AD-A047 673 WHARTON ECONOMETRIC FORECASTING ASSOCIATES PHILADELP--ETC F/G 5/9
INTERIM REPORT FOR OFFICE OF NAVAL RESEARCH. VOLUME I,(U)
NOV 77 M WACHTER, R M YOUNG N00014-76-C-0782

UNCLASSIFIED NL

| OF |

AD
A047673

END
DATE
FILMED
1 -78
DDC

WHARTON EFA, INC. WEFA

(10)

B.S.

ADA047673

INTERIM REPORT

FOR

OFFICE OF NAVAL RESEARCH

CONTRACT N00014-76-C-0782

NOVEMBER 1, 1977

Volume I

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Interim Report for Office of Naval Research, C~~ontract N00014-76-C-0782~~ Volumes I, II		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Michael Wachter (Univ. of Pennsylvania) and Richard M. Young (Wharton EFA, Inc.)		8. CONTRACT OR GRANT NUMBER(s) (15) N00014-76-C-0782 New
9. PERFORMING ORGANIZATION NAME AND ADDRESS Wharton Econometric Forecasting Associates, Inc. 4025 Chestnut Street Philadelphia, PA 19104		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of Navy Arlington, VA 22217		12. REPORT DATE (11) 1 November 1, 1977
		13. NUMBER OF PAGES Vol. I - 80 Vol. II - 103
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 83p.		15. SECURITY CLASS. (of this report) unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

unlimited distribution

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

unlimited distribution

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Demographics of Labor Market Behavior

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report documents an investigation on comparative methods of modeling the demographic characteristics of employment and the labor force, based on modeling stocks or net flows. The former is taken from research by the Urban Institute and the latter developed by WEFA and documented in the report. After comparing there characteristics the WEFA model is used to forecast labor markets through the second quarter of 1982.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

5C 392 960 LB

TABLE OF CONTENTS

Chapter 1 - Introduction

Research Goals

This report describes research completed during the first year of a multi-year research project. The desired end product of that research is the ability to project Naval manpower flows and evaluate the impact on those flows of alternate personnel policies.

An earlier interim report (undated) describes the initial stage of the research which was directed toward improving the relationships within the Wharton Quarterly Model used to explain total civilian employment by industry category. The work described in this report has been directed toward development of a capability of explaining the demographic composition of the labor force, employment and unemployment.

Research Strategy

Because of the specific demographic character of military interaction with the labor market, it is necessary to describe in detail the behavior of participation, employment and unemployment by age-sex groups before gross flows in and out of the military can be explained. At similar levels of unemployment for the total population, both enlistments and separations can be expected to vary in response to the distribution of unemployment by age-sex groups. If unemployment rates among young males and males in the 25-34 age group are high, it would normally be expected that enlistments would be higher and separations lower than for a similar aggregate unemployment rate with less unemployment in these age groups.

Our research stretegy for developing a demographic description of the labor force has involved the evaluation of two approaches to modeling labor markets. The first of these involved linking an already existing model of labor markets to the Wharton Quarterly Model. The second approach involved the development of an alternative set of relationships based on a distinctly different approach to describing labor market behavior.

The Urban Institute had in existence at the inception of the contract a monthly model of labor markets which was capable of generating considerable demographic detail. A detailed description of the model appears in Smith, Vanski and Holt, "Regression and the Employment of Demographic Groups", Brookings Papers on Economic Activity, 1974:3 and Smith, "A Simulation Model of the Demographic Composition of Employment, Unemployment and Labor Force Participation: Status Report," Working Paper 350-65 (Urban Institute, 1974). This model, labeled the RASST model (Race-Age-Sex-Search-Turnover), explains employment, unemployment and labor force stocks and flows for 16 race-age-sex groups on a monthly basis. The RASST model is distinguished by the characteristic of focusing on labor market flows rather than stocks. Thus, instead of directly estimating the stock of employed people in a specific demographic group, the number of people entering the employed in that group from either the ranks of the unemployed or those not in the labor force would be estimated and those leaving the employed to either drop out of the labor force or become unemployed would also be estimated; the stock of employed would then be derived as an identity from the lagged stock and the net flow. To quote Smith, "The size of each flow is determined by the size of the stock from which a potential flow can occur and the probability of its occurence. The transition probability is estimated to be a function of the state of the aggregate labor market, seasonal forces, and longer-run changes in behavior." This approach to modeling laobr markets is a divergence from more traditional approaches. The theoretical basis for the model and conceptual foundations can be traced from the work described in Microeconomic Foundations of Employment and Inflation Theory ed. Edmund S. Phelps.

To use the RASST model in conjunction with the Wharton Quarterly Model involves several substantive problems concerning:

(1) the differing levels of aggregation in the time dimension,

(2) developing relationships to explain the variables needed as inputs to the RASST model,

(3) developing compatible solution algorithms for the two models.

The techniques used to resolve these problems are described in Chapter 2.

As indicated above the RASST model differs substantially from prior efforts to model labor market phenomena particularly in its concentration on flow rather than stock phenomena, but also in its use of transitional probabilities to project those flows. In the present version of the model the values of the transitional probability matrix are based solely on aggregate labor market variables. No impact on labor market flows of real wages, retirement or unemployment benefits, school enrollment, training programs, etc. could be captured by the model. While the absence of these variables is probably not a severe problem when the horizon of interest is several months, it could become a severe problem over a period of several years when these structural effects can vary a great deal.

While early indications, such as the results cited by Smith, Vanski and Holt in the Brookings Papers are encouraging, we have simultaneously pursued an alternate line of research along more traditional lines to produce a demographic description of the labor force.

WEFA has developed labor force participation rate and unemployment rate equations by demographic group. We have not pursued explanation of labor force phenomena along racial grounds. The results of this investigation are described in Chapter 3.

Chapter 4 of the report describes the results of historical simulation of these two labor market models and evaluates them on this basis.

Chapter 5 describes the forecasting strategy and results of our initial forecast with the demographic model. The actual forecast appears in Volume II of this report.

Chapter 2 - Linking the RASST Model to the Wharton Model

As noted above the initial problems with respect to linking the RASST model to the Wharton Quarterly Model involve (1) developing explanations of variables which should be endogenous in an economy wide model but are exogenous in the RASST model and (2) dealing with the differing levels of time aggregation of the two models.

(1) The basic exogenous variables neeed for the RASST model are monthly GNP and unfilled orders in constant dollars, and a set of population variables. This latter group can be projected in a variety of ways but since population will be insensitive to economic variables over the five year horizon with which we are currently concerned leaving them exogenous presents no problem. Both GNP and unfilled orders, will be sensitive to economic conditions in the very short run and must be endogenized if the description of labor markets is to accurately reflect the evolving economy. In order to provide the appropriate input variables for the RASST model it was necessary to develop an explanation of aggregate unfilled orders for the Wharton Quarterly Model.

Two approaches were considered for endogenizing unfilled orders:

(1) The natural approach is to model shipments and new orders as the variables which are the fundamental choice elements and derive unfilled orders from the identity

$$\Delta\text{Unfilled Orders} = \text{New Orders} - \text{Shipments}.$$

(2) Alternatively one could attempt to model unfilled orders directly. The appeal of this approach is that since ΔUnfilled Orders is small relative to New Orders and Shipments small errors in either of the latter two relationships can result in large errors in the former. It may be possible to obtain more accuracy by directly modeling unfilled orders.

We have, in fact, pursued both of these lines of investigation.

At the outset it was decided that it would be necessary to deal separately with durable and non-durable manufacturing. This approach was determined by the observation that unfilled orders behave much differently in these two sectors and by the expectation that the determinants of new orders and shipments are likely to be very different in the two industry segments.

The description of the model segment developed is abbreviated since major interest is in the labor market dynamics.

In addition to distinguishing between durables and non-durables, new orders for durables are disaggregated into five categories: nondefense and defense capital goods, household consumers' goods, transportation equipment and a residual category. Shipments for manufacturing durables are disaggregated into transportation equipment and all other. All data series for orders and shipments are Bureau of Census data. An abbreviated glossary, interpreting all variable mnemonics appears in Volume II of this report. The following is a brief description of the equations which are documented in the following pages.

<u>18.13 ORNACGN$.</u> New orders for nondefense capital goods are explained by variables which are relevant to decisions on investment, the level of expected output and the cost of capital. The expectations and the effect of the cost of capital are both captured with distributed lags on the appropriate variables.

<u>18.14 ORSMFD371$.</u> Shipments of cars (conventionally assumed equal to new orders, which are not measured separately) adjust toward a desired level which depends on current domestic new car sales, on the difference between dealers' inventories and the desired level of such inventories as proxied by a three year moving average, and on the relative price of automotive expenditures.

<u>18.15 ORNAHD$.</u> The desired level of this small category, new orders for household consumers' goods, is linked to non-auto consumers' expenditures.

<u>18.16 ORNMFDOTHER$.</u> New orders in this residual category, which includes many intermediate goods, are taken to adjust toward a desired level which depends on the intended change in work in progress (as a measure of the scale of production) as well as on expected sales. The estimated equation makes

expected sales depend on durable manufacturing output and on the relative price of durable inventories.

18.17 ORNMFN$. This is the simplest form of stock-adjustment equation supplemented by a depressing effect of nondurable trade inventories on new orders.

18.18,18.19 ORSMFDNA$,ORSMFN$. These are the behavioral shipments equations which are omitted if the alternate behavioral unfilled orders relations 18.25 and 18.26 are used. All new orders which are measured net of cancellations are assumed to translate eventually into shipments - hence the unit coefficients on lagged new orders. The translation takes longer when the backlog of unfilled orders is high: this is picked up by a negative impact effect of this backlog, subsequently exactly offset.

18.25,18.26 ORUMFD$,ORUNFM$. (Alternates) Just as the translation of new orders to shipments dictates that the long-run coefficient on new orders in 18.18, 18.19 should be 1.0, so it requires that here it be 0.0. For durable unfilled order backlog subsequently offset, with the important addition of the difference between current new orders and a moving average to pick up the short-run inelasticity of shipments; for nondurables, it is achieved by omitting new orders altogether and allowing only a mild influence of capacity utilization.

August 5, 1977

18.13 New Orders, Net, Capital Goods, Nondefense = ORNACGN$

$$\frac{\text{ORNACGN\$}^{1/}}{\text{PDIBFNE}} = \underset{(3.39)}{-49.5425} + \sum_{i=0}^{2} a_i * \underset{\{2,FAR\}}{\text{XMF}}_{t-i}$$

$$+ \sum_{i=0}^{3} b_i * \underset{\{2,FAR\}}{\text{XMF.PRIMF}}_{t-i} + .700 * U_{t-1}$$

$\bar{R}^2$ = .967 SEE = 5.2048 DW = 2.112

Period of Fit: 195402 197404

Estimated: July 11, 1977

LAG	a_i	T	b_i	T
0	.274235	2.92	.00181438	.27
1	.161339	4.55	.00309879	.95
2	.0699276	1.12	.00322453	.70
3			.00219160	.56
SUM	.505501		.0103293	

XMF.PRIMF = (XMFD * PXMFD/UCKMFD) + (XMFN * PXMFN/UCKMFN)

1/ Equation is solved for ORNACGN$ for forecasting.

18.14 Value, Durable Manufactured Shipments, Transportation Equipment, Motor Vehicles & Parts = ORSMFD371$

$$\frac{ORSMFD371\$}{PDCEDAVN} - \frac{ORSMFD371\$_{-1}}{PDCEDAVN_{-1}} = \underset{(6.42)}{1.66623} * CEDAVND - \underset{(4.38)}{3.71482}$$

$$* \left(KIBITDAV_{-1} - \frac{1}{12} * \sum_{i=1}^{12} KIBITDAV_{t-i}\right) - \underset{(5.75)}{.317786} * \frac{ORSMFD371\$_{-1}}{PDCEDAVN_{-1}}$$

$$+ \sum_{i=0}^{4} a_i * \underset{\{2,FAR\}}{(CEDAVND * CAR+GASRP)}_{t-i}$$

$\bar{R}^2$ = .673 SEE = 2.6546 DW = 1.821

Period of Fit: 195404 197404

Estimated: July 7, 1977

LAG	a_i	T
0	-35.4573	1.96
1	-22.5046	2.98
2	-12.4825	3.34
3	-5.39103	1.02
4	-1.23020	.28
SUM		

CAR+GASRP = (CEDA$ + CENG$)/((CEDA + CENG) * PDGNP)

1/ Equation is solved for ORSMFD371$ for forecasting.

August 5, 1977

18.15 New Orders, Net, Consumer Durable Goods = ORNAHD

$$\frac{ORNAHD\$^{1/}}{PDCED} = \underset{(3.74)}{5.48095} + \underset{(1.98)}{.273309} * \frac{ORNAHD\$_{-1}}{PDCED_{-1}} + \underset{(4.49)}{.277972} * (CED - CEDA)$$

$$+ .772 * U_{t-1}$$

$\bar{R}^2 = .982$ SEE = .75658 DW = 1.88

Period of Fit: 195802 197404

Estimated: July 11, 1977

1/ Equation is solved for ORNAHD\$ for forecasting.

August 5, 1977

18.16 New Orders, Net, Manufacturing, Other Durables = ORNMFDOTHER$

$$ORNMFDOTHER - ORNMFDOTHER_{-1}{}^{1/} = \underset{(3.74)}{34.9028} - \underset{(2.78)}{1.50565}$$

$$* \frac{KICIFPMFD\$_{-1}}{PDKICIFPMFD_{-1}} - \underset{(5.62)}{.753671} * ORNMFDOTHER_{-1} + \sum_{i=0}^{2} a_i * XMFD_{t-i} \quad \{2,FAR\}$$

$$+ \sum_{i=0}^{2} b_i * (XMFD * PDKICIFMFD/PDGNP)_{t-i} + .246 * U_{t-1} \quad \{2,FAR\}$$

$\bar{R}^2$ = .498 SEE = 6.1356 DW = 1.800

Period of Fit: 195802 197404

Estimated: July 11, 1977

LAG	a_i	T	b_i	T
0	1.56440	2.88	-.653719	1.10
1	.295110	1.24	-.142246	.63
2	-.226356	.59	-.0756603	.19
SUM	1.63315		-.720305	

$$ORNMFDOTHER = \frac{ORNMFD\$}{PWDMFD/1.21133} - \frac{ORSMFD371\$}{PDCEDAVN} - \frac{ORNAHD\$}{PDCED} - \frac{ORNACG\$}{PDIBFNE}$$

1/ Although ORNMFDOTHER is used for estimation, the equation is solved for ORNMFDOTHER$ for forecasting.

