8

Estimating and Testing the US Model

8.1 Introduction

The previous chapter discussed techniques for estimating and testing complete models, and this chapter applies these techniques to the US model. For the work in this chapter the model has been estimated by 2SLAD, 3SLS, and FIML in addition to 2SLS. 2SLAD is discussed in Section 4.4, and 3SLS and FIML are discussed in Section 7.2. Also, median unbiased (MU) estimates have been obtained for 18 lagged dependent variable coefficients using the procedure discussed in Section 7.4, and the 2SLS asymptotic distribution is compared to the exact distribution using the procedure discussed in Section 7.5. Section 8.3 presents the MU estimates; Section 8.4 examines the asymptotic distribution accuracy; and Section 8.5 compares the five sets of estimates.

The rest of this chapter is concerned with testing. In Section 8.6 the total variances discussed in Section 7.7 are computed and compared for the US, VAR5/2, VAR4, and AC models. Section 8.7 uses the procedure discussed in Section 7.8 to examine the information content of the forecasts from these models. Finally, Section 8.8 estimates event probabilities for the models and compares the accuracy of these estimates across the models using the procedure discussed in Section 7.9. A brief summary of the results is presented in Section 8.9.

Some of the tests in this chapter require a version of the US model in which there are no hard to forecast exogenous variables. This version is called US+, and it is discussed in the next section.

8.2 US+ Model

The US+ model is the US model with an additional 91 stochastic equations. Each of the additional equations explains an exogenous variable and is an eighth order autoregressive equation with a constant term and time trend added. Equations are estimated for all the exogenous variables in the model except the age variables, the dummy variables, the variables created from peak to peak interpolations, and variables that are constants or nearly constants. All the exogenous variables in the model are listed in Table A.2. Those for which autoregressive equations are *not* estimated are: $AG1$, $AG2$, $AG3$, CDA, $D691$, $D692$, $D714$, $D721$, $D794823$, $D811824$, $D831834$, $DD772$, $DELD$, $DELH$, $DELK$, HFS, HM, $IHHA$, $IKFA$, JJP, LAM, MUH, $P2554$, T, $TAUG$, $TAUS$, TI, $TXCR$, $WLDG$, and $WLDS$. Excluding these variables left 91 variables for which autoregressive equations are estimated. Logs were used for some of the variables. Logs were not used for ratios, for variables that were negative or sometimes negative, and for variables that were sometimes close to zero. The estimation technique was ordinary least squares.

The US+ model thus has no hard to forecast exogenous variables, and in this sense it is comparable to the VAR and AC models discussed in Section 7.6, which have no exogenous variables other than the constant term and time trend. Remember, however, from the discussion in Section 7.8 that this treatment of the exogenous variables may bias the results against the US model. Many of the exogenous variables may not be as uncertain as the autoregressive equations imply.

The covariance matrix of the error terms in the US+ model is 121×121, and for purposes of the stochastic simulation work it was taken to be block diagonal. The first block is the 30×30 covariance matrix of the structural error terms, and the second block is the 91×91 covariance matrix of the exogenous variable error terms. In other words, the error terms in the structural equations were assumed to be uncorrelated with the error terms in the exogenous variable equations. This assumption is consistent with the assumption in the US model that the structural error terms are uncorrelated with the exogenous variables.

8.3 MU Estimates of the US Model[1]

The procedure for obtaining median unbiased (MU) estimates of a model is explained in Section 7.4. This procedure was carried out for the US model, and the results are reported in this section. The starting point was the set of 2SLS estimates in Chapter 5. Starting from these values, median unbiased estimates of the lagged dependent variable (LDV) coefficients were obtained for 18 of the 30 stochastic equations. The estimates for the other 12 equations were fixed at their 2SLS values. The estimation period was 1954:1–1993:2, for a total of 158 observations. The number of repetitions per iteration (i.e., the value of J in step 3 in Section 7.4) was 500. After 3 iterations (i.e., after steps 3 and 4 in Section 7.4 were done 3 times), the largest difference between the successive estimates of any LDV coefficient was less than .001 in absolute value. Convergence thus occurred very quickly.[2]

The results for the LDV coefficient estimates are presented in Table 8.1. The bias for each coefficient estimate, defined as the difference between the 2SLS estimate and the MU estimate, is presented in the table. The "Andrews bias" in the table is the exact bias for an equation with a constant term, time trend, and lagged dependent variable and with the LDV coefficient equal to the 2SLS coefficient estimate presented in the table. These biases are interpolated from Table III in Andrews (1993).

Also presented in Table 8.1 are the 90 percent confidence values. The first 2SLS confidence value for each coefficient is minus 1.645 times the 2SLS estimate of the asymptotic standard error of the LDV coefficient estimate. The second 2SLS confidence value is the absolute value of the first value. The MU values are computed using the coefficient estimates from the 500 repetitions on the last iteration. The first MU confidence value for each coefficient is minus the difference between the median estimate and the estimate at which five percent of the estimates are below it. The second MU confidence value is minus the difference between the median estimate and the estimate at which five percent of the estimates are above it.

[1]The material in this section is taken from Fair (1994a). The results in this paper are the same as those in Table 8.1.

[2]To lessen stochastic simulation error, the same draws of the error terms were used for each iteration. The number of errors drawn per iteration is 2,370,000 = (500 repetitions) $\times$ (30 stochastic equations) $\times$ (158 observations). The model is solved dynamically over the estimation period for each repetition, and each of the 18 equations is estimated for each repetition.

Table 8.1
Estimated Bias of 2SLS Lagged Dependent Variable Coefficient Estimates

Eq.	2SLS	Bias	Andrews Bias	90% Confidence Values 2SLS[a]		MU[b]	
1. CS	.943	-.012	-.040	-.052	.052	-.033	.025
2. CN	.620	-.029	-.027	-.070	.070	-.074	.060
3. CD	.575	-.025	-.025	-.104	.104	-.094	.079
4. IHH	.532	-.020	-.025	-.091	.091	-.104	.084
5. L1	.776	-.049	-.031	-.082	.082	-.104	.078
6. L2	.987	-.003	-.051	-.008	.008	-.017	.011
7. L3	.890	-.040	-.036	-.059	.059	-.081	.050
8. LM	.863	-.027	-.034	-.055	.055	-.077	.047
9. MH	.896	-.050	-.036	-.064	.064	-.083	.053
10. PF	.919	-.002	-.036	-.010	.010	-.010	.009
11. Y	.293	-.000	-.020	-.074	.074	-.059	.055
12. IKF	-.040	.000	-.012	-.022	.022	-.020	.017
17. MF	.904	-.027	-.036	-.048	.048	-.067	.042
23. RB	.881	-.002	-.035	-.034	.034	-.035	.027
24. RM	.842	-.003	-.033	-.042	.042	-.048	.034
26. CUR	.957	-.003	-.043	-.018	.018	-.016	.012
27. IM	.872	-.032	-.034	-.054	.054	-.071	.053
30. RS	.892	-.003	-.035	-.031	.031	-.035	.027
Average		-.018	-.033	-.051	.051	-.057	.042

[a]The first number for 2SLS is minus 1.645 times the 2SLS estimate of the standard error of the LDV coefficient estimate. The second number for 2SLS is the absolute value of the first number.

