

Modeling Labor Force Participation



Presented by:

Vicki Myers

University at Albany, SUNY

Department of Biometry & Statistics

Background

- Data – Current Population Survey (CPS)
- Key Terms
- Variables
- Objective

Key Terms

- Civilian Non-Institutional Population
- Employed Persons
- Unemployed Persons
- Labor Force
- Participation Rate

Variables

Name

- Gender
- Race
- Age

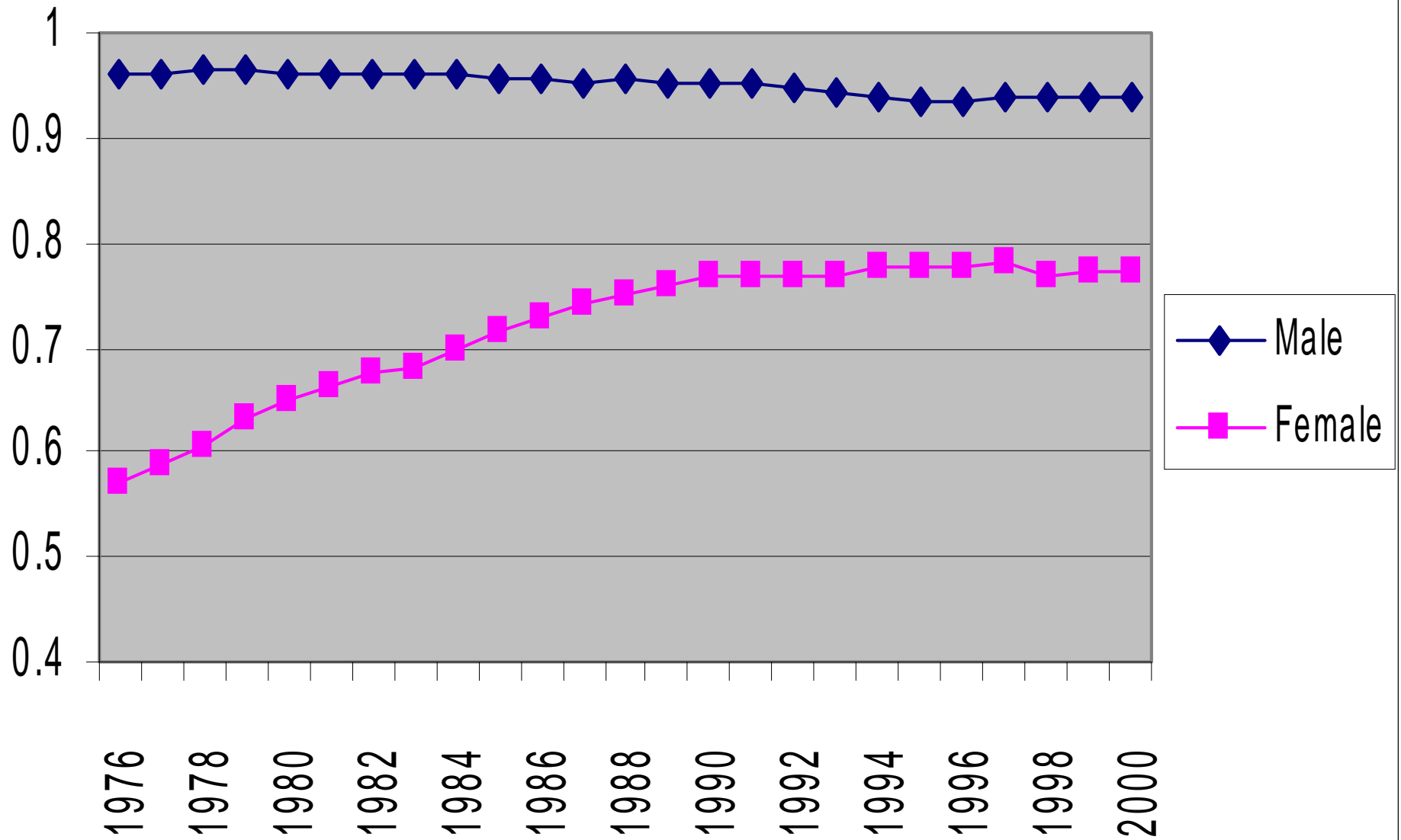
Categories

- Male & Female
- Black, White, & Other
- 16-19 20-24
25-34 35-44
45-54 55-64
65+

Objective

- Model labor force participation over time
- Use this model to predict participation into the short-term future

Observed Data - White persons aged 35-44



Model Building – Logistic Linear

Model:

$$\ln \text{ odds (time, gender)} = \beta_0 + \beta_1(\text{time}) + \beta_2(\text{gender}) + \beta_3(\text{time} * \text{gender})$$

Specifically:

$$\ln \text{ odds (male, time)} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)(\text{time})$$