August 5, 1977

18.17 New Orders Net, Nondurable Manufacturing = ORNMFN$

$$\frac{ORNMFN\$}{PWDMFN^{*}/1.14653} - \frac{ORNMFN\$_{-1}}{PWDMFN^{*}_{-1}/1.14653} \overset{1/}{} =- \underset{(3.34)}{38.7037} + \underset{(3.34)}{.770585} * CEN$$

$$- \underset{(.36)}{.227118} * KIBINWRN_{-1} - \underset{(4.31)}{.527603} * \frac{ORNMFN\$_{-1}}{PWDMFN^{*}_{-1}/1.14653}$$

$$+ .444^{*}\ U_{t-1}$$

$\bar{R}^2$ = .249 SEE = 3.9631 DW = 1.861

Period of Fit: 195303 197404

Estimated: July 11, 1977

1/ *Equation is solved for ORNMFN$ for forecasting.*

August 5, 1977

18.18 Shipments, Manufacturing, Nonauto Durables = ORSMFDNA$

$$ORSMFDNA^{1/} = \underset{(2.87)}{-\ 4.23059} + \sum_{i=0}^{6} a_i * \left(\frac{ORNACG\$}{PDIBFNE}\right)_{t-i}$$

$$+ \sum_{i=0}^{6} b_i * \left(\frac{ORNACG\$}{PDIBFNE} * \frac{ORUMFD\$}{PWDMFD/1.21133} / \frac{ORSMFD\$}{PWDMFD/1.21133}\right)_{t-i}$$

$$+ \sum_{i=0}^{3} c_i * ORNMFDOTHER_{t-i}$$

$$+ \sum_{i=0}^{3} d_i * \left(ORNMFDOTHER * \frac{ORUMFD\$}{PWDMFD/1.21133} / \frac{ORSMFD\$}{PWDMFD/1.21133}\right)_{t-i}$$

$$+ \frac{ORNAHD\$}{PDCED} + .663 * U_{t-1}$$

$\bar{R}^2$ = .874 SEE = 3.8267 DW = 1.896

Period of Fit: 195804 197404

Estimated: July 21, 1977

LAG	a_i 1/	b_i	c_i	d_i
0	.6012544	-1.2841234	.904664	-1.27196
1	.364823314	-.5503386	.3	0.0
2	.178571429	0.0	-.052332	.63598
3	.042498743	.3668924	-.152332	.63598
4	-.043394743	.5503386		
5	-.079109029	.5503386		
6	-.064644114	.3668924		
SUM	1.0	0.0	1.0	0.0

$$ORSMFDNA = \frac{ORSMFD\$}{PWDMFD/1.21133} - \frac{ORSMFD371\$}{PDCEDAVN}$$

1/ Although ORSMFDNA is used for estimation the equation is solved for ORSMFDNA$ for forecasting.

2/ The coefficients a_i, b_i, c_i and d_i are constrained so that they sum to 1.0, 0.0, 1.0 and 0.0 respectively. Individual T statistics are not calculated.

August 5, 1977

18.19 Shipments, Manufacturing, Nondurables = ORSMFN$

$$\frac{ORSMFN\$}{PWDMFN^*/1.14653}^{1/} = \underset{(1.85)}{-\ .271001} + \sum_{i=0}^{3} a_i * \left(\frac{ORNMFN\$}{PWDMFN^*/1.14653}\right)_{t-i}$$

$$+ \sum_{i=0}^{3} b_i * \left(\frac{ORNMFN\$}{PWDMFN^*/1.14653} * \frac{ORUMFN\$}{PWDMFN^*/1.14653} / \frac{ORSMFN\$}{PWDMFN^*/1.14653}\right)_{t-i}$$

$\bar{R}^2$ = .874 SEE = 1.0071 DW = 1.597

Period of Fit: 195303 197404

Estimated: July 20, 1977

LAG	a_i 2/	b_i
0	.825684	-1.658824
1	.3	0.0
2	-.012843	.829412
3	-.112842	.829412
SUM	1.0	0.0

1/ Equation is solved for ORSMFN$ for forecasting.

2/ The coefficients a_i and b_i are constrained so that they sum to 1.0 and 0.0 respectively. Individual T statistics are not calculated.

18.25 Unfilled Orders, Durable Manufacturers' Total, EOQ = ORUMFD$

$$ORUMFD\$ = ORUMFD\$_{-1} + .0460483 * (ORNMFDNA\$ - .2 * \sum_{i=0}^{4} ORNMFDNA\$_{t-i})$$

$$+ \sum_{i=0}^{6} a_i * (ORNACG\$ * ORUMFD\$/ORSMFDNA\$)_{t-i} + \sum_{i=0}^{3} b_i$$

$$* (ORNMFDOTHER\$ * ORUMFD\$/ORSMFDNA\$)_{t-i}$$

$\overline{R}^2$ = .948 SEE = .83256 DW = 1.827

Period of Fit: 195804 197404

Estimated: July 11, 1977

LAG	a_i 1/	b_i
0	.139500	.322515
1	.0597857	0.0
2	0.0	-.161258
3	-.039857	-.161258
4	-.0587857	
5	-.0597857	
6	-.039857	
SUM	0.0	0.0

1/ The coefficients a_i and b_i are constrained so that they sum to 0.0. Individual T statistics are not calculated.

18.26 Unfilled Orders, Nondurable Manufacturers' Total = ORUMFN$

$$\frac{ORUMFN\$}{PWDMFN^{*}/1.14653} - \frac{ORUMFN\$_{-1}}{PWDMFN^{*}_{-1}/1.14653}\ ^{\underline{1/}} = -\underset{(.15)}{.003954} + \underset{(5.54)}{.0278484} \ *$$

$$(100./(100. - CUWIPMFN) - 100./(100. - CUWIPMFN_{-1}))$$

$\bar{R}^2$ = .259   SEE = .23835   DW = 1.496

Period of Fit: 195303 197404

Estimated: July 20, 1977

1/ Equation is solved for ORUMFN$ for forecasting.

This set of relationships allows us to generate both unfilled orders and GNP on a quarterly basis. The remaining problem in the linkage is to produce monthly detail for the RASST model.

(2) In the development of the RASST model a series on monthly GNP in constant dollars was derived by using regression techniques to interpolate the quarterly data on current dollar GNP and using a linear interpolation of quarterly data to derive monthly data on the implicit deflator. Data on current dollar unfilled orders was, of course, available on a monthly basis. Constant dollar values were derived by deflating by the monthly value for the GNP implicit deflator.

The procedure used for GNP is not feasible for our purposes since it would require forecasting of additional monthly variables. Moreover since the only manner in which these two variables appear in the model is as ratios of one another and they are deflated by the same index it is only necessary to develop monthly totals for the current dollar values.

The technique used in developing these monthly totals is an interpolation technique developed by Leibenburg and Kaitz.[1/]

[1/] This technique is described in an undated U.S. Department of Commerce memorandum labeled Technical Note #1.

The method consists of applying a set of weights to quarterly totals for each of three successive quarters centered on the quarter to be interpolated. After a monthly series is generated for an initial quarter, weights are developed and applied to the succeeding three quarters and the third month of the preceding quarter.

The weights for the first quarter are based on a quadratic. For all succeeding periods weights are derived by assuming that a cubic parabola describes the flow over the time period to be interpolated. This cubic is fitted so:

(1) The area under it is equal to the given quarterly figure and the quarterly totals adjacent to it.

(2) The area under the curve for the last month prior to the quarter to be interpolated is equal to the last monthly figure previously computed.

This last condition assures that the resulting smoothed series does not exhibit jumps between quarters and requires that computations for preceding quarters be completed before the series for the current quarter can be computed. A complete derivation of the weights appears in Appendix I.

This interpolation technique is not useful for interpolating stocks since it depends on dividing flows over the period to be interpolated. It is therefore necessary to interpolate the flow variables, shipments and orders, and derive the stock of unfilled orders for each month form the appropriate identity.

CHAPTER 2

APPENDIX I

It is necessary to derive two sets of weights, the first for the initial quarter when no preceding monthly total is available and the second for all succeeding quarters where the constraint that the interpolated value for the last month of the preceding should be the same as the area under the curve of the interpolating surface for the current quarter, for that month. The first set of weights is derived in detail. For the second set the problem is essentially the same.

For the initial quarter, assume that the quadratic

$$f(t) = b_1 + b_2 t + b_3 t^2$$

is such that

$$Y(1) = \int_0^1 f(t)\ dt$$

$$Y(2) = \int_1^2 f(t)\ dt$$

$$Y(3) = \int_2^3 f(t)\ dt$$

where Y(2) is the value of the variable to be interpolated in the initial quarter to be interpolated, and Y(1) and Y(3) are respectively the values for the preceding and following quarters.

Integrating, we have three cubic equations which can be evaluated to yield the following three equations.

$$Y(1) = b_1 + b_2 + b_3$$

$$Y(2) = b_1 + 3b_2 + 7b_3$$

$$Y(3) = b_1 + 5b_2 + 19b_3$$

or $$Y = Xb$$

where $$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 7 \\ 1 & 5 & 19 \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

It follows that

$$b = X^{-1} Y$$

or $$b_1 = 11/6\ Y(1) - 7/6\ Y(2) + 1/3\ Y(3)$$

$$b_2 = -\ Y(1) + 3/2\ Y(2) - 1/2\ Y(3)$$

$$b_3 = 1/6\ Y(1) - 1/3\ Y(2) + 1/6\ Y(3)$$

Given values for Y(1), Y(2) and Y(3), b_1, b_2 and b_3 are determined. Values for each of the months in quarter 2 can be determined by evaluating the integral over the appropriate interval. Thus

$$m_1 = \int_1^{4/3} f(t)\, dt$$

$$m_2 = \int_{4/3}^{5/3} f(t)\, dt$$

$$m_3 = \int_{5/3}^{2} f(t)\, dt$$

Beyond the initial quarter it is assumed that f(t) is cubic rather than quadratic, i.e.

$$f(t) = b_1 + b_2 t + b_3 t^2 + b_4 t^3$$

Again $$Y(1) = \int_0^1 f(t)\, dt$$

$$Y(2) = \int_1^2 f(t)\, dt$$

$$Y(3) = \int_2^3 f(t)\, dt$$

but in addition

$$m_3 = \int_{2/3}^{1} f(t)\, dt$$

where the value of m_3 is determined initially as above and beyond the first period is the last monthly value from interpolation of the preceding quarter.

CHAPTER 3 - WEFA DEMOGRAPHIC LABOR FORCE MODEL

In order to provide an alternative approach to labor market explanation along more traditional lines, to eliminate the interface problem discussed in Chapter 2 and provide a more detailed age structure explanation of labor markets, Wharton has developed a quarterly model of labor markets based on explanation of participation and unemployment rates for 14 age-sex groups. When these are fully integrated into the model it enables us to explain both the rates and the levels for labor force, employment and unemployment for these age-sex groups.

Age-Sex Labor Force Equations

The civilian labor force participation rate equations are disaggregated into the 12 age-sex groups regularly calculated by the Bureau of Labor Statistics. The behavior across some of the age-sex groups is sufficiently marked to make the disaggregation useful. The specific age groups, for males and females, are ages 16-19, 20-24, 25-34, 35-44, 45-54, 55-64, and 65+. All variables are in logarithmic form so that the estimated coefficients are the short-run elasticities. For developing an initial set of equations, the equations were estimated using ordinary least squares on a single equation basis. This structural form was then introduced into the Wharton model for further testing and validation. This is discussed in Chapter 4.

Before describing individual equations, it is useful to introduce those variables which were entered into all equations. First, one or two lagged dependent variables were introduced to capture the dynamic, autoregressive changes in participation rates. This type of variable operates as a distributed lag generator for the other explanatory variables in the equation. For example, where one lagged dependent variable is included, the long-run elasticities are calculated to be $\alpha/(1-\lambda)$ where α is the short-run elasticity and λ is the coefficient on the lagged dependent variable.

The real wage variable captures a twofold effect. First, it is closely related to a time trend within the sample period. As such, it represents all of those factors, whether economic or not, which have increased along a trend rate of growth. The problem with this approach, which is true when any proxy variable is utilized, is that a post-sample break in the relationship between the time trend and the true underlying mechanism will not be captured by the equation. This tendency for post-sample residuals to grow can only be corrected by a continuous reestimation of the model. Second, the real wage variable is measured in terms of the "effective" real wage. That is, the employment rate of prime-age males, viewed as a measure of the probability of finding a job, is multiplied by the real wage variable. The prime-age male employment rate is utilized because it is the purest age-sex employment measure of cyclical fluctuations in job opportunities. The movement in real wages is only important to those on the margin of labor force participation if it can be converted into an actual job. Hence, the real wage multiplied by the employment rate captures this effect.

Third, a "previous lag" variable is constructed so as to measure a cohort effect. If a cohort, that is, a given age group, has a high participation rate when they are (say) 25-34, they are likely to carry over that high participation rate into the next age group. This variable essentially represents a lagged participation rate of a previous age-sex group. If this variable is significant it suggests that high participation rates of a given age group can be predicted looking back at that group's participation rate over its life-cycle.

The price variable, p/p*, measures changes in participation caused by fluctuations in the inflation rate. P is the price level and p* is a distributed lag on past price levels. This can represent money illusion or uncertainty. In the former case, workers are assumed to calculate their real wage by deflating by p* rather than by p. Hence, when the inflation rate is relatively large, p* is low relative to the actual price level p, and workers are "tricked" into believing that their real wages are higher than their actual level. With the uncertainty view, workers believe that a high rate of inflation causes or is associated with greater uncertainty about near term future movements in labor market conditions. The result is that marginal workers choose to search for a job now because they are concerned about future job opportunities.

To an important extent, the p/p* variable captures the short-run cyclical fluctuations in the economy. It is used in place of a straight discouraged worker effect (which is somewhat built into the real wage variable) because of the a priori notion that marginal workers are more responsive to uncertainties and problems created by inflation than they are to simple job availability. Estimating the participation equations confirms the relative importance of p/p* compared with the unemployment rate.

Each of the above variables was introduced into the 14 age-sex equations. They were omitted if their T values fell below the 95 percent confidence interval. Additional variables were added for males 16-19 and 20-24 and for females 16-19, 20-24, 25-34 and 35-44. For young males and females 16-24 the school enrollment rate was included. Since school is a frequent alternative to market work, the school enrollment rate should be inversely associated with labor force participation. Although the school enrollment is assumed to be exogenous in this model, a more complete economic-demo graphic model would include a separate school enrollment variable.