[b]The first number for MU is minus the difference between the median estimate and the estimate at which five percent of the estimates are below it. The second number for MU is minus the difference between the median estimate and the estimate at which five percent of the estimates are above it.

The results in Table 8.1 show that the estimated biases are zero to three decimal places for 2 of the 18 coefficients and negative for the rest. The average bias across the 18 estimates is $-.018$. The average Andrews bias, on the other hand, is $-.033$, and so the results suggest that the bias of a typical macroeconometric equation is on average less than the bias of an equation that includes only a constant term, time trend, and lagged dependent variable. In only four cases in the table is the Andrews bias smaller in absolute value—equations 2, 5, 7, and 9.

The 2SLS and MU confidence values in Table 8.1 are fairly similar. The average of the left tail values is $-.057$ for MU and $-.051$ for 2SLS. The average of the right tail values is .042 for MU and .051 for 2SLS. It is clear that the MU confidence interval is not symmetric around the median estimate. For all the coefficient estimates the right tail value is less than the left tail value in absolute value. The left tail of the distribution is thus thicker than the right tail, although the differences are fairly minor.

An interesting question is whether the biases in Table 8.1 are quantitatively important regarding the properties of the model. This question is examined in Sections 8.5 and 11.3.5. In Section 8.5 the sensitivity of the predictive accuracy of the model to the use of the MU estimates is examined, and in Section 11.3.5 the sensitivity of the multiplier properties of the model to the use of the estimates is examined. It will be seen that the use of the MU estimates has little effect on the predictive accuracy of the model and on its multiplier properties. These results thus suggest that macroeconometric model builders have not missed much by ignoring the Orcutt and Hurwicz warnings 40 years ago, although work with other models should be done to see if the present results hold up. With hindsight, the present results are perhaps not surprising. What they basically say is that if one changes a LDV coefficient estimate by about half of its estimated standard error and then reestimates the other coefficients in the equation to reflect this change, the fit and properties of the equation do not change very much. This is something that most model builders probably know from experience.

8.4 Asymptotic Distribution Accuracy[3]

The procedure for examining the accuracy of asymptotic distributions was discussed in Section 7.5. It is carried out in this section for the US model. Again, the 2SLS estimates in Chapter 5 were used as the base estimates. For the present results the US model was simulated and estimated 800 times. There are 166 coefficients to estimate in the model, and so the results from this exercise consist of 800 values of 166 coefficients. A summary of these results is presented in Table 8.2. Detailed results are presented for the same 18 coefficients that were examined in Table 8.1, namely the LDV coefficients of the 18 equations, and summary results are presented for all 166 coefficients.

The bias results for the 18 coefficients show, as in Table 8.1, that the 2SLS estimates of the LDV coefficients are biased downwards,[4] with the average

[3]The material in this section is also taken from Fair (1994a). The results in this paper are the same as those in Table 8.2.

[4]The bias estimates are slightly different in Table 8.2 than in Table 8.1 because they are

Table 8.2
Asymptotic Distribution Accuracy

			Med.-	Left Tail			Right Tail		
Eq.	2SLS	Med.	2SLS	5	10	20	5	10	20
1. CS	.943	.931	-.012	0.4	1.8	7.9	0.0	0.4	3.3
2. CN	.620	.595	-.025	4.0	10.1	19.5	1.6	4.8	14.6
3. CD	.575	.554	-.021	3.8	8.6	17.6	0.8	4.9	11.1
4. IHH	.532	.516	-.016	8.3	12.5	22.0	1.9	7.5	18.0
5. L1	.776	.731	-.045	8.9	13.6	22.0	5.6	11.6	22.3
6. L2	.987	.984	-.003	13.9	18.5	26.8	8.9	13.8	24.4
7. L3	.890	.856	-.034	9.4	13.8	22.9	1.4	5.8	15.9
8. LM	.863	.839	-.024	9.3	15.6	24.8	2.5	8.1	19.8
9. MH	.896	.850	-.046	9.4	15.1	23.3	2.5	8.3	18.1
10. PF	.919	.919	-.000	6.6	11.5	20.8	2.8	7.5	17.6
11. Y	.293	.292	-.001	2.3	6.0	12.0	1.4	5.0	14.4
12. IKF	-.040	-.039	.001	2.4	6.3	14.4	2.1	5.6	15.0
17. MF	.904	.882	-.022	13.1	18.8	27.1	3.4	9.1	21.1
23. RB	.881	.877	-.004	5.8	10.0	20.1	3.8	8.4	18.5
24. RM	.842	.836	-.006	6.1	11.6	20.3	4.6	8.5	19.0
26. CUR	.957	.954	-.003	4.0	7.5	14.9	1.8	4.4	11.8
27. IM	.872	.846	-.026	10.5	16.4	25.4	4.6	10.8	23.9
30. RS	.892	.889	-.003	6.9	12.4	21.6	3.0	7.4	17.9
MEAN(18)			-.016	5.5	9.9	19.4	4.0	8.1	17.4
MAE(18)				3.3	3.9	4.2	2.5	3.4	4.4
MEAN(166)				5.0	9.3	18.3	4.4	8.7	17.9
MAE(166)				2.8	3.6	4.3	2.4	3.2	4.1

bias being $-.016$. This is as expected.

The main point of Table 8.2 is to compare the left tail and right tail estimated probabilities to the values implied by the asymptotic distribution. Let p_{ik} be the estimated probability for coefficient i for the asymptotic value of k percent. Remember from Section 7.5 how these percentages are computed. Given for a particular coefficient estimate the 2SLS estimate of its asymptotic standard error, one can compute the value above which k percent of the coefficient estimates should lie if the asymptotic standard error is accurate. For k equal to 20, this value is the median plus 0.84 times the estimated asymptotic standard error. For k equal to 10 the multiplier is 1.28, and for k equal to 5 the multiplier is 1.64. From the 800 coefficient estimates one can compute the actual percent of the coefficient estimates that lie above this value. These are the right tail

based on 800 rather than 500 repetitions and because the iterations done for the results in Table 8.1 were not done for the results in Table 8.2.

percents. A similar procedure can be followed for the left tail percents. For each tail and each coefficient i, one can thus compute values of p_{i5}, p_{i10}, and p_{i20}. Values of these probabilities for each tail are presented in Table 8.2 for the 18 LDV coefficient estimates. Also reported in the table are the means of the probabilities across the 18 coefficients and across the 166 coefficients. In addition, the mean absolute errors around the means are presented for the 18 and 166 coefficients. For example, the mean absolute error for the left tail p_{i5} for the 18 coefficients is the sum of $|p_{i5} - 5.5|$ across the 18 coefficients divided by 18, where 5.5 is the mean.