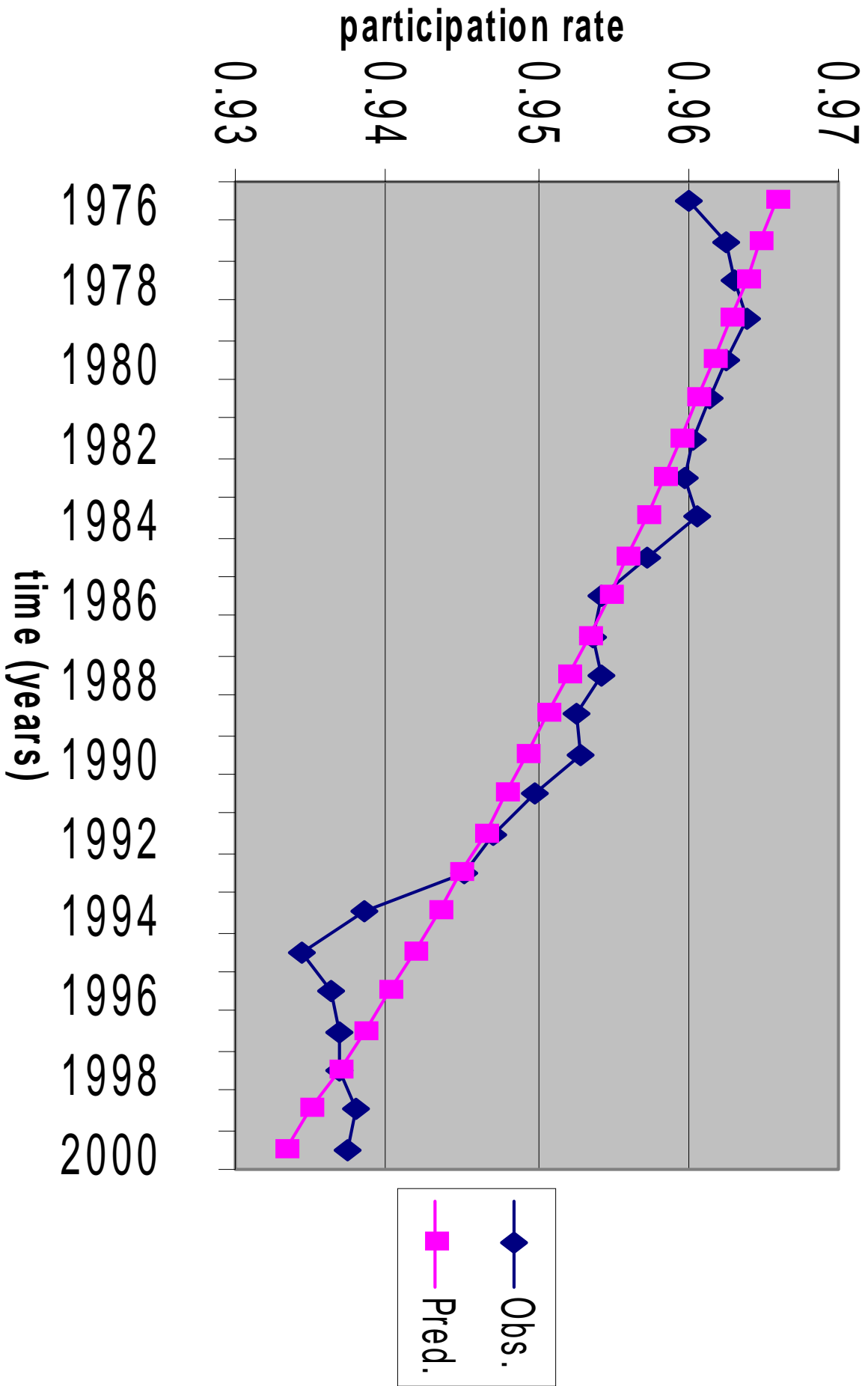
$$\ln \text{ odds (female, time)} = \beta_0 + \beta_1(\text{time})$$