The age-specific fertility rate is included in the participation rate equations for young females. Child rearing, especially when the children are very young, is another alternative to market work. When the fertility rate is high (low), female participation is low (high). As with school enrollment, a more fully specified economic-demographic model would include a fertility rate equation. A fertility rate equation, which is compatible with the above specification of the labor force participation equations appears in Michael L. Wachter, "A Time Series Fertility Equation: The Potential for a Baby Boom in the 1980s." <u>International Economic Review</u> 16, No. 3 (October 1975), pp. 609-624.

The results for the various age-sex participation rate equations are discussed below:

<u>Males 16-19</u>:

The participation rate of this group is mostly explained by school enrollment rates and its own lagged dependent variable. Since equations are estimated in log-form, each coefficient denotes elasticities. A 1% increase in school enrollment rates decreases the participation rate by 0.14% in the short-run, and its long-run effect will be a 0.72% decrease. The participation rate turns out to be responsive to the wage

Table 3.1

AGE-SEX LABOR FORCE PARTICIPATION EQUATIONS

1954:1 - 1975:4

Group	Constant	School Enrollment Rate	Total Fertility Rate	Wage	p/p*	Provious lag	Lag (-1) Dependent Variable	Lag (-2) Dependent Variable	$\bar{R}^2$/DW	SEE
Males:										
16-19	-0.425692 (-2.47)	-0.144146 (-1.96)		0.058775 (2.25)			0.804110 (11.43)		0.853/1.925	0.0175
20-24	-0.346144 (-6.17)	-0.107323 (-6.31)			0.191056 (2.71)	-0.152928 (-3.87)	0.334465 (3.78)		0.857/2.103	0.0081
25-34	0.005100 (2.28)					0.090445 (3.43)	0.514916 (4.94)	0.224237 (2.18)	0.873/2.120	0.0026
35-44	0.108633 (6.15)			-0.025743 (-6.10)		0.533059 (5.34)	0.333478 (3.11)	-0.142046 (-1.35)	0.901/1.923	0.0023
45-54	0.030231 (1.92)			-0.006799 (-1.68)		0.394488 (1.78)	0.803783 (11.34)		0.962/2.067	0.0028
55-64	0.043990 (1.62)			-0.0081423 (-1.21)		0.819466 (3.01)	0.832855 (13.04)		0.988/1.940	0.0049
65+	-0.007363 (-0.57)					0.125062 (0.84)	0.984382 (47.46)		0.990/2.057	0.0183
Females:										
16-19	-0.368331 (-3.49)	-0.170866 (-2.35)	-0.123215 (-2.53)		0.672038 (1.91)		0.739748 (9.58)		0.905/1.907	0.0266
20-24	-0.571860 (-3.77)		-0.105688 (-5.02)	0.077523 (2.99)	0.512375 (2.96)		0.575406 (7.64)		0.938/2.134	0.0127
24-34	-0.315566 (-2.22)		-0.135677 (-3.72)	0.031221 (1.30)	0.576075 (2.49)		0.683815 (8.38)		0.991/2.107	0.0134
35-44	-0.750984 (-4.40)		0.057439 (-3.47)	0.102366 (4.01)	0.621536 (3.76)		0.555871 (6.31)		0.990/2.026	0.0093
45-54	-0.026528 (-2.52)					0.038407 (1.43)	0.796997 (7.78)	0.114062 (1.15)	0.984/1.999	0.0095
55-64	-0.063532 (-4.20)					0.0685244 (1.56)	0.875900 (20.36)		0.980/2.280	0.0140
65+	-0.026222 (-0.23)			-0.091864 (-3.29)			0.811769 (14.78)		0.834/2.256	0.0324

rate, but the size of the coefficient is small. A 1% increase in the wage rate results in an increase of about 0.6% in participation rates. While small, the impact of this is not negligible. For example, a 10% increase in the real wage rate in mid-1977 could have been expected to increase the labor force for this group by 30,000. The actual decrease in participation rates over the sample period is caused by an increase in the school enrollment rate whose impact outweighs the effect of the wage rate causing the participation rate to decline.

<u>Males 20-24</u>:

Two more variables, p/p* and previous lag, were significant in explaining the behavior of this group. The elasticity of the school enrollment rate is 0.11 which is less than that of males 16-19. The long-run effect is 0.16, that is, one percentage increase in the enrollment rate will decrease the participation rate by 0.16%.

The rate of inflation has a positive effect on participation rates. As was mentioned earlier, future uncertainty makes people join the labor force now because it will reduce the perceived future real income and increase potential search costs.

The previous lag variable has a very interesting meaning for this age group. The expected sign of this variable is generally positive, because the characteristics of the group are assumed to be carried over. However, for males 20-24, the sign of the previous lag variable turned out to be negative. The previous lag variable of males 20-24 represents the participation rates of the males 16-19 age group of three years ago. As was shown, the school enrollment rate plays an important role in explaining the behavior of that group. It is quite obvious that higher enrollment rates mean lower participation rates. Higher enrollment rates mean that people are investing in human capital, and therefore, it is more likely that they will join the labor force in the next period. Consequently, a lower participation rate due to a higher enrollment rate at time t will produce a higher participation rate at time t+1, when the investment effects are realized.

For the above two age groups, school enrollment rates play an important role in explaining the participation rates. In fact, these are not purely behavioral relations because participation rates and enrollment rates add up to approximately some constant. Enrollment rates work like a behavioral reaction variable.

The economic rationale of enrollment rates is explained in part by the "relative income model." This model is explained below. For a detailed explanation see Michael L. Wachter, "A Labor Supply Model for Secondary Workers," Review of Economics and Statistics 54, No. 2 (May 1972), pp. 141-151 and "The Demographic Impact on Unemployment: Past Experience and the Outlook for the Future," Demographic Trends and Full Employment (a special report of the National Commission for Manpower Policy), Special Report No. 12 (December 1976), pp. 27-99. The data show a decline in enrollment rates in recent years. This decline is largely attributed to a recent decline in the rates of return to education. The influx of younger educated workers need not have caused this decline, since the increased supply can affect skilled and unskilled markets alike. To this extent, the decline in the return to education may simply be due to an increase in the supply of educated manpower (above its trend rate of growth) relative to the trend rate of increase in the demand for such workers. If this is the case, and assuming that perception lags are not too long, a reversal can be expected over the next several years.

The dynamics of the relative income model would imply that a drop in permanent income (or discounted lifetime earnings) due to the decline in the rates of return to education would cause an increased difficulty in financing education. More important, to the extent that education represents a (luxury) consumption good as well as an investment good, the reduction in permanent income of the young cohort group would result in a cutback in education.

Males 25-34, 35-44, 45-54:

These three groups compose the prime-age male groups. In general, workers in these groups have high specific skills and show fairly stable participation rates over time. The wage rate has a negative sign for these groups. It represents a decreasing trend in participation rates of these groups within the sample period. This may be due to noneconomic forces or a negative income effect associated with rising real wages. Therefore the autoregressive scheme largely suffices to explain their participation behavior. In addition, the previous lag variables have positive signs and the magnitude of coefficients are significantly large. This means that if these groups have experienced high participation rates in previous periods, these characteristics will probably follow the cohort group as it ages and moves into the higher age brackets. Cyclical variables, such as p/p*, however, are not significant in explaining the participation rate of these age groups. This confirms the notion that prime-age males have stable labor market attachment and that changes in their participation rates are best explained by trend and autoregressive variables.

Males 55-64, 65+:

For these older age groups, the autoregressive scheme explains most of their behavior. The estimated results showed the participation rate for males 55-64 is not sensitive with respect to the wage rate.

Female Groups:

For young female workers, the decline in the fertility rate is an important variable in explaining the timing of the recent upswing in participation rates. Although the variable is treated as exogenous, it is useful to very briefly outline the relative income model which may explain the movement in that variable.

In the relative income model, young adults have a desired income level determined by the experience of their earlier years. If their present income is high, relative to the desired level, they have more children, ceteris paribus, and if it is low, they have fewer children. Since it is expected that the greater the number of children the lower the rate of labor market participation of females, the relative income model has indirect implications for labor supply behavior. The upswing in the fertility rate beginning in the 1940s created a subsequent rise in the size of the cohort of young workers in the 1960s. The increase in the relative supply of young workers caused a reduction in their expected permanent income or wage. The aspiration levels, in terms of desired or expected living standards, of these younger workers were formed during their childhood years when they lived with their parents. Anticipating the "normal" growth in income and status that their parents had accomplished, the adverse labor market conditions for young people in the 1960s caused a change in living patterns. Perhaps the most dramatic effect of the decline in actual living standards relative to aspiration levels was the drop in fertility rates. With fewer children and additional workers, families were able to salvage their expected living standards. The fluctuations in fertility rates with the ratio of actual/expected living standards sketches out the basics of an intermediate-to-long swings model. That is, cyclical swings of lower frequency or longer duration than the traditional business cycle can be expected in the demographic variables.

The relative income model does not purport to explain the entire drop in fertility or the rise in young female participation rates. Other factors, including exogenous changes in social mores, developments in contraceptive technology, and other economic variables, also had an impact. The data do support the model, however, and this suggests that changes in participation and fertility rates should not be treated as exogenous in economic models.

Fertility rates, p/p* and wage variables, together with the lagged dependent variable explain the participation behavior of females 20-44. They consistently show a similar pattern in terms of all these variables. Females 45 and older are different from the younger four groups because they are not child-bearing age groups. Relative income theory no longer applies to these groups. Previous lag variables are very large for these older female groups (though older male groups also show large coefficients), consequently, these cohorts are fairly stable over time (because one exogenous shock does not quickly dampen the cycle).

Females 16-19:

The short-run elasticity of participation rates with respect to school enrollment rates is -0.17%. The long-run effect, however, which takes into account a multiplier effect, turns out to be -0.65. In the short-run, the female participation rate is slightly more sensitive to enrollment rates than males, but the long-run effects are the reverse.

Among all the age groups, this cohort turned out to be the most sensitive to price inflation. The short-run elasticity is 2.58. That is, a one percent increase in price inflation will result in a 2.58 percent increase in participation rates. (It is important to remember that the numbers are elasticities and do not represent percentage points. For example, if the values of p/p* and participation rates in time t are 1.00 and 0.50 respectively, the effect of a change in p/p* to 1.01 in time t+1 (one percent increase) will mean an increase in participation rates to 0.5129 (i.e., 0.50 x 1.0258).)

A one percent increase in the total fertility rate of this group will decrease the participation rate of the group by 0.12 percent in the short-run and by 0.47 in the long-run.

Females 20-24:

A one percent increase in price inflation will lead to a 0.51 percent decrease of the participation rate in the short-run, and a 1.21 percent decrease in the long-run.

The short-run elasticity of the participation rate with respect to the total fertility rate turns out to be -0.11, and the long-run elasticity, -0.25. Comparing these results with the female 16-19 age group, the short-run effects do not vary significantly, but the female 20-24 cohort is much less sensitive to total fertility rates in the long-run. This is reasonable because the female 16-19 age group has a relatively much higher school enrollment rate than do females 20-24.

A positive trend effect appears to exist, but since the coefficient is small, the participation rate is inelastic with respect to the wage rate. However, as was explained in the male 16-19 age group, the effects of the wage rate can still be significant. The behavior of this group in the long-run is fairly stable.

Females 25-34:

The short-run elasticities with respect to price inflation of this group turns out to be 0.58 and the long-run effect is 1.82. The long-run effects of price inflation to this group are much larger than those of females 20-24. The trend effect is least evident here and the total fertility rates effects are not small for this group. The short-run elasticity is -0.14 and the long-run elasticity is -0.43.

Females 35-44:

For this age group, the total fertility rate effects are not very large. Since the relative income theory is primarily utilized to explain the behavior of females in the child-bearing age groups, the small effects of the total fertility rate are not surprising. The short-run elasticity is -0.06, the long-run, -0.13.

The positive trend effect is also apparent in this age group. The effect is larger than that of younger-age female groups. The price inflation effects continue to be significant, with the short-run elasticity of 0.62 and the long-run at 1.40.

Females 45-54:

Two lagged dependent variables and the previous lag variable were utilized to explain the behavior of this age group. Autoregressive schemes were sufficient to explain the partipation behavior of this cohort. The experiences from the previous group are positively carred over. The short-run effect of this characteristic seems rather insensitive, but the long-run effect is much greater because large coefficients in lagged dependent variables imply large multiplier effects.

Females 55-64:

The behavior of this group is similar to that of females 45-54. The positive carry over of the previous experience characteristics is still maintained, and the long-run effect is much greater than the short-run effect.

Females 65+:

The participation rates of this group have consistently decreased since the early 1950s. Social welfare programs are probably responsible for this early retirement trend. The wage variable validates this trend effect.

Age-Sex Unemployment Equations

Given the prime-age male unemployment rate, a very simple structure is formulated for predicting the unemployment rate of the other age-sex groups. The equations are of the form

$$U_i = \alpha_0 + \alpha_1 U_{pm} + \alpha_2 RP_y$$

where U_i is the age-sex unemployment rate, U_{pm} is the prime-age male unemployment rate, and RP_y is the proportion of young workers 16 to 24 relative to the total population of working age.

The implicit hypothesis in the above set of equations is that one can distinguish older workers with continuous labor-market attachment from younger workers and workers with discontinuous attachment in terms of their specific training. These labor groups become imperfect substitutes for one another so that the relative abundance of one group should alter wage differentials. If wage differentials among demographic groups are not sufficiently flexible, unemployment rates will change as well (or instead). In fact, in a world in which the labor requirements for capital equipment and the like are largely fixed, it may be difficult for relative wages to clear the market. And since young workers have a tendency to age over time, firms must anticipate demographic swings in the labor force. Of special importance in preventing relative wages from adjusting, and hence in thrusting the adjustment process onto unemployment rates, however, has been government policy. First, the major extension of minimum wages has prevented adjustments in demand that would favor lower-skilled workers. Second, changes in unemployment compensation and welfare have steadily increased the relative reservation price of labor, thereby lowering the cost of being unemployed.

Specifically, this model suggests that as RP_Y has increased, in response to the baby boom generation entering the labor force, the relative unemployment rates of young males and most female groups have increased for a given level of prime-age male unemployment. Unfortunately, time series on the various transfer payments and minimum wage laws that encompass both dollars per claimant and coverage are difficult to construct. Consequently, it is not possible to directly test the hypotheses that increases in the benefits available and especially in the coverage of these programs have increased the relative unemployment of young workers. The data that are available indicate that the cost of being unemployed and minimum wage coverage have changed along a time path similar to RP_Y. Hence that variable is utilized to capture these influences as well as the demographic shift.