Consider the results for the 166 coefficients in Table 8.2. The means of the 5, 10, and 20 percent left tail values are 5.0, 9.3, and 18.3, with mean absolute errors of 2.8, 3.6, and 4.3, respectively. The corresponding right tail means are 4.4, 8.7, and 17.9, with mean absolute errors of 2.4, 3.2, and 4.1, respectively. These mean values are less than the asymptotic values (except for the equality for the 5 percent left tail value), and so on average the asymptotic distribution has thicker tails than does the exact distribution. These differences are, however, fairly small. In general the asymptotic distribution seems to be a good approximation, although the mean absolute errors reveal that there is some dispersion across the coefficients. The overall results suggest that the use of the asymptotic distribution is not in general likely to give misleading conclusions.

The closeness of the asymptotic distribution to the exact distribution is an important result. If this result holds up for other models, it means that the unit root problems that have received so much attention in the econometric literature are not likely to be of much concern to macro model builders. While the existence of unit roots can in theory cause the asymptotic distributions that are relied on in macroeconometrics to be way off, in practice the asymptotic distributions seem fairly good.

8.5 A Comparison of the Estimates

Section 8.3 examined the closeness of the 2SLS and MU estimates. This section compares the closeness of the 2SLS, 2SLAD, 3SLS, and FIML estimates. It also compares the predictive accuracy of the model for all five sets of estimates.

The first step for the results in this section was to compute the 2SLAD, 3SLS, and FIML estimates. There are some computational tricks that are needed to obtain these estimates. These tricks are discussed in Fair (1984),

Table 8.3
Comparison of 2SLS, 2SLAD, 3SLS, and FIML Estimates

	Number of estimates greater than .5, 1.0, 1.5, 2.0, and 3.0 standard errors away from the 2SLS estimates					Number of sign changes from 2SLS estimates
137 Coefficients:						
	.5	1.0	1.5	2.0	3.0	
3SLS	69	22	4	2	0	2
FIML	101	70	51	32	13	6
166 Coefficients:						
2SLAD	62	16	4	2	1	3
Average ratio of 2SLS standard error to 3SLS standard error (137 coefficients)					= 1.28	
Average ratio of 3SLS standard error to FIML standard error (137 coefficients)					= 0.81	

and this discussion will not be repeated here.[5] Of the 166 coefficients, 137 were estimated by 3SLS and FIML, with the remaining coefficients being fixed at their 2SLS values.[6] All 166 coefficients were estimated by 2SLAD. The first stage regressors that were used for 3SLS are listed in Table A.7 in Appendix A.[7] The same first stage regressors were used for 2SLAD as were used for 2SLS, and these are also listed in Table A.7.

A comparison of the four sets of estimates is presented in Table 8.3. The main conclusion from this comparison is that the estimates are fairly close

[5]The 2SLAD computational problem is discussed in Section 6.5.4, the 3SLS problem in Section 6.5.3, and the FIML problem in Section 6.5.2 in Fair (1984). The Parke (1982) algorithm was used for the 3SLS and FIML estimates.

[6]The equations whose coefficients were fixed for 3SLS and FIML are 15, 18, 19, 20, 21, 25, 28, and 29. (Remember that the coefficients for equations 19 and 29 were obtained in the manner discussed in Section 5.9 rather than by 2SLS.) In addition, the following other coefficients were fixed: the two autoregressive coefficients in equation 4, the coefficients of T and $DD772 \cdot T$ in equations 13 and 14, and the four dummy variable coefficients in equation 27. These coefficients were fixed to lessen potential collinearity problems. See Fair (1984), Section 6.4, for a discussion of sample size requirements and the estimation of subsets of coefficients.

[7]The choice of first stage regressors for 3SLS is discussed in Fair (1984), Section 6.3.3.

to each other, with the FIML estimates being the farthest apart. Of the 137 3SLS estimates, only 22 were greater than one 2SLS standard error away from the 2SLS estimate, and only 2 were greater than two standard errors. For the FIML estimates, 70 were greater than one standard error away from the 2SLS estimate, and 32 were greater than two standard errors. Of the 166 2SLAD estimates, 16 were greater than one standard error away from the 2SLS estimate, and 2 were greater than two standard errors. There were 2 sign changes for 3SLS, 6 for FIML, and 3 for 2SLAD. The closeness of these estimates is encouraging, since one would not expect for a correctly specified model that the use of different consistent estimators would result in large differences in the estimates.

The second to last result in Table 8.3 shows the efficiency gained from using 3SLS over 2SLS. The average ratio of the 2SLS standard error to the 3SLS standard error across the 137 coefficients is 1.28. In other words, the 2SLS standard errors are on average 28 percent larger than the 3SLS standard errors.

The last result in Table 8.3 shows that the 3SLS standard errors are on average smaller than the FIML standard errors. The average ratio of the 3SLS standard error to the FIML standard error across the 137 coefficients is .81. In other words, the 3SLS standard errors are on average 19 percent smaller than the FIML standard errors. The smaller 3SLS than FIML standard errors is a typical result, and a possible reason for it is discussed in Fair (1984), pp. 245–246. This discussion will not be repeated here.

Another way to compare the different sets of coefficient estimates is to examine the sensitivity of the predictive accuracy of the model to the different sets. This examination is presented in Table 8.4. One, two, three, four, six, and eight quarter ahead RMSEs are presented for four variables for each set of estimates. The prediction period is the same as the estimation period, namely 1954:1–1993:2. These predictions are all within sample predictions.[8] There are 158 one quarter ahead predictions, 157 two quarter ahead predictions, and so on through 151 eight quarter ahead predictions, where each of the 158

[8] If different models were being compared, the use of RMSEs in the manner done here would not be appropriate and one should use a method like the one in the next section. The RMSE procedure ignores exogenous variable differences and possible misspecifications. These problems are less serious when it is simply different estimates of the same model being used. There are no exogenous variable differences except for the fact that different coefficients multiply the same exogenous variables across versions. There are also no specification differences, and so misspecification effects differ only to the extent that misspecification is differentially affected by the size of the coefficients across versions.