Model Building – Logistic Linear

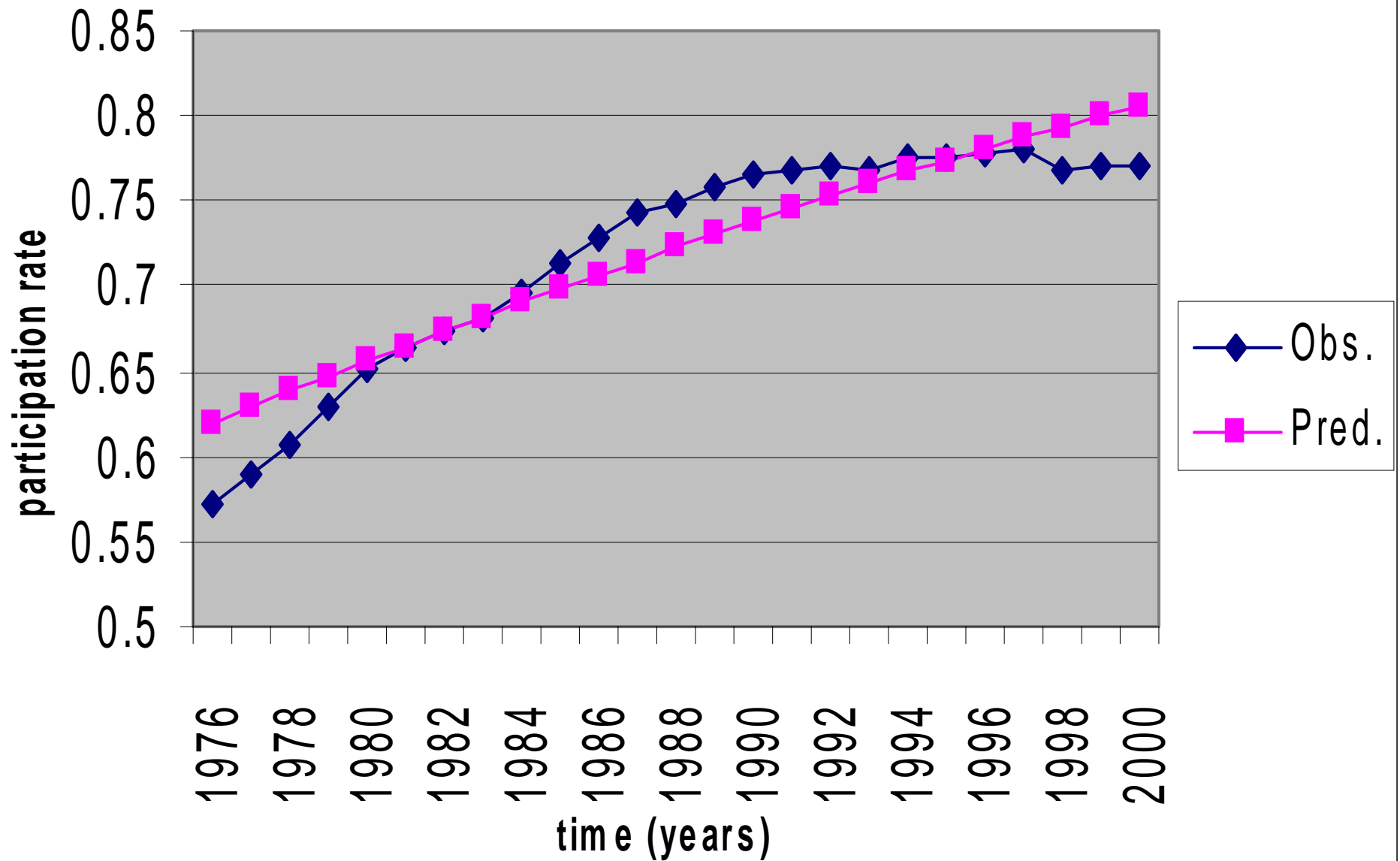
Analysis of Parameter Estimates

Parameter	Estimate	Standard Error	Chi-Square	Pr > Chi-Sq
Intercept	-2.4668	0.0471	2747.42	<.0001
time	0.0389	0.0005	5387.56	<.0001
gender	8.0437	0.1122	5139.73	<.0001
time*gender	-0.0683	0.0012	3040.44	<.0001

Males - Linear Model



Females - Linear Model



Model Building – Logistic Linear

Criteria For Assessing Goodness of Fit

Criterion	DF	Value	Value/DF	Pr > Chi-Sq
Deviance	46	985.7583	21.4295	1.428E-176
Pearson Chi-Sq	46	993.2491	21.5924	3.655E-178
Log Likelihood		-284951.1997		

Model Building – Logistic Quadratic

Model:

$$\ln \text{ odds (gender, time)} = \beta_0 + \beta_1(\text{time}) + \beta_2(\text{gender}) + \beta_3(\text{time} * \text{gender}) + \beta_4(\text{time}^2) + \beta_5(\text{gender} * \text{time}^2)$$

Specifically:

$$\ln \text{ odds (male, time)} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)(\text{time}) + (\beta_4 + \beta_5)(\text{time}^2)$$

