yinger, J., "Racial Preference and Residential Segregation in an Urban Model, JUE 3, 1976, pp. 383-396.

109. Yinger, John and Sheldon Danyiger, "An Equilibrium Model of Urban Population and the Distribution of Income." _Urban Studies_ 15, 1978, pp. 201-214.