Table 8.4
RMSEs for Five Sets of Coefficient Estimates for 1954:1–1993:2 for the US Model

	Number of Quarters Ahead					
	1	2	3	4	6	8
			$GDPR$: Real GDP			
2SLS	0.69	1.05	1.30	1.45	1.55	1.59
2SLAD	0.69	1.07	1.36	1.54	1.72	1.77
3SLS	0.68	1.02	1.27	1.42	1.53	1.58
FIML	0.70	1.02	1.24	1.40	1.56	1.68
MUE	0.68	1.04	1.28	1.42	1.52	1.54
			$GDPD$: GDP Deflator			
2SLS	0.40	0.60	0.78	0.97	1.29	1.52
2SLAD	0.40	0.60	0.78	0.98	1.33	1.60
3SLS	0.40	0.62	0.81	1.00	1.34	1.58
FIML	0.52	0.90	1.28	1.64	2.28	2.80
MUE	0.40	0.60	0.78	0.97	1.29	1.53
			UR: Unemployment Rate			
2SLS	0.30	0.56	0.73	0.87	1.02	1.06
2SLAD	0.30	0.57	0.75	0.90	1.09	1.16
3SLS	0.29	0.52	0.68	0.79	0.91	0.95
FIML	0.32	0.58	0.76	0.90	1.03	1.11
MUE	0.30	0.57	0.75	0.89	1.05	1.10
			RS: Bill Rate			
2SLS	0.54	1.02	1.20	1.40	1.62	1.72
2SLAD	0.54	1.01	1.20	1.42	1.67	1.78
3SLS	0.55	0.98	1.15	1.33	1.52	1.58
FIML	0.63	1.06	1.28	1.46	1.71	1.82
MUE	0.55	1.03	1.21	1.39	1.61	1.71

Errors are in percentage points.

simulations is based on a different starting point.

The results in Table 8.4 show that the RMSEs are very similar across the five sets of estimates. No one set of estimates dominates the others, and in general the differences are quite small. The largest differences occur for the FIML predictions of the price deflator, which are noticeably less accurate than the others. My experience with the FIML estimation of macroeconometric models is that FIML estimates are the most likely to differ in large ways from other estimates and that when they do differ they generally lead to a poorer fitting model. For example, 3SLS estimates are generally closer to 2SLS estimates than are FIML estimates, and they tend to lead to a better fitting overall model. The 3SLS estimates in Table 8.4 do in fact quite well. They are slightly worse than the 2SLS estimates for the price deflator, but slightly better for the other three variables. Again, however, these differences are small.

The closeness of the results in Table 8.4 is again encouraging, since one would not expect there to be large differences of this sort for a model that is a good approximation of the economy.

The fact that the MU results are similar to the others in Table 8.4 is consistent with the properties of a simple equation with only the lagged dependent variable as an explanatory variable, say $y_t = \alpha y_{t-1} + \epsilon_t$. Malinvaud (1970), p. 554, shows for this equation that the expected value of the prediction error is zero when the distribution of ϵ_t is symmetric even if the estimate of α that is used to make the prediction is biased. The present results show that even for much more complicated models, prediction errors seem to be little affected by coefficient estimation bias.

8.6 Predictive Accuracy

This section uses the method discussed in Section 7.7 to compare the US model to the VAR5/2, VAR4, and AC models. The latter three models are discussed in Section 7.6. The method computes forecast error variances for each variable and period ahead that account for the four main sources of uncertainty of a forecast. The variances can thus be compared across models. The results for the four models are presented in Table 8.5 for four variables: real GDP, the GDP deflator, the unemployment rate, and the bill rate. Standard errors rather than variances are presented in the table because the units are easier to interpret.

There are considerable computations behind the results in Table 8.5, and most of this section is a discussion of this table. Consider the a and b rows for the US model first. The simulation period was 1991:1–1992:4, and 1000 repetitions were made for each row. For the a row, only the structural error terms were drawn, and for the b row, both the structural error terms and the coefficients were drawn. In the notation in Section 7.7, each value in a b row is the square root of $\tilde{\sigma}^2_{itk}$.

The 2SLS estimates in Chapter 5 were used for this work. The estimated covariance matrix of the error terms, $\hat{\Sigma}$, is 30×30. Remember from the discussion at the end of Section 5.9 that equations 19 and 29 are taken to be

Table 8.5
Estimated Standard Errors of Forecasts for Four Models

	1991				1992			
	1	2	3	4	1	2	3	4
				$GDPR$: Real GDP				
US:								
a	.61	.98	1.29	1.49	1.62	1.70	1.78	1.81
b	.63	1.03	1.36	1.58	1.74	1.84	1.93	1.98
c	.72	1.22	1.64	1.95	2.20	2.38	2.48	2.52
d	.86	1.52	2.14	2.56	2.86	2.98	3.05	3.07
VAR5/2:								
a	.80	1.20	1.44	1.55	1.69	1.86	2.04	2.21
b	.83	1.24	1.53	1.77	1.99	2.22	2.42	2.65
d	.96	1.73	2.23	2.62	2.80	2.90	2.93	2.97
VAR4:								
a	.75	1.15	1.40	1.47	1.60	1.74	1.91	2.07
b	.82	1.32	1.57	1.71	1.94	2.12	2.32	2.49
d	1.08	2.01	2.45	2.91	3.35	3.64	3.82	3.89
AC:								
a	.51	.80	.99	1.18	1.34	1.42	1.49	1.53
b	.52	.87	1.15	1.36	1.51	1.64	1.74	1.81
d	.73	1.18	1.61	1.91	2.17	2.39	2.64	2.85
				$GDPD$: GDP Deflator				
US:								
a	.34	.51	.64	.74	.82	.89	.97	1.05
b	.36	.56	.69	.79	.87	.99	1.10	1.18
c	.48	.73	.92	1.08	1.20	1.32	1.41	1.52
d	.43	.70	.92	1.14	1.40	1.70	2.00	2.33
VAR5/2:								
a	.27	.40	.53	.67	.84	1.01	1.17	1.32
b	.27	.44	.60	.78	.97	1.18	1.42	1.64
d	.29	.58	.80	1.05	1.36	1.75	2.14	2.53
VAR4:								
a	.30	.44	.58	.75	.96	1.17	1.36	1.55
b	.31	.49	.65	.88	1.14	1.38	1.64	1.93
d	.33	.62	.86	1.14	1.49	1.89	2.31	2.77

Table 8.5 (continued)

	1991				1992			
	1	2	3	4	1	2	3	4
			UR: Unemployment Rate					
US:								
a	.27	.44	.58	.70	.80	.87	.96	1.03
b	.31	.49	.64	.77	.90	.98	1.07	1.14
c	.31	.52	.70	.87	1.02	1.13	1.21	1.30
d	.27	.55	.79	1.03	1.22	1.30	1.31	1.28
VAR5/2:								
a	.24	.44	.58	.66	.71	.76	.83	.90
b	.25	.47	.63	.75	.85	.93	.99	1.07
d	.29	.60	.86	1.08	1.23	1.30	1.30	1.27
VAR4:								
a	.23	.42	.54	.62	.65	.69	.75	.81
b	.24	.46	.62	.71	.79	.84	.91	.96
d	.34	.72	1.00	1.24	1.45	1.55	1.59	1.54
				RS: Bill Rate				
US:								
a	.56	.87	1.01	1.11	1.18	1.23	1.30	1.37
b	.54	.89	1.07	1.14	1.24	1.37	1.47	1.53
c	.57	.96	1.17	1.32	1.47	1.60	1.75	1.85
d	.82	1.57	1.88	2.28	2.74	3.03	3.35	3.63
VAR5/2:								
a	.67	1.08	1.24	1.35	1.46	1.53	1.63	1.65
b	.66	1.11	1.33	1.53	1.72	1.87	1.95	2.00
d	1.15	2.02	2.46	3.01	3.58	4.02	4.52	4.87
VAR4:								
a	.63	1.03	1.21	1.34	1.45	1.52	1.63	1.67
b	.65	1.12	1.37	1.56	1.74	1.91	2.02	2.06
d	1.14	2.11	2.51	3.05	3.77	4.40	4.91	5.31

a = Uncertainty due to error terms.
b = Uncertainty due to error terms and coefficient estimates.
c = Uncertainty due to error terms, coefficient estimates, and exogenous variable forecasts.
d = Uncertainty due to error terms, coefficient estimates, exogenous variable forecasts, and the possible misspecification of the model.
Errors are in percentage points.