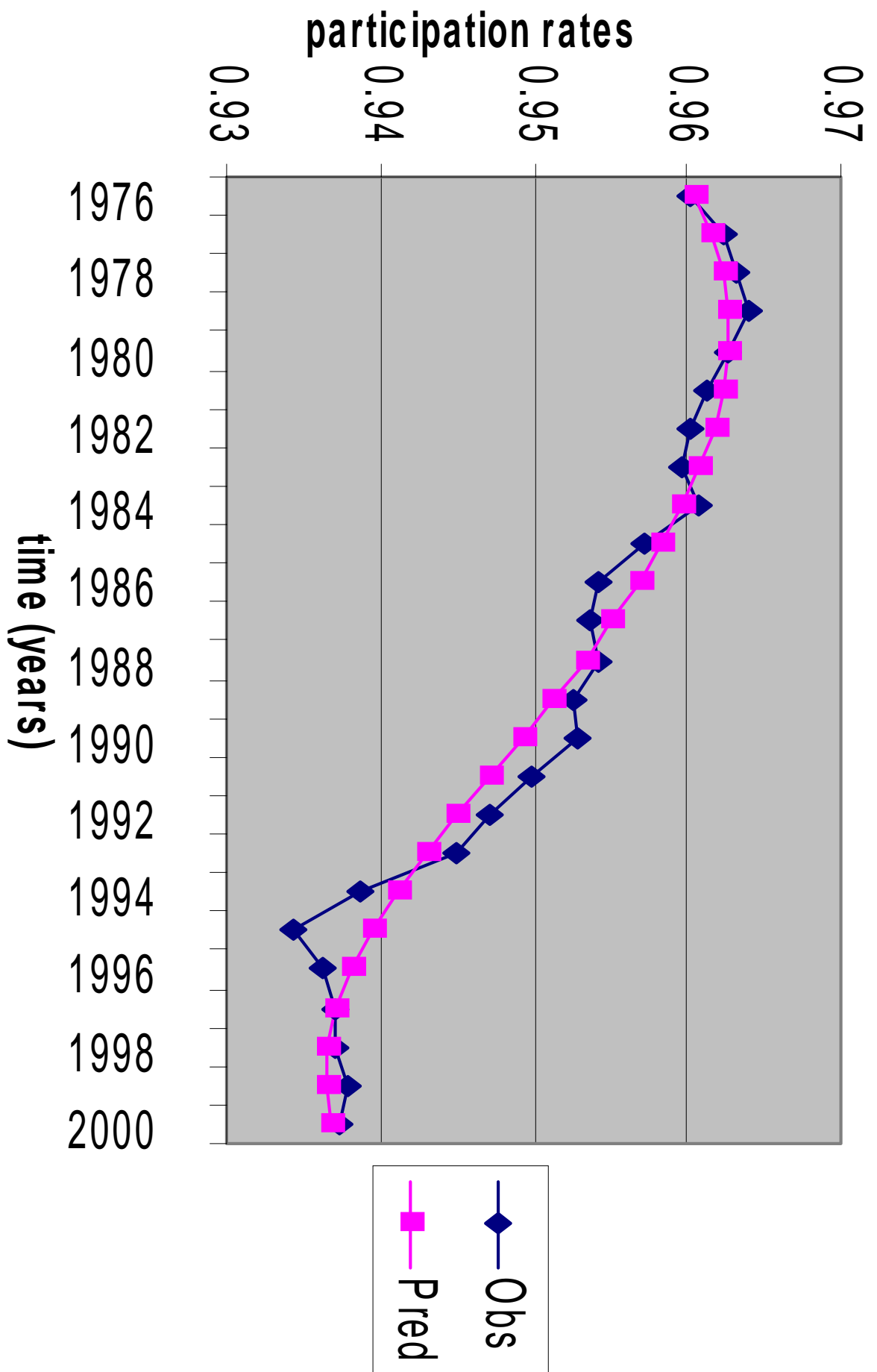
$$\ln \text{ odds (female, time)} = \beta_0 + \beta_1(\text{time}) + \beta_4(\text{time}^2)$$