Table 3.2

UNEMPLOYMENT RATE EQUATIONS, BY AGE AND SEX

1954:1 - 1975:4

Group	Constant	Prime-Age Male Unemployment Rate	RELPOP	$\bar{R}^2$/DW/SEE
Males:				
16-19	-0.0889392 (-0.50)	0.464031 (22.15)	0.748533 (13.23)	0.855/0.851/0.0647
20-24	-0.517042 (-1.84)	0.877549 (26.58)	0.533824 (5.99)	0.893/0.658/0.1019
25-34	-1.18398 (-8.20)	1.06414 (62.63)	0.422494 (9.22)	0.980/1.241/0.0525
35-44	0.449355 (3.27)	0.974934 (60.36)	-0.179345 (-4.11)	0.981/1.321/0.0499
45-54	1.55698 (9.32)	0.923054 (46.97)	-0.519845 (-9.80)	0.973/1.269/0.0607
55-64	2.69496 (9.44)	0.718786 (21.40)	-0.775566 (-8.55)	0.901/0.767/0.1037
65+	1.22766 (3.63)	0.584069 (-0.20)	-0.199044 (-1.86)	0.764/1.331/0.1227
Females:				
16-19	-1.21205 (-4.97)	0.282291 (9.85)	1.20149 (15.53)	0.747/0.684/0.0885
20-24	-1.74367 (-8.56)	0.508166 (21.20)	1.09208 (16.88)	0.861/1.222/0.0740
25-34	-0.596925 (-3.19)	0.513717 (23.33)	0.603675 (10.16)	0.862/1.031/0.0680
35-44	-0.437322 (-2.10)	0.561387 (22.87)	0.445664 (6.73)	0.858/1.008/0.0758
45-54	0.001035 (0.01)	0.616547 (23.50)	0.202232 (2.86)	0.873/1.551/0.0810
55-64	0.94309 (2.97)	0.618693 (16.55)	-0.160002 (-1.59)	0.799/1.394/0.1154
65+	-2.05468 (-4.28)	0.519806 (9.20)	0.874235 (5.73)	0.503/1.375/0.1745

The increase in coverage is of special importance since it largely affects the low-skilled, mostly young, workers who are disproportionately involved in cyclical unemployment and in the high turnover rates of the young and of (married) females. Reductions in the cost of being unemployed facilitate movements into and out of employment. Whether he is eligible for certain transfer payments helps an individual to choose between being unemployed and withdrawing from the labor force.

Taking RP_y as an indicator of demographic imbalance for all gorups in the labor market is not a strong assumption. Demographic trends being what they are, a relative increase in young people in the population implies quite directly a decrease in the relative population of older people. The equation then identifies which demographic groups are substitutes for younger workers ($\alpha_2 > 0$) and which groups are complements ($\alpha_2 < 0$). Essentially, almost all female groups and the young male groups have $\alpha_2 > 0$, while the older male groups have $\alpha_2 < 0$.

CONTRIBUTION RATES BY EACH VARIABLE

Age-Sex Group	Predicted Participation Rate (1954:1)	Predicted Participation Rate (1975:4)	School Enrollment Rate	Real Wage	p/p*	Total Fertility Rate	Previous Lag	Lag (-1) Dependent Variable	Lag (-2) Dependent Variable
Males:									
16-19	0.59972 (0.61689)	0.59141 (0.58049)	0.58390	0.61558				0.59178	
20-24	0.87892 (0.87328)	0.84497 (0.83925)	0.835284		0.389252		0.883673	0.87405	
24-34	0.97375 (0.97257)	0.95694 (0.95475)					0.970422	0.96457	0.969368
35-44	0.97857 (0.97842)	0.95999 (0.95723)		0.967448			0.97460	0.971338	0.98225
45-54	0.96345 (0.96153)	0.92616 (0.92521)		0.96054			0.958025	0.934222	
55-64	0.88170 (0.88860)	0.75612 (0.75178)		0.87852			0.861494	0.776661	
65+	0.39893 (0.40824)	0.21191 (0.20849)					0.394568	0.214245	
Females:									
16-19	0.39158 (0.41916)	0.49664 (0.48915)	0.373281		0.408018	0.415236		0.471517	
20-24	0.44140 (0.43589)	0.64284 (0.63801)		0.456858	0.45545	0.475966		0.55820	
25-34	0.35039 (0.35548)	0.54997 (0.55067)		0.355285	0.362962	0.380952		0.48160	
35-44	0.41851 (0.41270)	0.55766 (0.56253)		0.43798	0.43473	0.43495		0.49360	
45-54	0.40746 (0.41124)	0.54592 (0.54642)					0.412056	0.521679	0.421632
55-64	0.29106 (0.29163)	0.41113 (0.41229)					0.29799	0.401579	
65+	0.092527 (0.089345)	0.082013 (0.082997)		0.088828				0.085428	

NOTES ON CONTRIBUTION RATES TABLE:

The values in parentheses are the actual participation rates.

The Contribution Rates are interpreted as follows:

This rate implies the expected participation rate when only the value of an independent variable has changed from the 1954:1 value to the 1975:4 value. For example, in the male 16-19 age group, 0.58390 indicates that if only the school enrollment rate had changed from 0.594 (the value in 1954:1) to 0.715 (the value in 1975:4), all other variables remaining constant, the participation rate would have been 0.58390 in the 1975:4 period. The combined effects can also be calcu-ated. For example, the combined effects of school enrollment rate and real wage for the male 16-19 age group should be 0.59934 (0.58390 x 0.61558/0.59972). Because the estimations have been made in logarithmic form the effects are multi-plicative. The total effects are: (0.59141/0.59972) = (0.58390/ 0.59972) x (0.61558/0.59972) x (0.59178/0.59972).

NOTES ON TABLES 1 AND 2

I. Explanation of Variables

SER_i = School Enrollment Rate of group i, where i = males 16-19, males 20-24, females 16-19. (The values of this variable have a step-like form.)

TFR_i = Total Fertility Rate of group i, where i = females 16-19, 20-24, 25-34. (For the 35-44 age group, aggregate TFR was used.)

p/p^*:

p = Consumer Price Index

$$p^* = \sum_{i=-1}^{-16} (h_i P_i), \qquad h_i = \text{harmonic weight}, \quad \sum_{-1}^{-16} h_i = 1$$

$$\text{Wage} = \{\sum_{i=0}^{-7} [h_i(1-U_{pm_i})]\} \cdot \text{Real Wage}/8.0$$

where

h_i = harmonic weight, $\sum_{i=0}^{-7} h_i = 1$

U_{pm} = Prime-Age Male Unemployment Rate

Real Wage = Real Compensation per manhour

$$\text{Previous Lag}(i) = \sum_{t=-5}^{-k} [(\frac{LF}{POP})_{i-1,t}]/k - (5 - 1)$$

where

i = specific age-sex group; male, female, 20-24, ..., 65+

k = 12 for i = 20-24

k = 16 for i = 24-34

k = 20 for i = 35-44, ..., 65+

RELPOP = Percentage of young workers (age 16-24, both sexes) over total civilian noninstitutional population of working age.

II. Sources of data

SER: Handbook of Labor Statistics, 1975. Reference Edition. 1975 data were collected from Current Population Reports, March 1976 and Special Labor Force Report, March 1976.

TFR: Vital Statistics of the United States, Vol. I, Natality Tables 1-6.

All other data were either from the Wharton Econometric Forecasting Associates data bank or transformed by the author.

Chapter 4. Model Validation.

Even if one had a set of well-behaved estimates of the system parameters, the problem of generating goodness-of-fit measures and test statistics for large simulataneous systems is generally unresolved.[1/] Given that the two systems of interest have been estimated by ordinary least squares[2/] and the properties of the OLS estimator, i.e. that they can be expected to be inconsistent and biased, we must fall back on some rule of thumb measures for evaluating these models. Current practice in this situation is to validate models on the basis of historical simulation, multiplier analysis and/or forecasting results. We have benchmarked the RASST and WEFA models by simulating them over an historical period. Error statistics for these simulations appear in Tables 4.1 and 4.2. Before examining these results several points should be considered.

(1) The simulations were dynamic. That is, actual values of all endogenous variables for all periods prior to October, 1967 were used in solving the models, but solution values were used for any period beyond that time. This is a severe

---

[1/] See Dhrymes, P.J. Econometrics, Statistical Foundations and Applications, (New York: Harper and Row), 1970 pp. 240-277.

[2/] The RASST Model imposed certain cross-equation constraints on the system but yields the same statistical characteristics as ordinary least squares.

test, since both of these models have a strong dynamic character and depend heavily on lagged dependent variables.

(2) Actual values for all variables exogenous to the demographic labor market models were used for the historical simulations. While it would have been possible to simulate the models with the Wharton Quarterly Model solutions for all variables endogenized in the Wharton Model this would involve evaluating them not only on the basis of their internal characteristics but on the ability of the Wharton Model to forecast the variables necessary as inputs. Since the two sets of variables necessary are discrete, GNP are unfilled orders for RASST and prices and wages for the WEFA model, we have used actual values to avoid this problem. We are, however, interested in the question, how well would these two models have described labor markets if we could predict exogenous variables perfectly. To this end while we have used the ac data for quarterly variables, we have interpolated to the monthly level using the technique described in Chapter 3. This procedure biases the results against the RASST model to an unknown extent, since the WEFA model is using exogenous data that is the same as sample period data. This seems unavoidable if we are attempting to test the two models for forecasting characteristics among other things.

(3) One final caveat regarding the degree of exogeneity of the two models is in order.

Given a set of parameter estimates the solution to any econometric model through time, whether it is a single linear relationship or a large system of equations, will depend on the values taken by the exogenous variables and, if the system is dynamic, the set of initial conditions on the lagged variables. While this is obvious the importance this assumes in comparing two models or in specifying a model for forecasting purposes is often overlooked.

A simple example might illustrate the problem. Suppose that the behavior of the economic phenomena in which we are interested can be described by the recursive set of relationships

$$y_{t1} = \alpha_0 + \alpha_1 y_{t2} + \alpha_2 X_t + \varepsilon_{t1} \tag{1}$$

$$y_{t2} = \beta_0 y_{t-1,1} + \varepsilon_{t1} + \alpha_1 \varepsilon_{t2} \tag{2}$$

If the variable of ultimate interest is y_{t1} then three approaches to estimating this system are available. Equation (1) may be estimated as a single relationship assuming y_{t2} to be exogenous to the system when it actually is not. The equation set (1) and (2) may be estimated. Finally, we could substitute equation (2) out of the system and estimate the reduced form equation

$$y_{t1} = \alpha_0 + \alpha^*_1 y_{t-1,1} + \alpha_2 X_t + \varepsilon^*_{t1}$$

$$\alpha^*_1 = \alpha_1 \beta_0; \ \varepsilon^*_{t1} = \varepsilon_{t1} + \alpha_1 \varepsilon_{t2}$$

Any one of these procedures would be operational from an estimation point of view but they have different implications for required resources and usefulness in forecasting exercises. They also have different implications for bias and other statistical properties of the estimates.

If (1) is estimated as a single equation then it will be necessary to supply projections of the variables y_{t2} and X_t over the forecast horizon in order to use the model for forecasting y_{t1}.

If either the set of equations (1) and (2) or equation (3) is used to forecast y_{t1} the information requirement for y_{t2} over the forecast is eliminated and only X_t and an initial value for $y_{t-1,1}$ is required for forecasting. The difference between these two is the ability the former provides to monitor the feedback process through y_{t2}.

Notice also the difference in comparing these models on the basis of within sample simulations, the strategy we are using in comparing models for the purposes of selecting one for forecasting. In case (1), historical simulation will simply reproduce the residual errors from the least squares regression calculation. In case (2), simulation of the system dynamically and simultaneously allows error to enter the system both in terms of the estimation error and because both contemporaneous and historical values are produced by the system. The value of y_{t2} used in determining y_{t1} will not be the actual value used in estimation but the value predicted by equation (2).

Once the system is used to predict more than one period into the future, additional error may accumulate because the values for $y_{t-1,1}$ used to predict y_{t2} will be solution rather than historical values. Similar comments apply to case 3. The point to be recognized here is that in order to approximate a valid test of two forecasting systems which differ in their degree of exogeneity it would be necessary to attempt to approximate the methodology that will actually be used in projecting exogenous variables for forecasting purposes.

(4) Four measures of error are reported in Tables 4.1 and 4.2

MAE = Mean Absolute Error

$$\sum_{t=1}^{N} (|P_t - A_t|)/N$$

A_t = actual

P_t = predicted or simulated value

N = number of solution periods

MAPE = Mean Absolute Percentage Error

$$\sum_{t=1}^{N} (|(P_t - A_t)/A_t| * 100.)/N$$

RMSE = Root Mean Square Error

$$\left(\sum_{t=1}^{N} (P_t - A_t)^2/N\right)^{1/2}$$

RMSPE = Root Mean Square Percentage Error

$$\left(\sum_{t=1}^{N} ((P_t - A_t)/A_t)^2 / N \right)^{1/2}$$

While alternative measures of error are available these are easily comparable, serve our purposes, and are unlikely to give misleading indications of the relative accuracy of the two models.

(5) While the behavioral variables in the RASST model are labor market flows and in the WEFA model participation and unemployment rates, appropriate identities and the exogenous population data can be used to generate levels for labor force, employment and unemployment for each model. Tables 4.1 and 4.2 compile error statistics for the number of unemployed and the number in the labor force. These two variables would appear to be the most significant because:

(1) the labor force and unemployment pools essentially circumscribe the choice environment for those considering enlistment and separation and appear to be among the most likely candidates for demographic support of military flows.

(2) these two and their opposites, not in the labor force and employed, fully describe the state of the population.