stochastic for purposes of computing $\hat{\Sigma}$ even though their coefficients are not estimated in a traditional way. Remember also that equation 19 is divided through by $|AF + 10|$ and that equation 29 is divided through by $|AG|$ before computing the error terms to be used in computing $\hat{\Sigma}$.

The estimation period for $\hat{\Sigma}$ was 1954:1–1993:2. This is the estimation period used for estimating all the equations except 15, which explains HO. The estimation period for equation 15 begins in 1956:1 rather than 1954:1. However, for purposes of computing $\hat{\Sigma}$, the period beginning in 1954:1 was used for equation 15. Data for HO prior to 1956:1 were constructed in the manner discussed in Section 3.2.3.

The estimated covariance matrix of the coefficient estimates, $\hat{V}_2$, is 166×166. The formula for this matrix is given in equation 4.5 in Chapter 4. For purposes of computing $\hat{V}_2$, the coefficients in equations 19 and 29 were taken to be fixed. There are five of these coefficients. Also, four of the coefficients in the wage equation 16 are constrained and thus not freely estimated. There are thus a total of 175 coefficients in the model, but only 166 freely estimated. The dimension of $\hat{V}_2$ is thus 166×166 rather than 175×175.

Consider next the c row for the US model. For this row, structural errors, coefficients, *and* exogenous variable errors were drawn, and again 1000 repetitions were made. The procedure that was used for the exogenous variable errors is the following. First, an eighth order autoregressive equation with a constant and time trend was estimated for each of 91 exogenous variables. These are the same equations that are used for the US+ model discussed in Section 8.2 except that all the equations here are linear whereas many of the equations for US+ are in logs. The estimation period was 1954:1–1993:2. Let $\hat{s}_i$ denote the estimated standard error from the equation for exogenous variable i. Let v_{it} be a normally distributed random variable with mean zero and variance $\hat{s}_i^2$: $v_{it} \sim N(0, \hat{s}_i^2)$ for all t. Let x_{it}^a be the actual value of exogenous variable i for period t. Finally, let x_{it}^* be the value of variable i used for a given repetition. Then for prediction period 1 through T, the values for x_{it}^* for a given repetition were taken to be

$$x_{i1}^* = x_{i1}^a + v_{i1}$$

$$x_{i2}^* = x_{i2}^a + v_{i1} + v_{i2}$$

$$\cdot$$

$$\cdot$$

$$x_{iT}^* = x_{iT}^a + v_{i1} + v_{i2} + \cdots + v_{iT}$$

where each v_{it} $(t = 1, \cdots, T)$ is drawn from the $N(0, \hat{s}_i^2)$ distribution. This treatment implies that the errors are assumed to pertain to *changes* in the exogenous variables. The error v_{i1} is carried along from quarter 1 on, the error v_{i2} is carried along from quarter 2 on, and so forth. Given the way that many exogenous variables are forecast, by extrapolating past trends or taking variables to be unchanged from their last observed values, it may be that any error in forecasting the level of a variable in, say, the first period will persist throughout the forecast period. If this is true, the assumption that the errors pertain to the changes in the variables may be better than the assumption that they pertain to the levels. Given that the simulation period is 8 quarters in length and given that there are 91 exogenous variables, 728 exogenous variable errors are drawn for each repetition.

Turn next to the d row for the US model. This row required by far the most computational work. In the notation in Section 7.7, each value in a d row is the square root of $\hat{\sigma}^2_{itk}$. Put another way, the square of each d row value is equal to the square of the c row value plus $\bar{d}_{ik}$, where $\bar{d}_{ik}$ is the mean of the d_{isk} values discussed in Section 7.7. In computing the d_{isk} values, the model was estimated and stochastically simulated 68 times. All estimation periods began in 1954:1 (except for equation 15, where the beginning was 1956:1). The first estimation period ended in 1976:2, the second in 1976:3, and so on through 1993:1. The estimation technique was 2SLS. For each estimation period the covariance matrix of the structural error terms, Σ, and the covariance matrix of the coefficient estimates, V_2, were estimated along with the coefficients. For this work V_2 was taken to be block diagonal.

Dummy variables whose nonzero values begin after 1976:2 obviously cannot be included in the version of the model estimated only through 1976:2. Dummy variables were thus added when appropriate as the length of the estimation period increased. The variable $D794823 \cdot PCM1_{-1}$ in equation 30 was added for the first time for the estimation period ending in 1979:4. The variable $D811824$ in equation 21 was added for the first time for the period ending in 1981:1, and the variable $D831834$ in the same equation was added for the first time for the period ending in 1983:1. Finally, the variables involving $DD772$ in equations 13 and 14 were added for the first time for the period ending in 1983:1.

Given the 68 sets of estimates, 68 stochastic simulations were run. Each simulation period was of length 8 quarters subject to the restriction that the last quarter for predictions was 1993:2. All simulations were outside the estimation period. The first simulation period began in 1976:3, the second in 1976:4, and so on through 1993:2. Both structural error terms and coefficients were drawn for these simulations (using the appropriate estimates of Σ and V_2), and the number of repetitions per each of the 68 stochastic simulations was 250. For the one quarter ahead prediction ($k = 1$), these calculations allowed 68 values of d_{isk} to be computed for each endogenous variable i, from which the mean $\bar{d}_{ik}$ was computed. For the two quarter ahead prediction, there were 67 values of d_{isk} computed, and so on. Given these means and given the c row values in Table 8.5, the d row values could be computed.

The same procedure was followed for the other three models except that the other models have no exogenous variables and so no c row values are needed. For these models the number of repetitions per stochastic simulation was 1000 even for the 68 stochastic simulations involved in getting the d_{isk} values. The estimation technique was ordinary least squares. As was the case for the US model, the covariance matrices of the coefficient estimates were taken to be block diagonal.

Once these calculations have been done and the d row values computed, one can compare the models. As discussed in Section 7.7, each model is on an equal footing with respect to the d row values in the sense that the four main sources of uncertainty of a forecast have been accounted for. The d row values can thus be compared across models.