Model Building – Logistic Quadratic

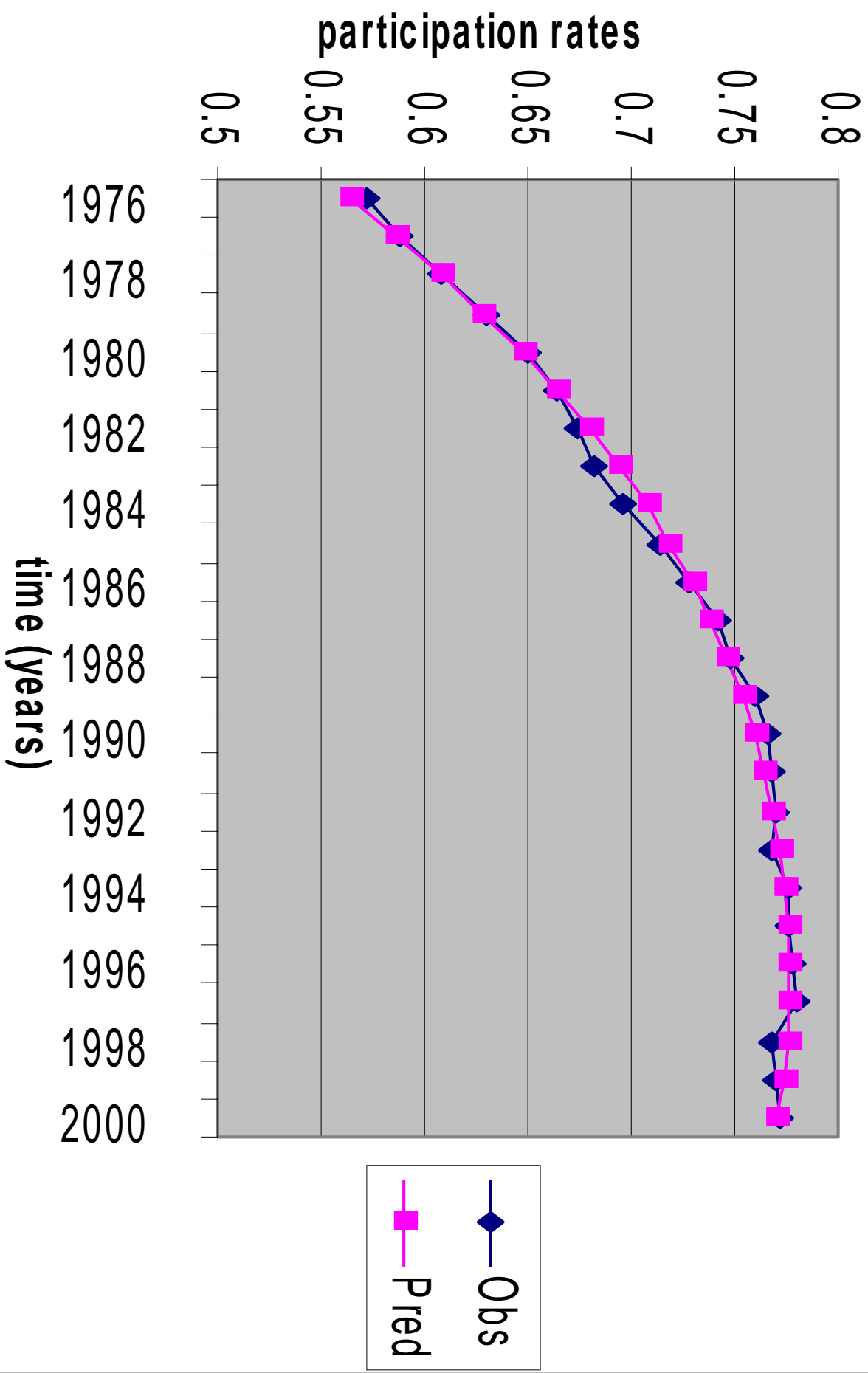
Analysis of Parameter Estimates

Parameter	Estimate	Standard Error	Chi-Square	Pr > Chi-Sq
Intercept	-21.0552	0.6258	1132.01	<.0001
time	0.4632	0.0143	1056.31	<.0001
gender	25.2276	1.4932	285.44	<.0001
time*gender	-0.4608	0.0337	186.94	<.0001
time2	-0.0024	0.0001	888.03	<.0001
gender*time2	0.0022	0.0002	138.40	<.0001

Males - Quad Model



Females - Quad Model



Model Building – Logistic Quadratic

Criteria For Assessing Goodness of Fit

Criterion	DF	Value	Value/DF	Pr > Chi-Sq
Deviance	44	102.6368	2.3327	1.379E-06
Pearson Chi-Sq	44	103.6789	2.3563	1.001E-06
Log Likelihood		-284509.6389		

Model Building – Logistic Cubic

Model:

$$\begin{aligned} \ln \text{ odds (gender, time)} = & \beta_0 + \beta_1(\text{time}) + \beta_2(\text{gender}) + \\ & \beta_3(\text{time} * \text{gender}) + \beta_4(\text{time}^2) + \beta_5(\text{gender} * \text{time}^2) \\ & + \beta_6(\text{time}^3) + \beta_7(\text{gender} * \text{time}^3) \end{aligned}$$

Specifically:

$$\begin{aligned} \ln \text{ odds (male, time)} = & (\beta_0 + \beta_2) + (\beta_1 + \beta_3)(\text{time}) + \\ & (\beta_4 + \beta_5)(\text{time}^2) + (\beta_6 + \beta_7)(\text{time}^3) \end{aligned}$$

$$\begin{aligned} \ln \text{ odds (female, time)} = & \beta_0 + \beta_1(\text{time}) + \beta_4(\text{time}^2) + \\ & \beta_6(\text{time}^3) \end{aligned}$$

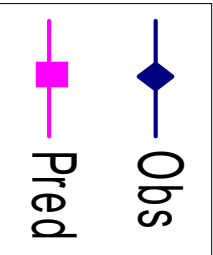
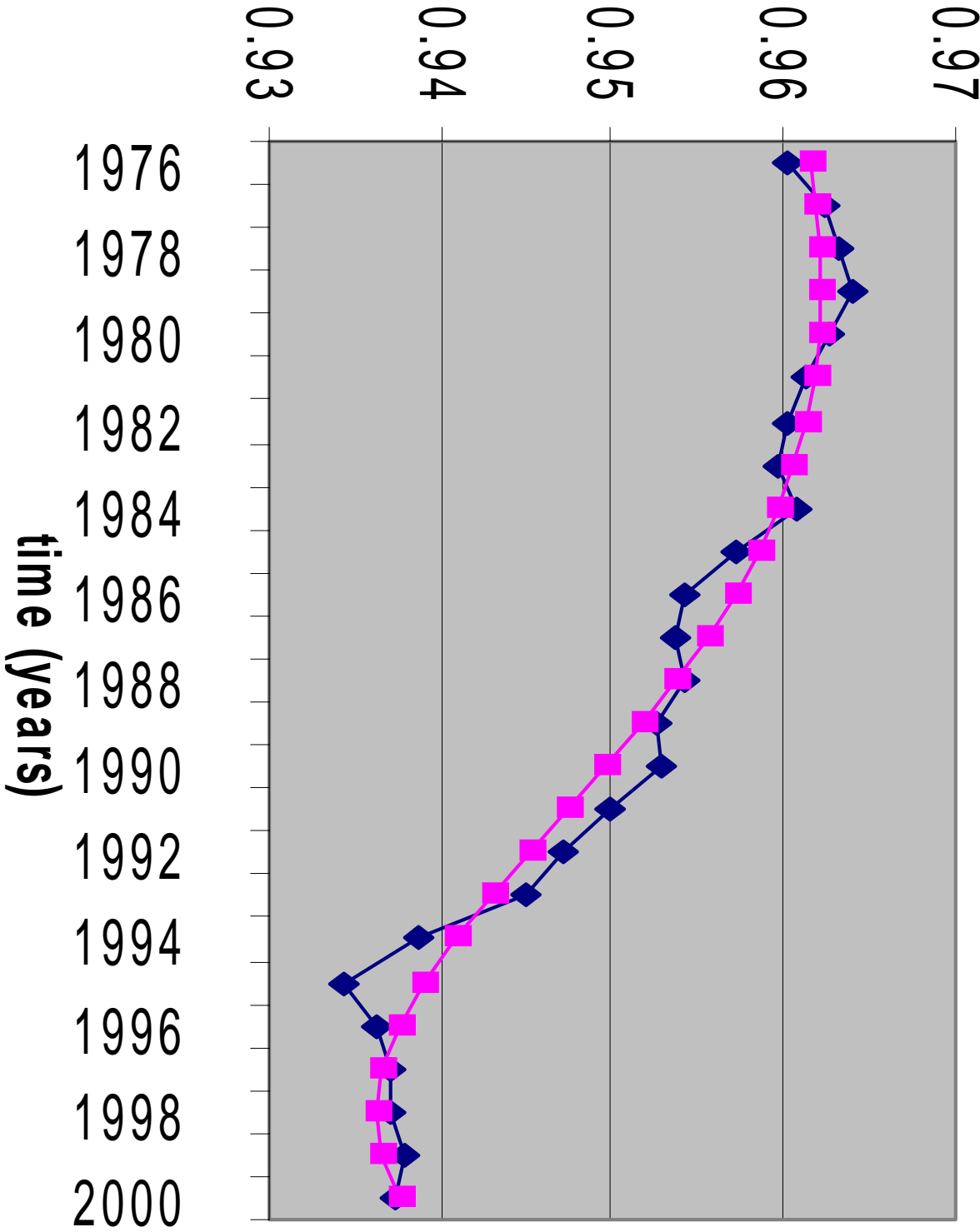
Model Building – Logistic Cubic