Turning to the simulation results several observations can be made about both models. First, not unexpectedly, both models display more accuracy in tracking labor force behavior in the prime age groups. For the RASST model both the RMSE and MAPE for the labor force are smaller in the 25-29 year groups for every age-sex category. The WEFA model has its smallest per-

RASST MODEL
LABOR FORCE AND UNEMPLOYMENT
HISTORICAL SIMULATIONS 1967.10-1973.12
(MONTHS)

	MAE (Mil. of People)	MAPE	RMSR (Mil. of People)	RMSPE
NML16.19	.0200	4.56	.0244	5.69
NML20.24	.0155	2.07	.0200	2.58
NML25.59	.0236	0.66	.0299	0.84
NML60+	.0148	3.58	.0181	4.44
WML16.19	.1357	3.75	.1524	4.16
WML20.24	.0490	0.93	.0617	1.14
WML25.59	.2372	0.72	.2434	0.74
WML60+	.0625	1.36	.0740	1.61
NFL16.19	.0239	7.16	.0282	8.64
NFL20.24	.0204	3.25	.0251	4.02
NFL25.59	.0286	1.01	.0359	1.27
NFL60+	.0101	3.90	.0130	5.06
WFL16.19	.0773	2.69	.0947	3.33
WFL20.24	.0454	1.09	.0621	1.52
WFL25.59	.1930	1.05	.2377	1.29
WFL60+	.0395	1.63	.0498	2.05
NMUL16.19	.0119	9.67	.0156	11.76
NMUL20.24	.0169	17.09	.0221	20.16
NMUL25.59	.0219	13.41	.0289	16.40
NMUL60+	.0044	28.62	.0055	36.53
WMUL16.19	.0447	9.07	.0579	11.12
WMUL20.24	.0668	17.14	.0866	20.50
WMUL25.59	.1381	15.92	.1871	19.58
WMUL60+	.0190	13.94	.0248	16.41
NFUL16.19	.0111	9.98	.0139	12.51
NFUL20.24	.0101	10.82	.0122	13.51
NFUL25.59	.0197	10.85	.0243	12.87
NFUL60+	.0022	38.91	.0027	60.58
WFUL16.19	.0310	7.83	.0406	9.59
WFUL20.24	.0718	10.01	.0938	12.44
WFUL25.59	.0303	9.80	.0384	11.90
WFUL60+	.0129	17.95	.0167	21.56

MAE = Mean Absolute Error; MAPE - Mean Absolute Percentage Error; RMSE = Root Mean Square Error; RMSPE = Root Mean Square Percentage Error

1/ For translation of mnemonics see Appendix 4.1.

Table 4.2

WEFA DEMOGRAPHIC MODEL[1]
LABOR FORCE AND UNEMPLOYMENT
HISTORICAL SIMULATIONS 1967.4-1973.4
(QUARTERS)

	MAE (Mil. of People)	MAPE	RMSR (Mil. of People)	RMSPE
NLCM16.19	.0444	1.09	.0573	1.42
NLCM20.24	.0374	0.64	.0463	0.78
NLCM25.34	.0161	0.14	.0191	0.16
NLCM35.44	.0176	0.17	.0218	0.21
NLCM45.54	.0274	0.26	.0319	0.31
NLCM55.64	.0544	0.77	.0658	0.93
NLCM65+	.0778	3.84	.0909	4.48
NLCF16.19	.0958	2.86	.1201	3.52
NLCF20.24	.0449	0.94	.0552	1.16
NLCF25.34	.0718	1.26	.0865	1.55
NLCF35.44	.0368	0.61	.0490	0.82
NLCF45.54	.1045	1.65	.1237	1.95
NLCF55.64	.0773	1.90	.0916	2.25
NLCF65+	.0463	4.63	.0525	5.28
NUTM16.19	.019	3.12	.026	4.12
NUTM20.24	.032	7.69	.040	8.98
NUTM25.34	.012	3.62	.018	5.02
NUTM35.44	.009	4.26	.012	5.37
NUTM45.54	.008	3.65	.010	5.04
NUTM55.64	.013	7.18	.015	8.22
NUTM65+	.007	10.59	.008	13.86
NUTF16.19	.016	3.19	.019	3.98
NUTF20.24	.018	4.43	.020	4.94
NUTF25.34	.014	4.30	.018	6.18
NUTF35.44	.015	5.96	.018	7.05
NUTF45.54	.011	5.55	.013	6.73
NUTF55.64	.007	6.20	.010	8.20
NUTF65+	.004	11.82	.005	16.78

MAE = Mean Absolute Error; MAPE = Mean Absolute Percentage Error; RMSE = Root Mean Square Error; RMSPE = Root Mean Square Percentage Error.

[1] For translation of mnemonics see Appendix 4.1.

centage error either in the 25-34 or 35-44 age group. Moreover, errors for the 16-19 and 60+ (RASST) or 65+ (WEFA) groups are systematically larger than those for other groups. This behavior seems to reflect both less stable labor force attachment of these extreme age groups and the lack of structural variables relevant to the labor force choice. This seems likely to be particularly severe in the older age groups where neither model succeeds in capturing the opportunity cost of transfer income.

The second characteristic the two models share in their explanation of the labor force is a tendency toward greater accuracy in explaining the behavior of males than females. This is particularly striking in the case of the WEFA model where the percentage of error measures for the females in the labor force exceeds those for males in every age group. In the RASST model the only exception to this is for whites in the 16-19 age group.

When error measures for the number of unemployed are examined this systematic relationship of error to age-sex characteristics largely disappears. The one, unexpected, characteristic that the models do share is that for each sex/race group the smallest percentage errors appear in the 16-19 age category. This occurs despite the fact that the labor force total on which this estimate is based in the WEFA model exhibits larger errors than any other age group except the 65+ category. For the 16-19 age group the participation rate equations and unemployment rate equations in the WEFA model tend to make offsetting errors. In the RASST model the unemployment total is determined by netting flows into and out of the stock of the unemployed and the link between labor force and unemployment is not as close. It would appear that the explanation for this phenomena which appears in both models is to be found in the simple arithmetic fact that the number of unemployed is large in this category and so errors that are larger in numerical terms than for much larger labor force categories translate into smaller percentages. For example, within the WEFA model while the percentage errors in unemployment are smallest for the 16-19 age group in terms of thousands of people the error in this group is exceeded only by that for the 20-24 age group.

Given the similarity of patterns between the two models, reflecting the underlying data, there are substantial differences in the levels of accuracy obtained by the two models over this period. While direct comparisons cannot be made by cell, on the basis of these tables both the MAPE and RMSPE reported show a tendency to be much lower for the WEFA model than for the RASST model. For example the MAPE of 1.09 reported for males 16-19 in the WEFA model compares to RASST errors for the same age-sex group of 4.56% for non-whites and 3.75% for whites. For the 25-29 age group RASST has an MAPE of 0.66% for non-white males and 0.72% for white males. Of the the four age groups in this span in the WEFA model the percentage errors for 25-34, 35-44 and 45-54 run approximately 1/3 of these values, ranging from 0.14-0.26, while the error for the 55-64 age group is of a similar order of magnitude at 0.77%.

For females the order of magnitude of the errors in labor force totals appears to be much closer and the RASST model may exhibit some superiority in the 25-59 year range. MAPE for both white and non-white female labor force totals runs near 1.0%. Only the 35-44 year age group exhibits a lower error in the WEFA model with the other three age groups having errors from 1.26%-1.90%.

Turning to unemployment measures, if the 60+ and 65+ categories are dropped from consideration the largest MAPE in the WEFA categories is 7.69% for males 20-24. The smallest MAPE in the RASST categories is the 7.83% for white females 16-19.

While we have not run the RASST model with the original exogenous inputs to compare those results with those from the interpolated variables it does appear that a substantial amount of the error in the male unemployment arises from this source. Table 4.3 presents a comparison of the MAPE reported by Smith on an earlier version of RASST with those reported here. The errors for non-white males 20-24 and white males 20-24 and 25-59 show particularly striking increases of from 6.0 to 10.0 percentage points. However, even if Smith's error measures are used for comparison purposes the WEFA model appears to exhibit superior accuracy.

A final caveat should be noted in comparing these two models. The basic source data and simulation data for RASST is not seasonally adjusted monthly and for WEFA seasonally adjusted quarterly. The former is much "noisier". The process

Table 4.3

RASST

UNEMPLOYMENT MAPE

	Smith[1]	Current
NMUL16.19	13.8	9.7
NMUL20.24	11.2	17.2
NMUL25.59	8.8	13.4
NMUL60+	31.4	28.6
NMUL16.19	8.1	9.1
NMUL20.24	7.7	17.1
NMUL25.59	7.9	15.9
NMUL60+	12.6	13.9
NFUL16.19	11.2	10.0
NFUL20.24	11.4	10.8
NFUL25.59	14.3	10.9
NFUL60+	34.9	38.9
WFUL16.19	7.1	7.8
WFUL20.24	9.6	10.0
WFUL25.59	6.1	9.8
WFUL60+	15.9	18.0

---

[1] Smith, R. E. "A Simulation Model of the Demographic Composition of Employment, Unemployment, and Labor Force Participation: Status Report," Urban Institute Working Paper 350-65, p.36.

of seasonal adjustment and averaging to the quarterly level yields much smoother data. Even accounting for seasonal factors with dummy variables, as the RASST model does, will not completely eliminate the advantage of using a smoother series. An indication of the much greater variation in the RASST data is contained in the scatter diagrams of Figures 4.1-4.6.

It is our opinion that the WEFA approach to modeling these markets is likely to yield better forecasting properties given the current state of the art but there is still room for a difference of opinion.

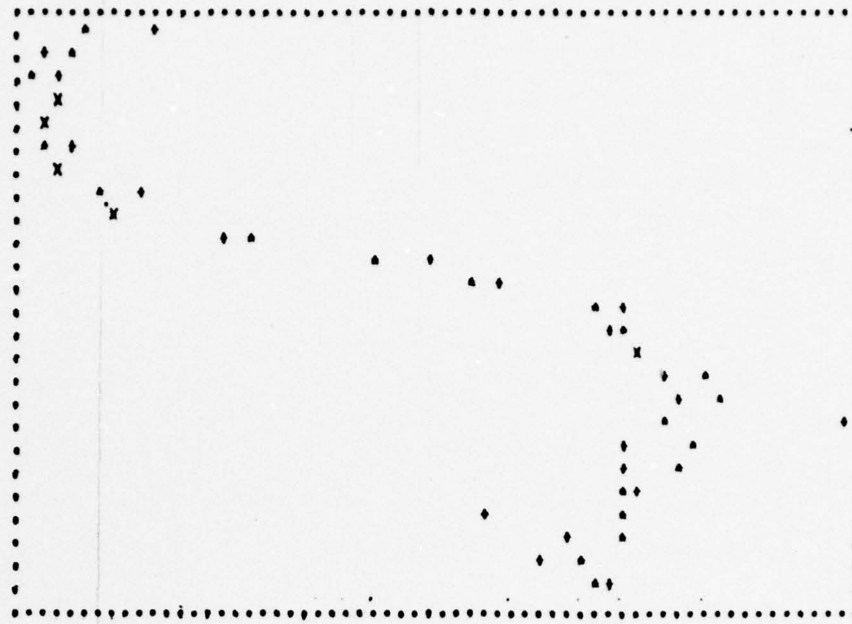
HISTORICAL SIMULATION - ONR PROJECT 1977

FIGURE 4.1

VARIABLE GRAPHED: NUTMIN,19

DATE	SIMULATION (+)	ACTUAL (+)	ERROR (TIE = X)	%ERROR
196704	0,433	0,463	-0,030	-6,97
196801	0,430	0,417	0,012	2,88
196802	0,410	0,422	-0,012	-3,02
196803	0,420	0,423	-0,003	-0,68
196804	0,418	0,419	-0,001	-0,28
196901	0,416	0,429	-0,013	-3,18
196902	0,425	0,420	0,005	1,16
196903	0,440	0,462	-0,023	-5,12
196904	0,449	0,445	0,004	0,94
197001	0,510	0,497	0,013	2,52
197002	0,570	0,590	-0,021	-3,63
197003	0,611	0,626	-0,015	-2,47
197004	0,667	0,678	-0,011	-1,65
197101	0,679	0,674	0,005	0,78
197102	0,687	0,684	0,003	0,47
197103	0,716	0,699	0,017	2,37
197104	0,726	0,707	0,018	2,52
197201	0,698	0,781	-0,083	-11,86
197202	0,711	0,683	0,028	3,89
197203	0,706	0,681	0,024	3,44
197204	0,682	0,686	-0,004	-0,65
197301	0,678	0,615	0,063	9,28
197302	0,679	0,656	0,023	3,43
197303	0,664	0,642	0,021	3,20
197304	0,665	0,676	-0,011	-1,66

GRAPH RANGE OF VALUES: 0,410 TO 0,781



SUMMARY STATISTICS: MEAN ABSOLUTE ERROR.......: 0,019
MEAN ABSOLUTE % ERROR.....: 3,12
ROOT MEAN SQUARED ERROR...: 0,026
ROOT MEAN SQUARED % ERROR.: 4,12

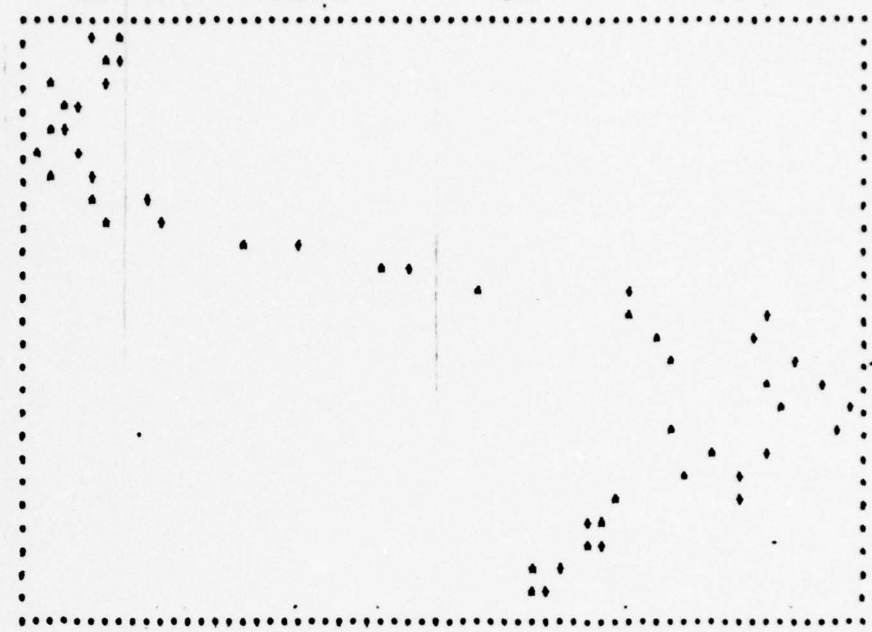
FIGURE 4.2

HISTORICAL SIMULATION - ONR PROJECT 1977

VARIABLE GRAPHED: OUTM20,24

DATE	SIMULATED (*)	ACTUAL (+)	ERROR (TIE = X)	%ERROR
196704	0,268	0,255	0,013	5,03
196801	0,267	0,271	-0,003	-1,21
196802	0,235	0,267	-0,032	-13,55
196803	0,238	0,252	-0,014	-5,92
196804	0,234	0,245	-0,010	-4,42
196901	0,227	0,252	-0,025	-10,92
196902	0,235	0,258	-0,023	-9,66
196903	0,255	0,283	-0,028	-10,80
196904	0,266	0,295	-0,030	-11,10
197001	0,337	0,363	-0,026	-7,83
197002	0,409	0,427	-0,018	-4,34
197003	0,463	0,540	-0,077	-16,60
197004	0,542	0,618	-0,076	-13,95
197101	0,557	0,608	-0,051	-9,24
197102	0,568	0,633	-0,064	-11,32
197103	0,616	0,649	-0,033	-5,33
197104	0,625	0,660	-0,035	-5,52
197201	0,568	0,653	-0,085	-15,04
197202	0,584	0,614	-0,030	-5,14
197203	0,570	0,604	-0,034	-6,04
197204	0,533	0,604	-0,071	-13,24
197301	0,526	0,523	0,003	0,54
197302	0,520	0,525	-0,005	-0,95
197303	0,492	0,506	-0,014	-2,84
197304	0,489	0,498	-0,009	-1,81