Turn now to the d row values in Table 8.5, and consider first the US model versus the two VAR models. For real GDP ($GDPR$) the US model is better than VAR5/2 for the first four quarters and slightly worse for the remaining four. The US model is better than VAR4 for all eight quarters. For the GDP deflator ($GDPD$) the US model is worse than VAR5/2 for the first five quarters and better for the remaining three. The US model is worse than VAR4 for the first three quarters, tied for quarter four, and better for the remaining four quarters. For the unemployment rate (UR) the US model is better than VAR5/2 for the first four quarters and essentially tied for the remaining four. The US model is better than VAR4 for all quarters. For the bill rate (RS) the US model is better than both VAR models for all quarters. Comparing VAR5/2 and VAR4, VAR5/2 is more accurate for all variables and all quarters except for the one quarter ahead prediction of the bill rate, where the two models essentially tie.

Using VAR5/2 as the better of the two VAR models, what conclusion can be drawn about the US model versus VAR5/2? For the first three variables the models are generally quite close, and one might call it a tie. For the fourth variable, the bill rate, the US model does substantially better. The US model may thus have a slight edge over VAR5/2, but only slight. Remember, however, that the present results are based on the use of the autoregressive equations for the 91 exogenous variables. As discussed earlier, these equations may exaggerate the uncertainty of the exogenous variables and thus bias the results against the US model.

Turning next to the AC model, it does very well in the $GDPR$ predictions. It has the smallest d row values in the table. There clearly seems to be predictive power in the lagged components of $GDPR$ that is not captured in the US and VAR models.

Comparing the a and b rows in Table 8.5 shows that coefficient uncertainty contributes much less to the variances than does the uncertainty from the structural error terms. In other words, the a row values are large relative to the difference between the b row and a row values. For the US model the differences between the c row values and the b row values are generally larger than the differences between the b row and a row values, which says that exogenous variable uncertainty (as estimated by the autoregressive equations) generally contributes more to the total variance than does coefficient uncertainty.

The differences between the d row and c row values are measures of the misspecification of the model not already captured in the c row values. On this score, the worst specifications for the models are for the bill rate and the best are for the unemployment rate. Again, the differences between the US model and VAR5/2 regarding misspecification are close except for the bill rate, where the US model is much better.

Outside Sample RMSEs

From the 68 stochastic simulations that are used for the d_{isk} calculations, one has for each endogenous variable i, 68 one quarter ahead outside sample error terms, 67 two quarter ahead outside sample error terms, and so on. (These errors are denoted $\hat{\epsilon}_{isk}$ in Section 7.7.) From these errors one can compute RMSEs, and the results of doing this for four variables are presented in Table 8.6. Remember, however, that comparing RMSEs across models has problems that do not exist when comparing the d row values in Table 8.5 across models. Exogenous variable uncertainty is not accounted for, which affects the comparisons between the US model and the others but not between the other models themselves. Also, the fact that forecast error variances change over time is not accounted for in the RMSE calculations. The RMSEs in Table 8.6 are, however, all outside sample, which is a least a crude way of accounting for misspecification effects.

For what they are worth, the results in Table 8.6 show that the US model is noticeable better than the VAR models for real GDP and the bill rate. The

Table 8.6
RMSEs of Outside Sample Forecasts for Four Models for 1976:3–1993:2

	Number of Quarters Ahead					
	1	2	3	4	6	8
			$GDPR$: Real GDP			
US	.79	1.39	1.95	2.33	2.64	2.74
VAR5/2	1.07	1.89	2.51	3.06	3.84	4.57
VAR4	1.15	2.05	2.61	3.22	4.33	5.15
AC	.79	1.23	1.64	1.95	2.48	2.99
			$GDPD$: GDP Deflator			
US	.34	.58	.82	1.23	2.32	3.21
VAR5/2	.31	.61	.87	1.18	1.98	2.89
VAR4	.33	.62	.88	1.18	2.01	3.02
			UR: Unemployment Rate			
US	.31	.61	.89	1.16	1.51	1.61
VAR5/2	.32	.65	.94	1.20	1.52	1.68
VAR4	.36	.74	1.04	1.29	1.68	1.84
			RS: Bill Rate			
US	.80	1.61	1.91	2.29	3.03	3.61
VAR5/2	1.18	2.08	2.52	3.07	4.13	5.10
VAR4	1.17	2.15	2.56	3.12	4.53	5.56

1. The results are based on 68 sets of coefficient estimates of each model.
2. Each prediction period began one quarter after the end of the estimation period.
3. For UR and RS the erors are in percentage points. For $GDPR$ and $GDPD$ the errors are expressed as a percent of the forecast mean (in percentage points).

results are fairly close for the GDP deflator and the unemployment rate. The AC model is about the same as the US model and noticeably better than the VAR models. Therefore, as expected, the US model does better relative to the other models when exogenous variable uncertainty is not taken into account.

This completes the comparison of the models using the d row values. The next two sections compare the models in two other ways, and the final section summarizes the overall comparison results.

8.7 Comparing Information in Forecasts[9]

Section 7.8 discussed a method for comparing the information in various forecasts, and this section uses this method to compare the forecasts from the US, US+, VAR5/2, VAR4, and AC models. The results of comparing the US and US+ models to the other three are presented in Table 8.7, and the results of comparing the AC model to the two VAR models are presented in Table 8.8. The rest of this section is a discussion of these two tables.

When using the method in Section 7.8, the forecasts should be based on information only up to the beginning of the forecast period. In other words, they should be "quasi ex ante" forecasts. The 68 sets of estimates that were used for the results in the previous section are used here to generate the forecasts. As was the case in the previous section, each forecast period begins one quarter after the end of the estimation period. There are 68 one quarter ahead forecasts, 67 two quarter ahead forecasts, and so on. All these forecasts are outside sample, and so they meet one of the requirements of a quasi ex ante forecast.

The other main requirement of a quasi ex ante forecast is that it not be based on exogenous variable values that are unknown at the time of the forecast. The VAR and AC forecasts meet this requirement because the models have no exogenous variables, but the forecasts from the US model do not. The 68 sets of forecasts that were computed for the US model are based on the actual values of the exogenous variables.[10] The US+ model, on the other hand, has no hard to forecast exogenous variables, and so it meets the exogenous variable requirement. Both the US and US+ models were used for the present results to see how sensitive the results for the US model are to the treatment of exogenous variables. For this work the US+ model was also estimated 68 times, including estimation of the 91 exogenous variable equations, and these

[9]The material in this section is an updated version of the material in Fair and Shiller (1990) (FS). In FS the US model was compared to six VAR models, eight AC models, and two autoregressive models, whereas for present purposes only two VAR and one AC model are used. In addition, the version of the US model that was used in FS was the version that existed in 1976, whereas the current version of the model is used here. Finally, only the results for real output were discussed in FS, whereas results for the GDP deflator, the unemployment rate, and the bill rate are also discussed here. The forecasts examined in this section are all *quasi* ex ante. The information content of *actual* ex ante forecasts for a number of models is examined in Fair and Shiller (1989) using the present method, but this material is not presented here.