Analysis of Parameter Estimates

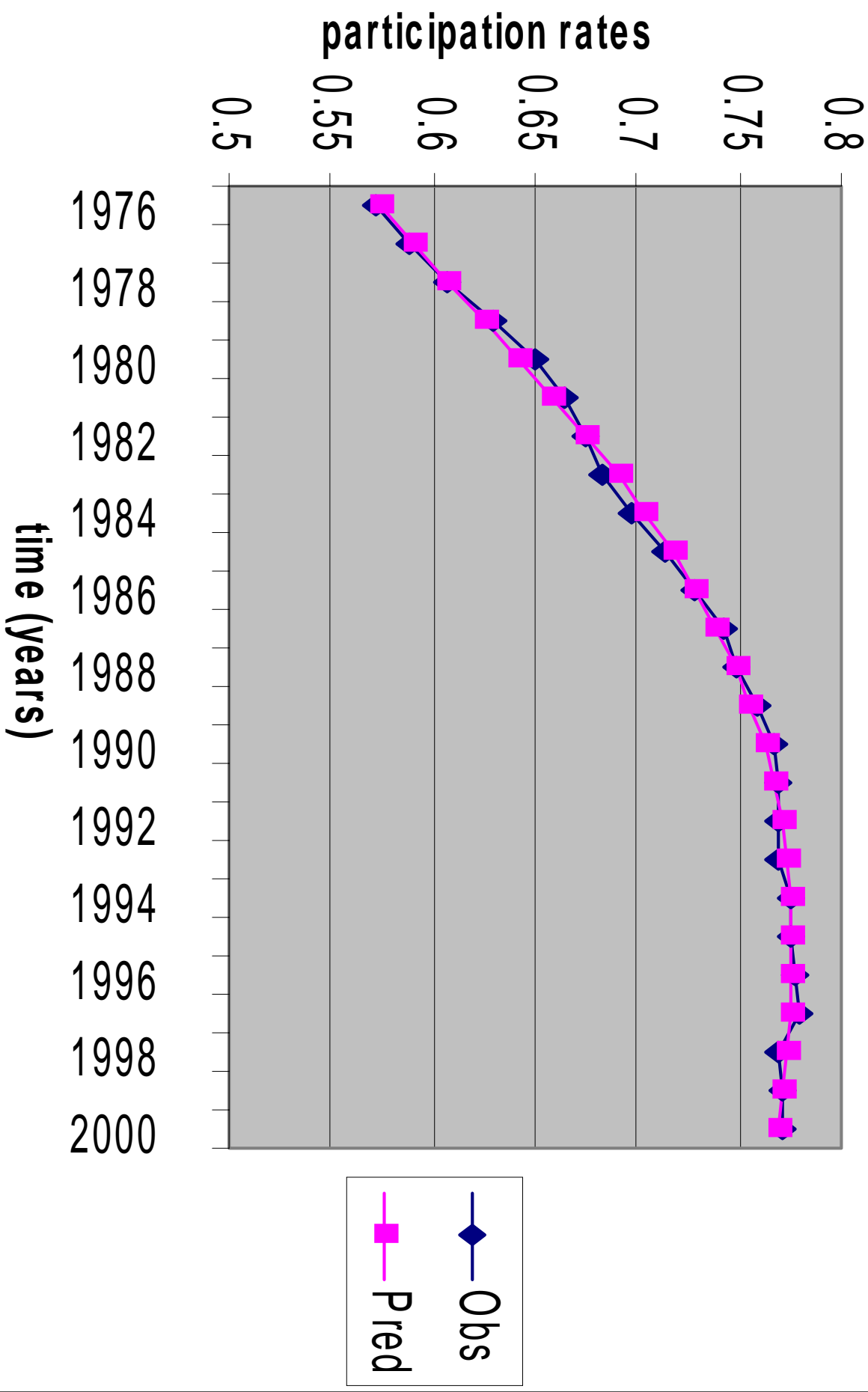
Standard Parameter	Chi- Estimate	Error	Square	Pr > Chi-Sq
Intercept	255.2808	88.9559	8.24	0.0041
time	-11.1703	3.8135	8.58	0.0034
gender	-104.087	20.7607	25.14	<.0001
time*gender	3.9538	0.7078	31.20	<.0001
time2	0.1712	0.0580	8.71	0.0032
gender*time2	-0.0478	0.0080	35.56	<.0001
time3	-0.0010	0.0003	8.61	0.0033
gender*time3	0.0000	0.0000	7.93	0.0049