GRAPH RANGE OF VALUES: 0,227 TO 0,660



SUMMARY STATISTICS: MEAN ABSOLUTE ERROR.......: 0,032
MEAN ABSOLUTE % ERROR.....: 7,69
ROOT MEAN SQUARED ERROR...: 0,040
ROOT MEAN SQUARED % ERROR.: 8,98

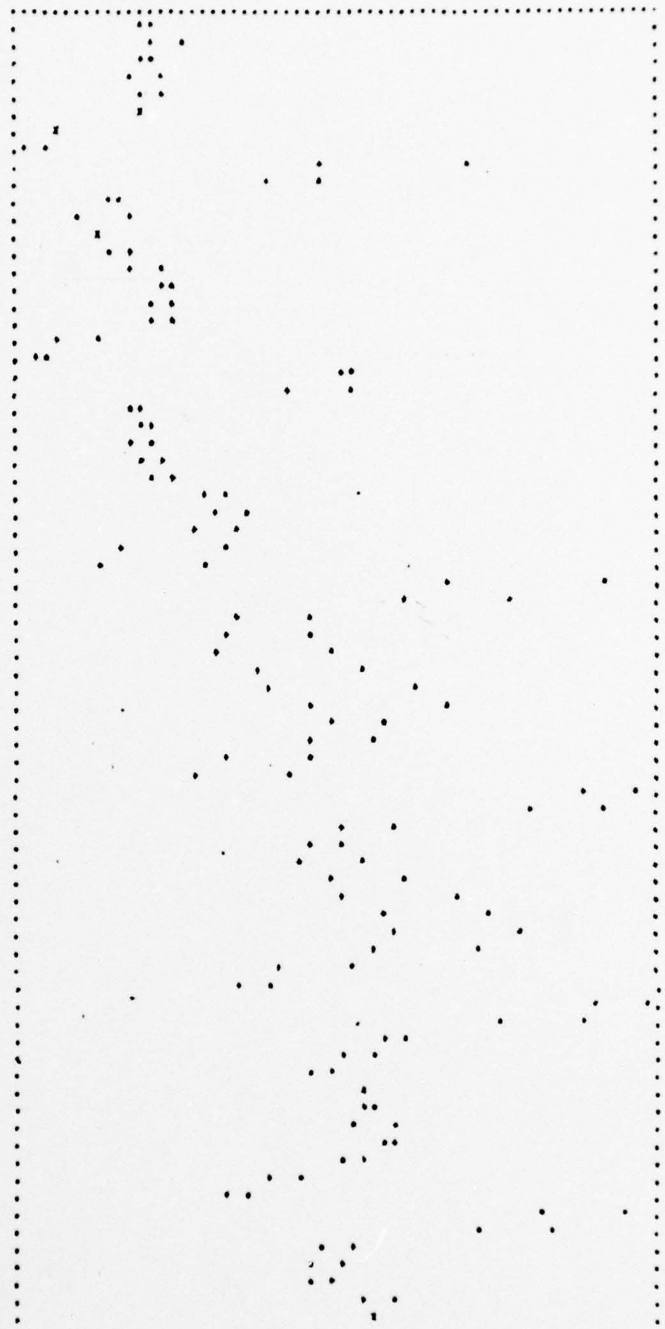
FIGURE 4.3

HISTORICAL SIMULATION - ONR PROJECT 1977

VARIABLE [illegible]

DATE	HISTORY ( )	SIMULATION (*)	DIFFERENCE ( )	%DIFFERENCE
196710	312.0	321.5	-9.5	-3.07
196711	347.0	326.0	21.0	6.05
196712	320.0	313.1	6.9	2.15
196801	308.0	330.5	-22.5	-7.35
196802	324.0	341.2	-17.2	-5.31
196803	308.0	302.5	5.5	1.83
196804	[illegible]	256.6	5.0	1.90
196805	232.0	210.1	21.9	9.45
196806	606.0	475.8	130.2	21.48
196807	471.0	425.1	45.9	9.75
196808	292.0	288.2	4.2	1.44
196809	268.0	308.4	-40.4	-15.07
196810	276.0	278.3	-2.3	-0.82
196811	[illegible]	305.8	-21.8	-7.52
196812	333.0	307.4	25.6	7.69
196901	340.0	333.7	6.3	1.86
196902	321.0	339.6	-18.6	-5.81
196903	340.0	320.4	19.6	5.77
196904	279.0	242.2	36.8	13.18
196905	228.0	219.3	8.7	3.82
196906	507.0	494.4	12.6	2.49
196907	506.0	443.3	62.7	12.39
196908	306.0	310.1	-4.1	-1.35
196909	317.0	323.8	-6.8	-2.15
196910	327.0	302.7	24.3	7.43
196911	318.0	335.6	-17.6	-5.54
196912	326.0	338.5	-12.5	-3.83
197001	390.0	370.7	19.3	4.95
197002	408.0	380.4	27.6	6.77
197003	398.0	365.5	32.5	8.16
197004	387.0	292.9	94.1	24.30
197005	371.0	272.4	98.6	26.58
197006	733.0	587.7	145.3	19.82
197007	646.0	543.5	102.5	15.81
197008	465.0	396.8	68.2	14.66
197009	463.0	393.3	69.7	15.06
197010	487.0	384.3	102.7	21.09
197011	514.0	421.4	92.6	18.02
197012	558.0	426.5	131.5	23.56
197101	587.0	469.0	118.0	20.11
197102	532.0	483.6	48.4	9.10
197103	528.0	468.9	59.1	11.20
197104	468.0	391.0	77.0	16.46
197105	442.0	358.9	83.1	18.81
197106	761.0	715.5	45.5	5.98
197107	730.0	667.2	62.8	8.61
197108	444.0	405.2	38.8	8.74
197109	[illegible]	447.5	20.5	4.79
197110	512.0	456.7	55.3	10.80
197111	549.0	486.7	62.3	11.35
197112	596.0	492.0	104.0	17.40
197201	626.0	532.4	93.6	14.95
197202	658.0	541.1	116.9	17.77
197203	611.0	518.9	92.1	15.08
197204	[illegible]	436.1	63.9	12.78
197205	428.0	398.2	29.8	6.96
197206	715.0	762.5	-47.5	-6.64
197207	633.0	714.2	-81.2	-12.82
197208	576.0	560.1	15.9	2.76
197209	526.0	497.0	29.0	5.51
197210	461.0	481.9	-20.9	-4.53
197211	510.0	508.2	1.8	0.36
197212	520.0	508.1	11.9	2.29
197301	498.0	540.2	-42.2	-8.48
197302	551.0	564.1	-13.1	-2.38
197303	495.0	516.1	-21.1	-4.27
197304	355.0	324.8	30.2	8.51
197305	408.0	387.3	20.7	5.07
197306	672.0	747.4	-75.4	-11.22
197307	617.0	683.9	-66.9	-10.84
197308	476.0	505.9	-29.9	-6.28
197309	465.0	491.7	-26.7	-5.75
197310	461.0	479.6	-18.6	-4.03
197311	536.0	514.0	22.0	4.10
197312	519.0	518.3	0.7	0.14

GRAPH RANGE OF VALUES: 210.1 TO [illegible]



SUMMARY STATISTICS: MEAN ABSOLUTE ERROR........: 44.7
MEAN ABSOLUTE % ERROR......: 9.37
ROOT MEAN SQUARED ERROR....: 57.9
ROOT MEAN SQUARED % ERROR.: 11.12

BEST AVAILABLE COPY

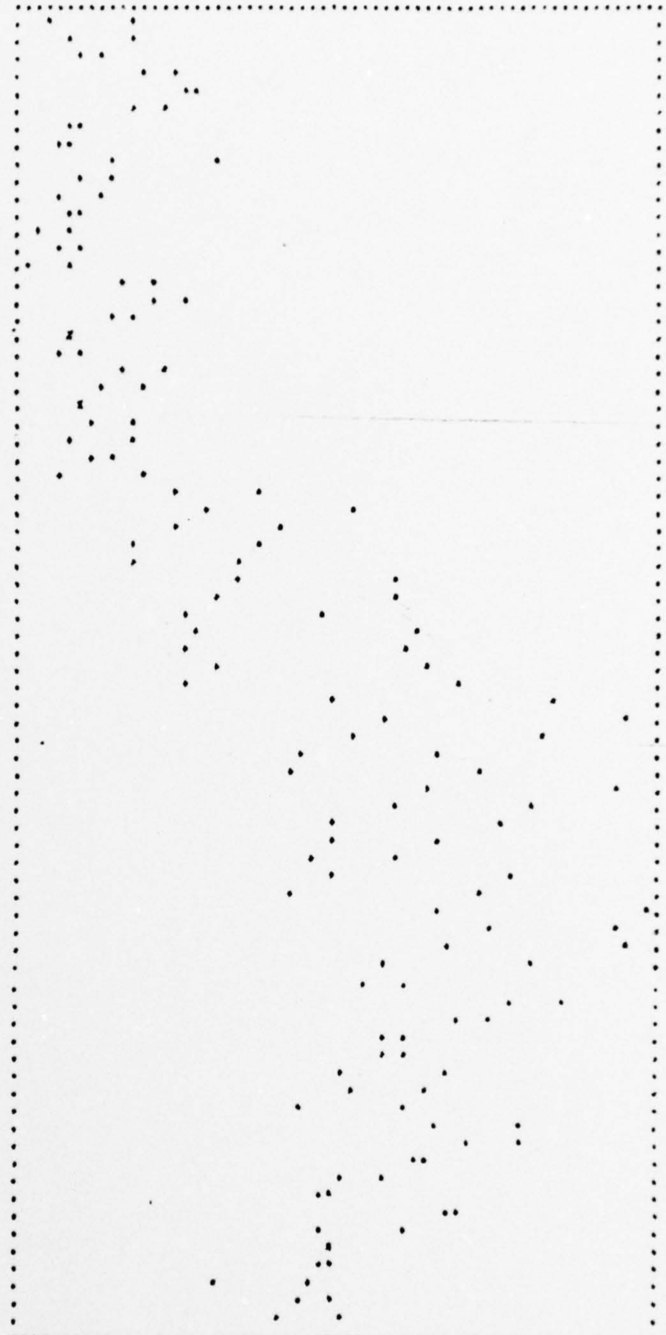
HISTORICAL SIMULATION - [illegible] PROJECT 1977

FIGURE 4.4

VARIABLE [illegible]

DATE	HISTORY ( )	SIMULATION ( )	DIFFERENCE (HIS - SIM)	%DIFFERENCE
196710	153.0	217.7	-64.7	-42.29
196711	172.0	216.7	-44.7	-25.99
196712	197.0	179.9	17.1	8.68
196801	226.0	250.5	-24.5	-10.84
196802	271.0	259.8	11.2	4.14
196803	247.0	220.5	26.5	10.73
196804	182.0	174.1	7.9	4.33
196805	168.0	165.5	3.5	2.09
196806	286.0	205.1	80.9	28.25
196807	[illegible]	161.0	[illegible]	9.97
196808	195.0	161.3	33.7	17.30
196809	176.0	167.4	10.6	5.94
196810	168.0	149.4	18.6	11.04
196811	177.0	162.0	15.0	8.48
196812	170.0	138.3	31.7	18.65
196901	233.0	211.2	21.8	9.37
196902	256.0	232.2	23.8	9.29
196903	218.0	205.6	12.4	5.75
196904	169.0	167.4	1.6	0.95
196905	176.0	163.5	12.5	7.13
196906	242.0	212.1	29.9	12.34
196907	229.0	195.6	33.4	14.58
196908	177.0	177.1	-0.1	-0.05
196909	221.0	186.8	34.2	15.49
196910	220.0	171.3	48.7	22.14
196911	205.0	186.7	18.3	8.94
196912	227.0	165.1	61.9	27.28
197001	313.0	249.3	63.7	20.36
197002	390.0	277.0	113.0	28.96
197003	331.0	252.7	78.3	23.65
197004	313.0	220.8	92.2	29.46
197005	303.0	220.8	82.2	27.14
197006	421.0	298.9	122.1	29.00
197007	424.0	285.7	140.3	33.08
197008	365.0	261.3	103.7	28.42
197009	435.0	271.7	163.3	37.54
197010	432.0	262.0	170.0	39.35
197011	447.0	284.4	162.6	36.37
197012	471.0	259.9	211.1	44.81
197101	548.0	372.5	175.5	32.03
197102	601.0	413.2	187.8	31.25
197103	532.0	388.3	143.7	27.02
197104	456.0	348.4	107.6	23.59
197105	489.0	341.7	147.3	30.12
197106	591.0	449.1	141.9	24.02
197107	527.0	420.4	106.6	20.22
197108	500.0	376.4	123.6	24.71
197109	453.0	375.2	77.8	17.17
197110	423.0	357.6	65.4	15.46
197111	508.0	369.6	138.4	27.24
197112	487.0	339.3	147.7	30.34
197201	617.0	455.8	161.2	26.13
197202	590.0	497.3	92.7	15.71
197203	600.0	461.1	138.9	23.15
197204	525.0	410.0	115.0	21.91
197205	434.0	395.8	38.2	8.80
197206	551.0	510.6	40.4	7.33
197207	496.0	473.9	22.1	4.45
197208	433.0	416.2	16.8	3.88
197209	430.0	410.2	19.8	4.61
197210	459.0	384.7	74.3	16.19
197211	446.0	392.9	54.1	12.18
197212	451.0	369.7	81.3	18.03
197301	512.0	451.9	60.1	12.75
197302	516.0	480.7	35.3	6.82
197303	446.0	441.0	5.0	1.12
197304	411.0	382.0	29.0	7.06
197305	370.0	365.5	4.5	1.20
197306	461.0	474.6	-13.6	-2.95
197307	368.0	431.3	-63.3	-17.21
197308	372.0	375.8	-3.8	-1.02
197309	362.0	377.4	-15.4	-4.26
197310	288.0	360.2	-72.2	-25.07
197311	351.0	374.0	-23.0	-6.57
197312	383.0	332.8	50.2	13.10