[10]Remember that the actual values of the exogenous variables were used in computing the d_{isk} values in the previous section. Exogenous variable uncertainty was handled through the c row calculations.

Table 8.7
US Model Versus Three Others: Estimates of Equation 7.12

Other Model	One Quarter Ahead Forecast cnst	US β	Other γ	SE	Four Quarter Ahead Forecast cnst	US β	Other γ	SE
				$GDPR$: Real GDP				
				US Model				
VAR5/2	-.0008	.781	-.051	.00691	-.0025	.753	-.103	.01727
	(0.45)	(5.30)	(0.34)		(0.41)	(4.87)	(0.72)	
VAR4	-.0008	.756	-.003	.00692	-.0021	.767	-.112	.01722
	(0.50)	(5.35)	(0.03)		(0.36)	(4.86)	(0.84)	
AC	-.0020	.620	.324	.00681	-.0101	.505	.578	.01629
	(1.11)	(3.48)	(1.45)		(1.56)	(3.74)	(2.30)	
				US+ Model				
VAR5/2	-.0002	.678	.006	.00825	.0069	.381	.153	.02121
	(0.10)	(3.90)	(0.04)		(0.53)	(1.01)	(0.60)	
VAR4	-.0000	.613	.064	.00823	.0053	.417	.124	.02123
	(0.02)	(3.02)	(0.52)		(0.43)	(1.08)	(0.52)	
AC	-.0020	.289	.758	.00770	-.0116	.335	.911	.01866
	(0.90)	(1.51)	(4.14)		(1.45)	(2.13)	(3.37)	
				$GDPD$: GDP Deflator				
				US Model				
VAR5/2	.0023	.454	.416	.00260	.0079	.519	.341	.01000
	(3.22)	(3.49)	(2.95)		(1.36)	(2.54)	(1.59)	
VAR4	.0027	.461	.387	.00264	.0082	.489	.377	.00981
	(3.71)	(3.49)	(2.67)		(1.54)	(2.39)	(1.89)	
				US+ Model				
VAR5/2	.0024	.394	.454	.00284	.0073	.261	.582	.01050
	(3.08)	(2.26)	(2.41)		(1.10)	(1.02)	(2.41)	
VAR4	.0027	.407	.428	.00282	.0071	.307	.556	.01021
	(3.53)	(2.96)	(2.82)		(1.14)	(1.43)	(2.94)	
				UR: Unemployment Rate				
				US Model				
VAR5/2	.0018	.579	.398	.00278	.0385	.689	-.200	.00909
	(0.97)	(4.25)	(2.84)		(3.96)	(2.82)	(0.85)	
VAR4	.0030	.730	.230	.00288	.0409	.761	-.305	.00892
	(1.68)	(6.23)	(1.89)		(4.89)	(3.20)	(1.48)	
				US+ Model				
VAR5/2	.0011	.595	.392	.00279	.0373	.556	-.071	.00996
	(0.54)	(4.14)	(2.67)		(3.06)	(2.30)	(0.28)	
VAR4	.0021	.748	.225	.00288	.0399	.625	-.176	.00990
	(1.10)	(5.83)	(1.73)		(3.67)	(2.49)	(0.75)	
				RS: Bill Rate				
				US Model				
VAR5/2	-.31	1.069	-.027	.795	1.69	.588	.184	2.180
	(0.88)	(6.55)	(0.20)		(0.92)	(1.74)	(1.21)	
VAR4	-.32	1.097	-.054	.795	1.63	.662	.121	2.209
	(0.94)	(6.82)	(0.37)		(0.86)	(1.83)	(0.75)	
				US+ Model				
VAR5/2	-.35	1.073	-.027	.822	2.28	.501	.186	2.223
	(0.92)	(6.20)	(0.19)		(1.34)	(1.55)	(0.95)	
VAR4	-.36	1.093	-.047	.821	2.22	.575	.123	2.247
	(0.99)	(6.54)	(0.31)		(1.27)	(1.64)	(0.60)	

Table 8.8
AC Versus VAR5/2 and VAR4

	One Quarter Ahead Forecast				Four Quarter Ahead Forecast			
Other Model	cnst	AC β	Other γ	SE	cnst	AC β	Other γ	SE
				$GDPR$: Real GDP				
VAR5/2	-.0010	.916	.106	.00778	-.0038	.938	.204	.01863
	(0.54)	(5.28)	(0.81)		(0.46)	(3.78)	(2.28)	
VAR4	-.0010	.881	.120	.00774	-.0048	.954	.181	.01873
	(0.54)	(4.81)	(1.18)		(0.60)	(3.79)	(2.22)	

68 sets of estimates were used. All the forecasts for the US+ model were also outside sample. Again, remember from the discussion in Section 7.8 that the treatment of the exogenous variables as in US+ may bias the results against the model. Many of the exogenous variables may not be as uncertain as the autoregressive equations imply.

Both one quarter ahead and four quarter ahead forecasts are examined in Table 8.7. In the estimation of the equations, the standard errors of the coefficient estimates were adjusted in the manner discussed in Section 7.8 to account for heteroskedasticity and (for the four quarter ahead results) a third order moving average process for the error term. Equation 7.12 was used for real GDP and the GDP deflator, where both variables are in logs, and the level version of equation 7.12 was used for the unemployment rate and the bill rate.

Turn now to the results in Table 8.7, and consider the forecasts of real GDP first. Also, ignore for now the results for the AC model. The results show that both US and US+ dominate the VAR models for real GDP. The estimates of the coefficients of the VAR forecasts are never significant, and the estimates of the coefficients of the US and US+ forecasts are significant except for the four quarter ahead forecasts for US+, where the t-statistics are about one. It is thus interesting to note that even though the standard errors of the forecasts in Table 8.5 (the d row values) are fairly close for real GDP for the US and VAR models, the results in Table 8.7 suggest that the VAR forecasts contain no information not already in the US forecasts. In this sense the method used in this section seems better able to discriminate among models.

The results for the GDP deflator show that both the US (and US+) forecasts and the VAR forecasts contain independent information. In most cases both coefficients are significant, the exceptions being US versus the VAR models for the four quarter ahead forecasts, where the VAR forecasts are not quite significant, and US+ versus the VAR models for the four quarter ahead forecasts, where the US+ coefficients are not quite significant.

For the unemployment rate US and US+ dominate the VAR models with the exception of the one quarter ahead forecasts from VAR5/2, which are significant in the US and US+ comparisons, although with t-values smaller than those for the US and US+ forecasts.