participation rates

Males - Cubic Model



Females - Cubic Model



Model Building – Logistic Cubic

Criteria For Assessing Goodness of Fit

Criterion	DF	Value	Value/DF	Pr > Chi-Sq
Deviance	42	51.7716	1.2627	0.14351
Pearson Chi-Sq	42	51.9254	1.2665	0.14024
Log Likelihood		-284484.2063		

Types of Residuals

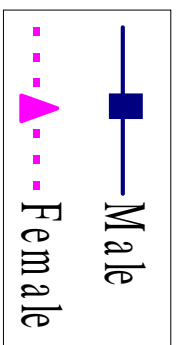
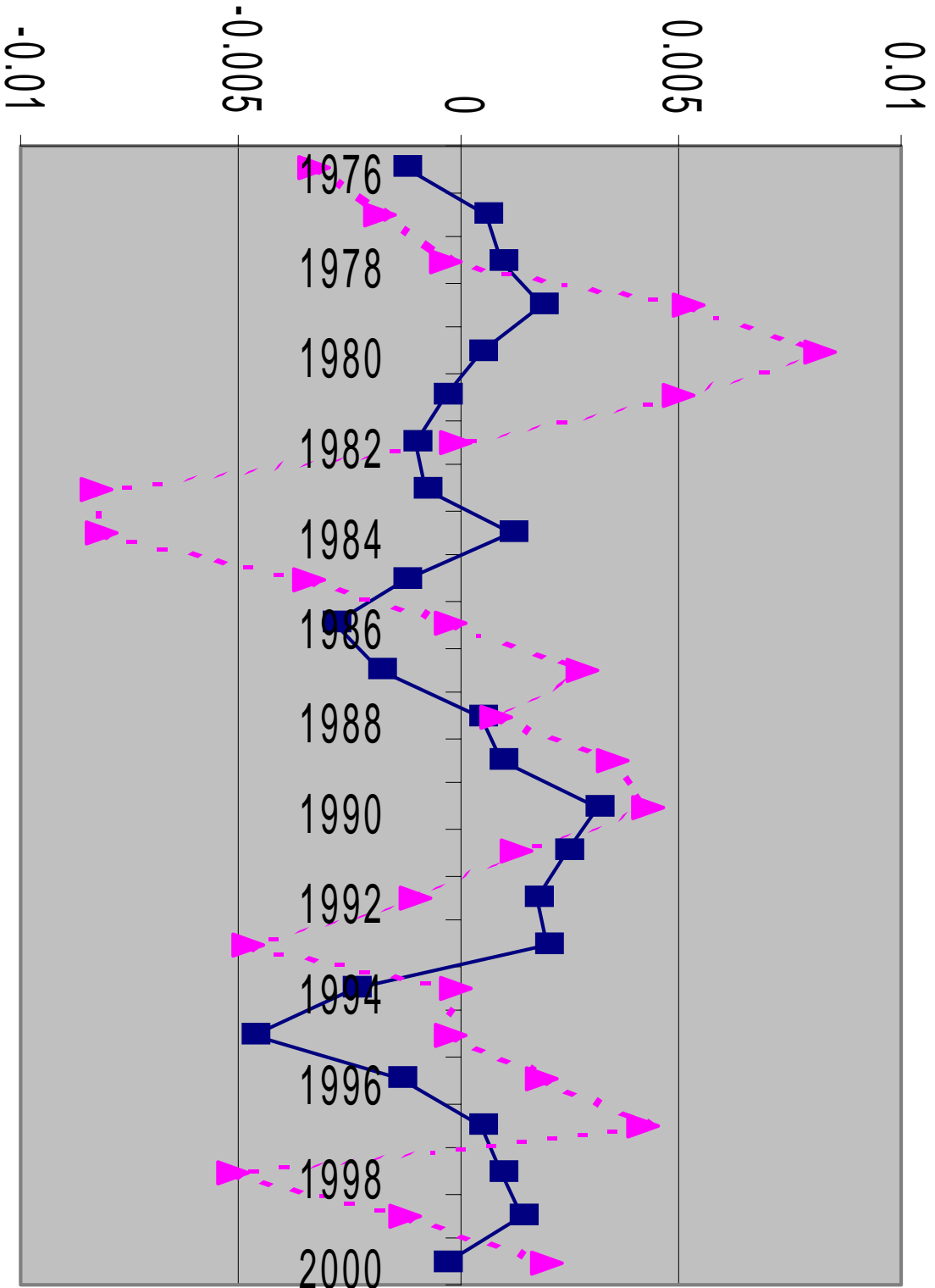
- Raw Residuals