GRAPH RANGE OF VALUES: 138.3 TO 617.0



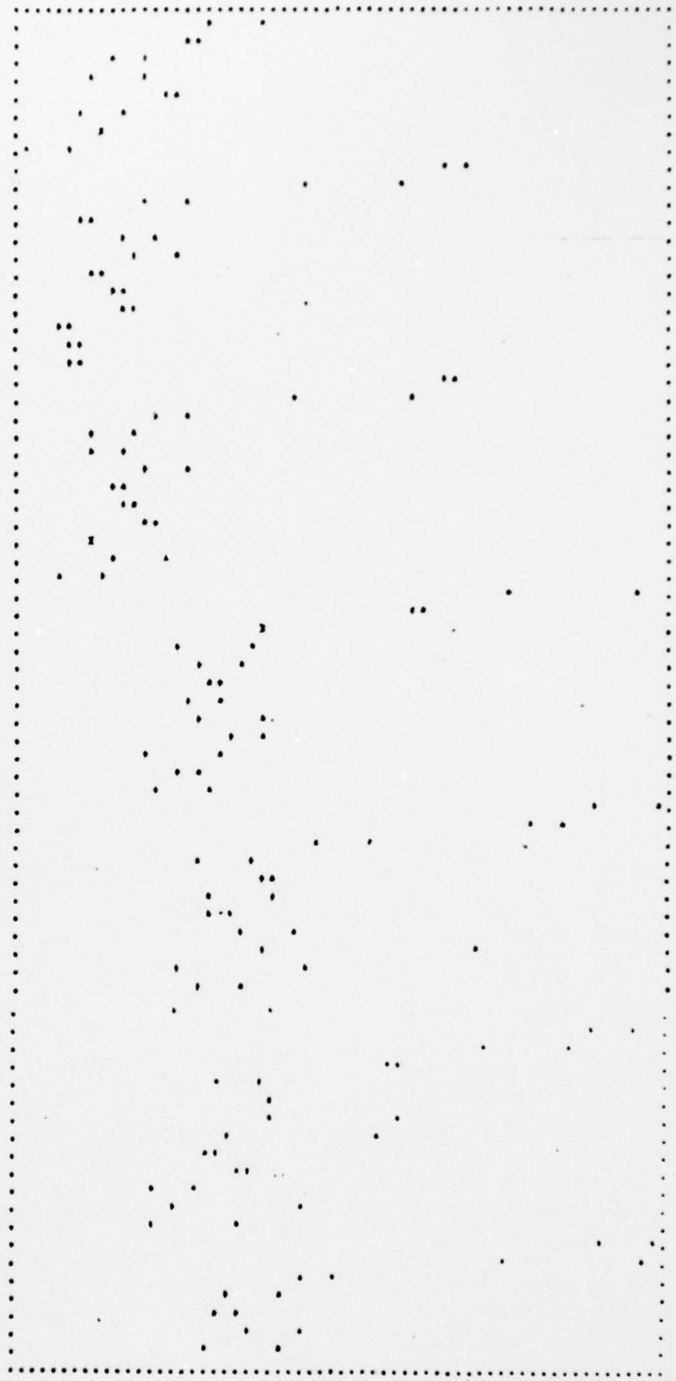
SUMMARY STATISTICS: MEAN ABSOLUTE ERROR........: 90.8
MEAN ABSOLUTE % ERROR......: 17.19
ROOT MEAN SQUARED ERROR...: 88.6
ROOT MEAN SQUARED % ERROR.: 20.50

BEST AVAILABLE COPY

VARIABLE: [illegible]

DATE	HISTORY (*)	SIMULATION (+)	DIFFERENCE (HIS - SIM)	%DIFFERENCE
196710	120.0	105.8	14.2	11.80
196711	102.0	104.1	-2.1	-2.06
196712	82.0	91.5	-9.5	-11.59
196801	76.0	90.5	-14.5	-19.09
196802	100.0	95.5	4.5	4.51
196803	84.0	74.6	9.4	11.19
196804	81.0	78.6	2.4	2.95
196805	61.0	72.1	-11.1	-18.19
196806	171.0	167.2	3.8	2.20
196807	156.0	129.8	26.2	16.80
196808	102.0	91.5	10.5	10.27
196809	78.0	75.2	2.8	3.59
196810	92.0	83.3	9.7	10.54
196811	100.0	87.7	12.3	12.31
196812	77.0	79.3	-2.3	-2.98
196901	86.0	81.6	4.4	5.17
196902	84.0	88.9	-4.9	-5.86
196903	73.0	69.8	3.2	4.41
196904	73.0	75.4	-2.4	-3.29
196905	75.0	70.8	4.2	5.59
196906	168.0	166.8	1.2	0.70
196907	157.0	129.5	27.5	17.51
196908	101.0	92.8	8.2	8.13
196909	89.0	77.1	11.9	13.38
196910	76.0	86.1	-10.1	-13.30
196911	102.0	90.1	11.9	11.66
196912	84.0	82.4	1.6	1.92
197001	89.0	85.4	3.6	4.08
197002	91.0	94.3	-3.3	-3.57
197003	78.0	76.3	1.7	2.18
197004	95.0	83.4	11.6	12.18
197005	70.0	79.9	-9.9	-14.11
197006	214.0	183.3	30.7	14.34
197007	160.0	157.2	2.8	1.78
197008	120.0	119.2	0.8	0.70
197009	117.0	97.5	19.5	16.65
197010	115.0	104.4	10.6	9.19
197011	104.0	109.4	-5.4	-5.22
197012	110.0	100.4	9.6	8.76
197101	119.0	104.3	14.7	12.39
197102	121.0	111.4	9.6	7.96
197103	110.0	91.3	18.7	16.95
197104	103.0	98.0	5.0	4.90
197105	109.0	95.2	13.8	13.69
197106	219.0	202.8	16.2	7.38
197107	195.0	188.8	6.2	3.32
197108	153.0	168.7	-15.7	-10.27
197109	103.0	115.9	-12.9	-12.56
197110	123.0	119.7	3.3	2.68
197111	108.0	122.5	-14.5	-13.39
197112	106.0	113.3	-7.3	-6.93
197201	129.0	115.4	13.6	10.55
197202	173.0	120.4	52.6	30.41
197203	132.0	98.1	33.9	25.67
197204	117.0	107.1	9.9	8.29
197205	122.0	99.4	22.6	18.56
197206	205.0	214.2	-9.2	-4.51
197207	177.0	197.7	-20.7	-11.69
197208	152.0	154.1	-2.1	-1.39
197209	110.0	121.1	-11.1	-10.11
197210	123.0	122.9	0.1	0.05
197211	155.0	123.6	31.4	20.28
197212	151.0	111.2	39.8	26.36
197301	107.0	110.7	-3.7	-3.45
197302	116.0	116.6	-0.6	-0.50
197303	105.0	93.6	11.0	10.87
197304	132.0	99.4	32.6	24.71
197305	114.0	94.0	20.0	17.54
197306	220.0	206.8	13.2	6.00
197307	217.0	183.1	33.9	15.69
197308	132.0	139.5	-7.5	-5.51
197309	125.0	112.2	12.8	10.21
197310	110.0	115.9	-5.9	-5.36
197311	130.0	118.7	11.3	8.72
197312	123.0	107.7	15.3	14.50

GRAPH RANGE OF VALUES: 61.0 TO 220.0



SUMMARY STATISTICS: MEAN ABSOLUTE ERROR........: 11.9
MEAN ABSOLUTE % ERROR.....: 9.67
ROOT MEAN SQUARED ERROR...: 15.6
R.M.S. MEAN SQUARED % ERROR.: 11.78

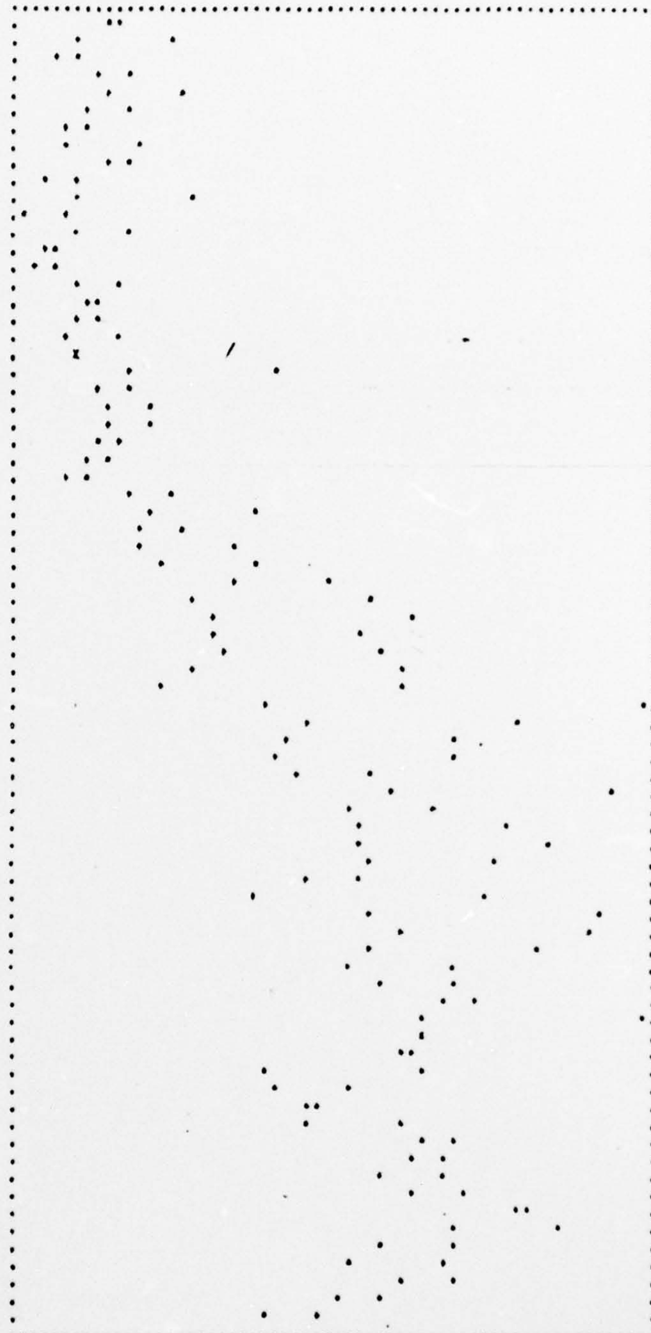
HISTORICAL SIMULATION - OHM PROJECT 1977

FIGURE 4.6

VARIANCE GRAPH OF: [illegible]

GRAPH RANGE OF VALUES: 37.0 TO 155.0

DATE	HISTORY (*)	SIMULATION (+)	DIFFERENCE ([illegible])	%DIFFERENCE
196710	53.0	54.1	-1.1	-2.14
196711	65.0	47.5	17.5	26.88
196712	47.0	42.0	5.0	10.61
196801	57.0	51.7	5.3	9.22
196802	66.0	53.1	12.9	19.58
196803	57.0	48.3	8.7	15.29
196804	49.0	45.6	3.4	7.03
196805	59.0	45.5	13.5	23.00
196806	57.0	52.8	4.2	7.39
196807	41.0	46.0	-5.0	-12.20
196808	69.0	48.3	21.7	31.44
196809	32.0	41.0	-9.0	-28.28
196810	56.0	46.4	9.6	17.19
196811	35.	33.5	1.2	3.40
196812	42.0	38.2	3.8	9.10
196901	55.0	47.2	7.8	14.20
196902	51.0	49.8	1.2	2.29
196903	50.0	46.7	3.3	6.60
196904	54.0	45.0	9.0	16.59
196905	47.0	47.5	-0.5	-0.98
196906	84.0	57.7	26.3	31.30
196907	56.0	52.0	4.0	7.14
196908	60.0	53.1	6.9	11.50
196909	61.0	53.2	7.8	12.71
196910	50.0	54.2	-4.2	-8.39
196911	53.0	48.8	4.2	7.91
196912	49.0	45.5	3.5	7.07
197001	65.0	57.0	8.0	12.37
197002	80.0	62.0	18.0	22.50
197003	67.0	59.0	8.0	11.94
197004	77.0	58.8	18.2	23.61
197005	80.0	62.6	17.4	21.75
197006	95.0	76.7	18.3	19.26
197007	103.0	70.0	33.0	32.04
197008	110.0	72.5	37.5	34.11
197009	101.0	73.2	27.8	27.50
197010	104.0	75.5	28.5	27.38
197011	109.0	68.7	40.3	36.95
197012	108.0	63.8	44.2	40.91
197101	154.0	83.1	70.9	46.04
197102	130.0	90.5	39.5	30.35
197103	119.0	86.7	32.3	27.15
197104	118.0	84.4	33.6	28.51
197105	102.0	89.8	12.2	11.96
197106	149.0	107.7	41.3	27.69
197107	115.0	99.6	15.4	13.36
197108	128.0	101.7	26.3	20.57
197109	137.	100.6	36.4	26.55
197110	137.0	111.1	25.9	18.95
197111	101.0	91.6	9.4	9.30
197112	126.0	81.0	45.0	35.64
197201	146.0	102.8	43.2	29.62
197202	135.0	99.2	35.8	26.52
197203	135.0	103.3	31.7	23.47
197204	119.0	99.0	20.0	16.83
197205	119.0	104.1	14.9	12.53
197206	116.0	123.5	-7.5	-6.55
197207	155.0	112.8	42.2	27.29
197208	112.0	113.6	-1.6	-1.42
197209	108.0	111.4	-3.4	-3.17
197210	82.0	112.2	-30.2	-36.84
197211	84.0	99.5	-15.5	-18.42
197212	93.0	90.8	2.2	2.37
197301	90.0	109.5	-19.5	-21.65
197302	113.0	118.5	-5.5	-4.85
197303	117.0	111.1	5.9	5.03
197304	117.0	105.5	11.5	9.85
197305	120.0	110.1	9.9	8.24
197306	132.0	132.0	0.0	0.02
197307	134.0	114.2	20.8	14.78
197308	105.0	118.2	-13.2	-12.61
197309	98.0	117.0	-19.0	-19.36
197310	109.0	118.6	-9.6	-8.82
197311	96.0	105.8	-9.8	-10.18
197312	83.0	92.7	-9.7	-11.69



SUMMARY STATISTICS: MEAN ABSOLUTE ERROR........: 18.9
MEAN ABSOLUTE % ERROR......: 17.09
ROOT MEAN SQUARED ERROR...: 22.1
ROOT MEAN SQUARED % ERROR.: 20.19

BEST AVAILABLE COPY

Appendix I.