$$r_i = p_i - \hat{p}_i$$

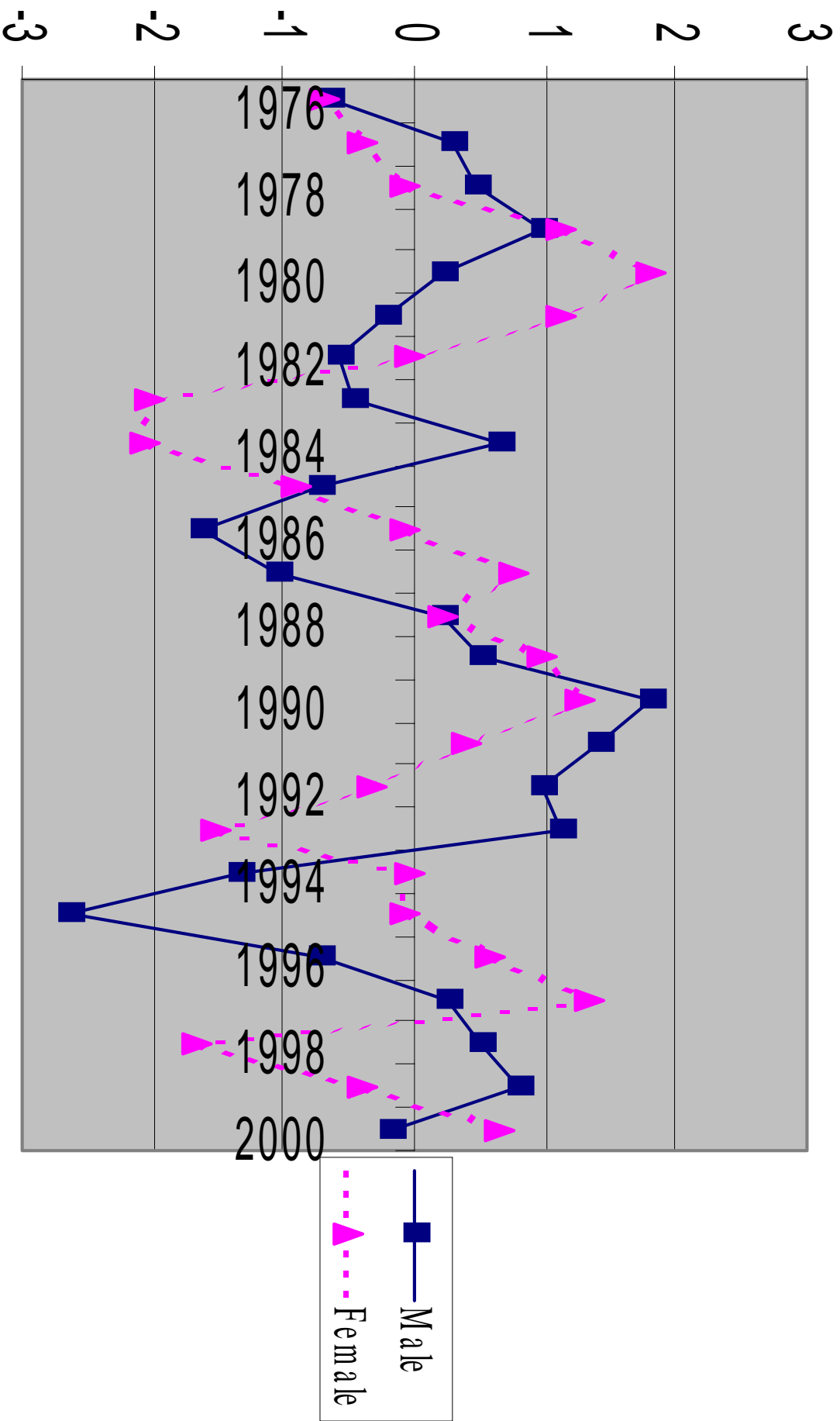
- Pearson Chi Residuals

$$S_i = \frac{x_i - n_i \hat{p}_i}{\sqrt{n_i \hat{p}_i (1 - \hat{p}_i)}}$$

Raw Residuals



Pearson Residuals



Residual Analysis

- Identify the cyclic trend

$$A \sin (B * \text{time} - C)$$

- Estimate the parameters

$$A, B, C$$

- Incorporate the trend into the log odds model

$$\begin{aligned} \ln \text{odds} (\text{gender}, \text{time}) = & \beta_0 + \beta_1(\text{gender}) + \\ & \beta_2(\text{time}) + \beta_3(\text{gender} * \text{time}) + \beta_4(\text{time}^2) + \\ & \beta_5(\text{gender} * \text{time}^2) + \beta_6(\text{time}^3) + \beta_7(\text{gender} * \text{time}^3) \\ & + A \sin (B * \text{time} - C) \end{aligned}$$

Future Work

- Explore other race & age groups
- Add cyclic term to model & assess goodness of fit
- Identify other possible models to predict labor force participation
- Find an overall model for the U.S.
- Use these models to predict participation into the future