Mnemonic Conventions

A. RASST Model

All labor force variables have attached a mnemonic that can be broken into four fields describing respectively race-sex-labor force status-age. The field values are

Race:

N = non-white
W = white

Sex:

M = male
F = female

Labor Force Status:

UL = unemployment level
E = employment
L = labor force
UR = unemployment rate
EN = flow from employment to not in the labor force
UN = flow from unemployment to not in the labor force
NL = flow from not in the labor force to the labor force
EU = flow from employment to unemployment
UE = flow from unemployment to employment
NE = flow from not in the labor force to employment
NU = flow from not in the labor force to unemployment
NE/NL = probability of successful labor force entry

Age:

16-19
20-25
25-59
60+

thus

NMUL20-24 = non white male level of unemployment between ages 20 and 24 inclusive

B. WEFA Model

All labor force variables begin with an "N". The second letter may indicate that the variable is a rate in which case the letter is "R". If no R appears the second letter indicates the component of the labor force alluded to, such as L for labor force, E for employment or U for unemployment. If an R does appear in the second place, the third place will assume these values. The next field indicates whether total (T) or civilian (C) labor force measures are being used. The next field the sex measure (M,F), if any, and the final field the age category if any.

thus

NRLTF65+ = labor force participation rate for females 65 and over

while

NRUT20.24 = unemployment rate for all persons 20-24

A complete alphabetical glossary of WEFA model variables appears in Volume II of this report.

Chapter 5 - Labor Market Forecast

Forecast Strategy

On the basis of the simulation results we have proceeded to use the WEFA model in an initial forecasting exercise. While the simulation results lead us to feel that the WEFA model is likely to produce superior forecasts, there are additional reasons for proceeding to forecast with this model. First we feel that the finer age breakdown in the WEFA model will be necessary to support explanation of separations when we turn to that portion of the study. Second, the distinctive labor market patterns for non-whites over the sample period seems very likely to be a function of the skill distribution among these groups rather than their race. Attempting to forecast far enough into the future to allow policies to affect this skill distribution is likely to mean that a substantial shift will occur in non-white labor market patterns. Age distinctions based on accumulated skill and experience and life cycle patterns are likely to be much more stable. Finally, forecasting provides an additional validation test for the model.

The final validation of any econometric model is its ability to produce "sensible" forecasts or at a minimum to be able to simulate with some degree of accuracy the path of economic phenomenon outside the sample period.

While it is not unusual for post sample performance to be advocated as a criterion for model validations, the extent to which the least squares estimation technique may obviate within sample simulation results appears to have been neglected. The obvious point to be made is that the least squares estimator picks the set of estimates which minimizes

$$\sum_{t=1}^{T} (Y_t - \hat{Y}_t)^2$$

no matter what the extent of specification error. Within sample simulations will in fact benefit from this characteristic. To choose an extreme example one could specify a completely recursive system

$$Y_{t1} = a_{11} + a_{21} X_{t1}$$

$$Y_{t2} = a_{12} + a_{22} X_{t2} + a_{32} Y_{t1}$$

Clearly so long as each individual equation is chosen to have a high R^2 the system as a whole is not likely to drift far from the sample experience.

Moreover for certain types of specification, response characteristics may approximate the underlying structure within the neighborhood of the sample mean. For example, suppose the correct specification of a consumer durables expenditures equation is

$$CD/Y = \beta_0 + \beta_1 P_r + \beta_2 r + \varepsilon$$

or $$CD = \beta_0 Y + \beta_1 P_r Y + \beta_2 r Y + \varepsilon Y$$

with CD = Expenditure on Consumer Durables

Pr = Relative Price

Y = Income

but the estimated equation is

$$CD = \beta_0^* + \beta_1^* P_r + \beta_2^* r + \beta_3^* Y + \varepsilon^*$$

Under some fairly stringent assumptions it is possible to show that the estimate of β_3^* that results from applying ordinary least squares to the mispecified equation has a probability limit that is approximately

$$\beta_0 + \beta_1 \bar{P} + \beta_2 \bar{r}$$

i.e., in the neighborhood of the means of the explanatory variables the estimated/simulated effect of an increase in Y on CD is likely to be approximately the same whichever version of the equation is estimated.[1/]

Since, in fact, the estimation technique is working so hard for us within the sample period, satisfactory reports on simulation results in periods should be treated with a great deal of circumspection.

[1/] For a proof of this statement see Appendix 5.I.

While we find the simulation results of the WEFA model encouraging, it must be recognized that the forecast discussed in this Chapter and appearing in Volume II is an initial trial at using this model. Until some experience is developed in forecasting with the model some circumspection is necessary in using it.

The Wharton Quarterly Model already has in it relationships for determining the total level of the labor force, total employment and by implication the number of unemployed.[2/] Substantial experience has been accumulated using these for forecasting. In order to make use of this experience we have used forecasts for these totals as controls in developing a forecast of the demographic detail of the labor market. The WEFA demographic model is thus essentially used to allocate the labor force into demographic cells after the totals are determined within the Quarterly Model.

For the labor force this is done by taking the predicted participation rates for each population group and scaling it by the ratio of total civilian labor force, NLC, predicted by the Quarterly Model to the labor force implied by the demographic subsector, i.e.,

$$NLC/\sum_i \gamma_i \; NPCNM_i$$

[2/] These were documented in the previous interim report on this contract.

where γ_i is the participation rate of the ith age-sex group and $NPCNM_i$ is the population of that group. This ratio is then applied to the γ_i to enforce the identity

$$NLC = \sum_i \gamma_i' \, NPCNM_i$$

The value of this ratio is displayed in Table 1, Line 19 of the demographic forecast tables in Volume II. A similar procedure is followed with respect to the unemployment rate and the value of the ratio used there appears Table 1.2, Line 27.

All inputs to the model for the forecast appear in the forecast tables. Exogenous variables are identified by the letter E following the mnemonic.

Given the primary concern with labor markets some comment is necessary concerning assumptions directly affecting this sector:

1. Population: The assumptions for total population and the age-sex detail are taken from Population Estimates and Projections, Bureau of Census, Series p-25 No. 704 July, 1977.

2. School Enrollment Rates and Fertility Rates: These have been extrapolated taking into account recent trends and with respect to the fertility rates some modification based on M. Wachter's earlier work in this area cited in Chapter 3. These procedures yield enrollment rates that are constant through 1981 and a slowly declining fertility rate.

The general procedure for establishing assumptions is to incorporate recent information on such things as government expenditures and taxes for the initial 8-12 quarters and allow exogenous variables to approach longer run trends beyond that period.

One final comment on the current forecast should be made. The level of military manpower is exogenous in this forecast. This means that we have assumed these will be the levels and that the total military budget and wage bill assumed is consistent with achieving this total. At a later stage of the research we would hope to endogenize military manpower so that given various parameters of military personnel policies, including pay, we can determine the resulting force levels and personnel costs.

Forecast Results

General Economic Outlook:

The Wharton Forecast is for a cyclical slowing in the growth rate of the economy over the next four years. While the growth rate for constant dollar GNP in 1977 is expected to be just under 5.0%, the outlook is for growth to slow to near 4.0% in 1978 and be below 4.0% in the 1979-81 period. This slowdown occurs despite a sizable tax cut which has been included as part of our assumptions. Without this tax cut the growth rate in 1979 would be forecast approximately 0.5% less than the 3.75% in the present forecast.

The basic cause of the slowdown is a convergence of weakness in the most highly cyclical sectors of the private sector which is exacerbated by high interest rates and inflationary pressures contained in current government policy. The cyclical patterns followed by consumer durables expenditure, residential investment, business investment in plant and equipment and inventories over the past two years would result in a slowing of growth as a result of a simple accelerator process even if no alterations in policy occurred. The adjustment to higher desired stocks of durables, housing, and fixed and working capital fueled growth above trend through the period 1975-77 and as that adjustment is completed we would normally expect some slowing in growth. This natural slowdown resulting from private sector behavior could be offset by stimulative public sector behavior. However, our current outlook is that monetary policy will be acting to inhibit growth during this period and that the effects of the final energy and Social Security bills will act to reduce incomes and increase prices, reducing total final demand. This latter effect is partially offset by a tax cut but our current expectations are that in toto the public sector will not be stimulative enough to offset the decline in growth rates of private demand.

The most dramatic shift in the composition of demand is expected to occur in housing. Following a growth of nearly 20.0% in 1977, constant dollar residential investment is expected to be nearly flat in 1978, decline 7.5% in 1979 and another 2.5% in 1980. Not until 1981 is an increase in real expenditures expected to occur. This sharp cycle is caused by a combination of higher interest rates and slower growth in real disposable income.

Less dramatic but exerting downward pressure on the economy are expected declines in the rate of inventory accumulation and a sharp slowing of the growth of plant and equipment investment. While the latter is affected by higher interest costs, the major determinant here is the general slowing in the growth of final sales.

The outlook for price inflation does not reflect the slowing of growth. Despite reduced demand pressures we expect inflation, measured by the GNP implicit deflator, to remain in the 5.5-6.5% range through 1981. In part this reflects the very slow pace at which inflation can be eased given sufficient demand pressures to stabilize the unemployment rate even at its current high level. However, it also reflects price increases that are being legislated into the economy in the form of increased taxes included in the energy and Social Security legislation included in this forecast. The impact of these increases is expected to be greatest about 1980 just as we would normally expect some slowing of inflation.

Labor Market Outlook:

Against this background we expect the unemployment rate to decline from the current level of 7.0% to the vicinity of 5.75% by mid-1982. This decline proceeds steadily despite the slowdown in the expansion rate of the economy. As Table 5.1 documents, the growth rate in employment does reflect this

TABLE 5.1

FORECAST GROWTH RATES

	GNP	Civilian Unemployment	Civilian Labor Force	Population 16 and over
1978	4.1	2.6	2.5	1.45
1979	3.8	2.3	2.0	1.43
1980	3.5	2.2	1.8	1.35
1981	3.4	2.1	1.7	1.21

decline in the rate of growth of the economy, but this is more than offset by the expected decline in the growth of the civilian labor force. As indicated by the Table, some small portion of this slowdown in labor force growth is attributable to a slowdown in population growth but the largest portion is caused by a leveling off of increases in participation rates. The aggregate participation rate (Table 1, Line 50) increases from .621 in 1977 to .628 in 1978, but increases an average of .003 in each of the succeeding years.

Of particular interest to the Navy is labor market behavior of the male 16-19 and 20-24 year old groups. The population in the 16-19 category is expected to decline throughout the forecast period. For the 20-24 group, population rises until mid-1981 after which it declines. Rising participation rates partially offset the population decline among the 16-19 age group and the labor force continues to rise into 1979, levels off, and begins to decline in late 1980. The 20-24 year old group is expected to behave in just the opposite fashion. Despite rising population, a decline in participation rates results in a labor force that peaks in late 1979 and declines beyond that point.

Reflecting the general improvement in labor market conditions we expect employment among these groups to rise throughout the forecast. Moreover, while the total number of unemployed in the economy is expected to decline by 10.0% from 1977 to 1981, the equivalent decline for males 16-19 is expected to be about 11.25% and for males 20-24 nearly 18.0%.

The implication of the forecast is that labor market competition for this prime group for military purposes is going to become more severe both as a result of the decline in population and as a result of the general improvement in the economy.

APPENDIX 5.I.

Let $Z = CD/Y$

$X = [\underline{1}\ P_r\ r]$

$Z^* = \underline{Y}x$

$X^* = [X\ Y]$

$$\underline{Y} = \begin{matrix} Y_1 & 0 & 0 \\ 0 & Y_2 \ldots & 0 \\ 0 & 0 \ldots\ldots & Y_T \end{matrix}$$

Then we can rewrite the correctly specified relationship as:

$$Z = X\beta + \varepsilon \qquad (1a)$$

and the misspecified equations:

$$Z^* = X^*\beta^* + \mu \qquad (2a)$$

where

$$\beta = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{vmatrix} \qquad \beta = \begin{vmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \\ \beta_3^* \end{vmatrix} \qquad \text{and } \mu = \underline{Y}\varepsilon$$

Assume for the sake of simplicity that X is non-stochastic if

plim $\underline{Y}\varepsilon = 0$ plim $X'\varepsilon = 0$ plim $\underline{Y} = \bar{Y}I$ and plim $YY' = \sigma^2 I_{txt}$

then applying OLS to (2a) yields

$$\hat{\beta}_3^* = -(Y'Y)^{-1}(Y'X)\ (X'MX)^{-1}X'\underline{Y}Z = (Y'Y)^{-1}(Y'\underline{Y}Z)$$
$$+ (Y'Y)^{-1}(Y'X)(X'MX)^{-1}(X'Y)(Y'Y)^{-1}(Y'\underline{Y}Z)$$

where $\quad M = I - Y(Y'Y)^{-1}Y' \qquad \text{plim}\, M = I - \frac{1}{T} I = \frac{T-1}{T} I$

and $\quad \text{plim}\,\hat{\beta}_3^* = - \frac{\bar{Y}^2}{T\sigma^2} \frac{T}{T-1} I \; i' X (X'X) \quad X'X \beta$

$$+ \frac{\sigma^2}{T\sigma^2} i'X\beta$$

$$+ \frac{\bar{Y}^2\sigma^2}{(T\sigma^2)^2 (T-1)} I \; i' X(X'X)^{-1}X'X \beta$$

$$= \frac{i'X\beta}{T} \left(- \frac{\bar{Y}^2}{\sigma^2} \frac{T}{T-1} + 1 + \frac{\bar{Y}^2}{\sigma^2} \frac{1}{T-1} \right)$$

$$= \frac{i'X\beta}{T} \left(1 + \frac{\bar{Y}^2}{\sigma^2} \frac{(1-T)}{(T-1)} \right) \sim \frac{i'X\beta}{T} \left(1 - \frac{\bar{Y}^2}{\sigma^2} \right)$$

for large T.