The results for the bill rate show that US and US+ dominate the VAR models for the one quarter ahead forecasts. For the four quarter ahead forecasts the US and US+ forecasts have larger coefficient estimates and larger t-values than do the VAR forecasts, although collinearity is such that none of the t-values are greater than two.

The results of these comparisons are thus encouraging for the US model. Only for the GDP deflator is there much evidence that even the US+ forecasts lack information that is contained in the VAR forecasts.

Consider now the AC model, where there are only results for real GDP. The US and US+ comparisons in Table 8.7 suggest that both the US or US+ forecasts and the AC forecasts contain independent information. There clearly seems to be forecasting information in the lagged components of GDP that is not captured in the US model, and this is an interesting area for future research.

The VAR versus AC comparisons in Table 8.8 show that the VAR forecasts appear to contain no independent information for the one quarter ahead forecasts, but at least some slight independent information for the four quarter ahead forecasts. As did the results in the previous section, these results suggest that the AC model may be a better alternative than VAR models for many purposes.[11]

---

[11]With a few exceptions, the results for real GDP here are similar to those in Fair and Shiller (1990) (FS). The US+ version is closest to the version used in FS, and so the following discussion focuses on the US+ results. The one quarter ahead results for US+ in Table 8.7 have the US model dominating the VAR models, which is also true in Table 2 in FS. For the four quarter ahead results neither the US+ nor the VAR forecasts are significant in Table 8.7 and both are significant in Table 2 in FS. However, in both tables the US forecasts have larger coefficient estimates and larger t-values than do the VAR forecasts. Regarding US+ versus AC, the results in Table 8.7 are more favorable for AC than they are in Table 2 in FS. In Table 2 in FS the US model dominates the AC models, whereas in Table 8.7 the AC model has a large and significant coefficient estimate for both the one quarter ahead and four quarter ahead forecasts for US+ versus AC. Finally, the VAR versus AC comparisons in Table 8.8 are similar to those in Table 3 in FS. In both tables the AC forecasts dominate the VAR forecasts for the one quarter ahead results and both forecasts are significant for the four quarter ahead results.

8.8 Estimating Event Probabilities[12]

The use of event probability estimates to compare models was discussed in Section 7.9. This comparison is made in this section for two events and five models. The five models are the US, US+, VAR5/2, VAR4, and AC models. The two events, labelled A and B are:

A = At least two consecutive quarters out of five of negative real GDP growth.

B = At least two quarters out of five of negative real GDP growth.

Event A is a recession as generally defined. Event B allows the two or more quarters of negative growth not to be consecutive.

The first 64 sets of estimates of each model that were used for the results in the previous section were used here. (Only 64 rather than 68 sets of estimates could be used because each forecast here has to be five quarters ahead.) There were 64 five quarter ahead outside sample stochastic simulations performed. The number of repetitions per five quarter forecast was 250 for US and US+ and 1000 for the VAR5/2, VAR4, and AC.

Regarding the US+ model, this is the first time that stochastic simulation of the model is needed. For the results in the previous section only deterministic outside sample forecasts were used. As discussed in Section 8.2, when stochastic simulation was performed using US+, the covariance matrix of all the error terms, which is 121×121, was taken to be block diagonal. For the results in this section this matrix was estimated 64 times, each estimate being used for each of the 64 stochastic simulations. The covariance matrices of the coefficient estimates are not needed for the work in this section because coefficients are not drawn.

From the stochastic simulation work one has five sets of values of P_t $(t = 1, \cdots, 64)$ for each of the two events, one for each model, where P_t is the model's estimate of the probability of the event for the period beginning in quarter t. One also has values of R_t for each event, where R_t is the actual outcome—one if the event occurred and zero otherwise. Given the values

[12] The material in this section is an updated and expanded version of the material in Section 3.3 in Fair (1993c). In Fair (1993c) only within sample forecasts were used and the only comparisons were to the constant model and a fourth order autoregressive model. In this section all the forecasts are outside sample and comparisons are made to two VAR models and an AC model in addition to the constant model. Also, no coefficients are drawn for the present results, whereas they were drawn in the earlier work. (See the discussion in Section 7.9 as to why coefficients were not drawn here.)

Table 8.9
Estimates of Probability Accuracy

Model	$\bar{p}$	QPS	LPS
Event A (Actual $\bar{p}$ = .188)			
Constant	.188	.305	.483
US	.175	.310	.477
US+	.173	.310	.472
VAR5/2	.310	.496	.844
VAR4	.264	.518	.972
AC	.154	.324	.510
Event B (Actual $\bar{p}$ = .234)			
Constant	.234	.359	.545
US	.211	.290	.438
US+	.238	.306	.465
VAR5/2	.416	.514	*
VAR4	.358	.521	*
AC	.237	.363	.537

*LPS not computable.

of R_t, another model can be considered, which is the model in which P_t is taken to be equal to $\bar{R}$ for each t, where $\bar{R}$ is the mean of R_t over the 64 observations. This is simply a model in which the estimated probability of the event is constant and equal to the frequency that the event happened historically. This model will be called "Constant." The results for this model are not outside sample because the mean that is used is the mean over the whole sample period.

The summary statistics are presented in Table 8.9. In two cases (both for the VAR models) the LPS measure could not be computed because either P_t was 1 and R_t was 0 or vice versa. This is a limitation of the LPS measure in that in cannot handle extreme errors of this type. It, in effect, gives an infinite loss to this type of error.

The results in Table 8.9 are easy to summarize. Either US or US+ is best for both events for both error measures except the case of the constant model and event A, where the QPS for the constant model is slightly smaller. This is thus strong support for the US model.

The results in Table 8.9 also show that the AC model completely dominates the VAR models. This is in keeping with the results in the previous two sections, which generally show the AC model out performing the VAR models.

Figures 8.1 and 8.2 plot the values of P_t and R_t for the US+ and VAR5/2 models for event A for the 64 observations. It is clear from the plots why US+ has better QPS and LPS values in Table 8.9. VAR5/2 has high probabilities too early in the late 1970s and comes down too fast after the recession started compared to US+. Note that both models do not do well predicting the 1990–1991 recession. No model seems to do well predicting this recession.

8.9 Summary of the Test Results

Overall, the results in Tables 8.5, 8.6, 8.7, and 8.9 are favorable for the US model. Even after correcting for exogenous variable uncertainty that may be biased against the model, the model does well in the tests relative to the VAR and AC models. The GDP deflator results are the weakest for the US model, and this is an area for future work. Also, the results in Table 8.7 show that there is information in the AC forecasts of real GDP not in the US forecasts, which suggests that the US model is not using all the information in the lagged components of GDP. Aside from the GDP deflator forecasts, there does not appear to be much information in the VAR forecasts not in the US forecasts.

The AC model generally does as well as or better than the VAR models. This suggests that there is useful information in the lagged components of GDP that the VAR models are not using. From another perspective, if one wants a simple, non structural model to use for forecasting GDP, an AC model would seem to be a better choice than a VAR model.

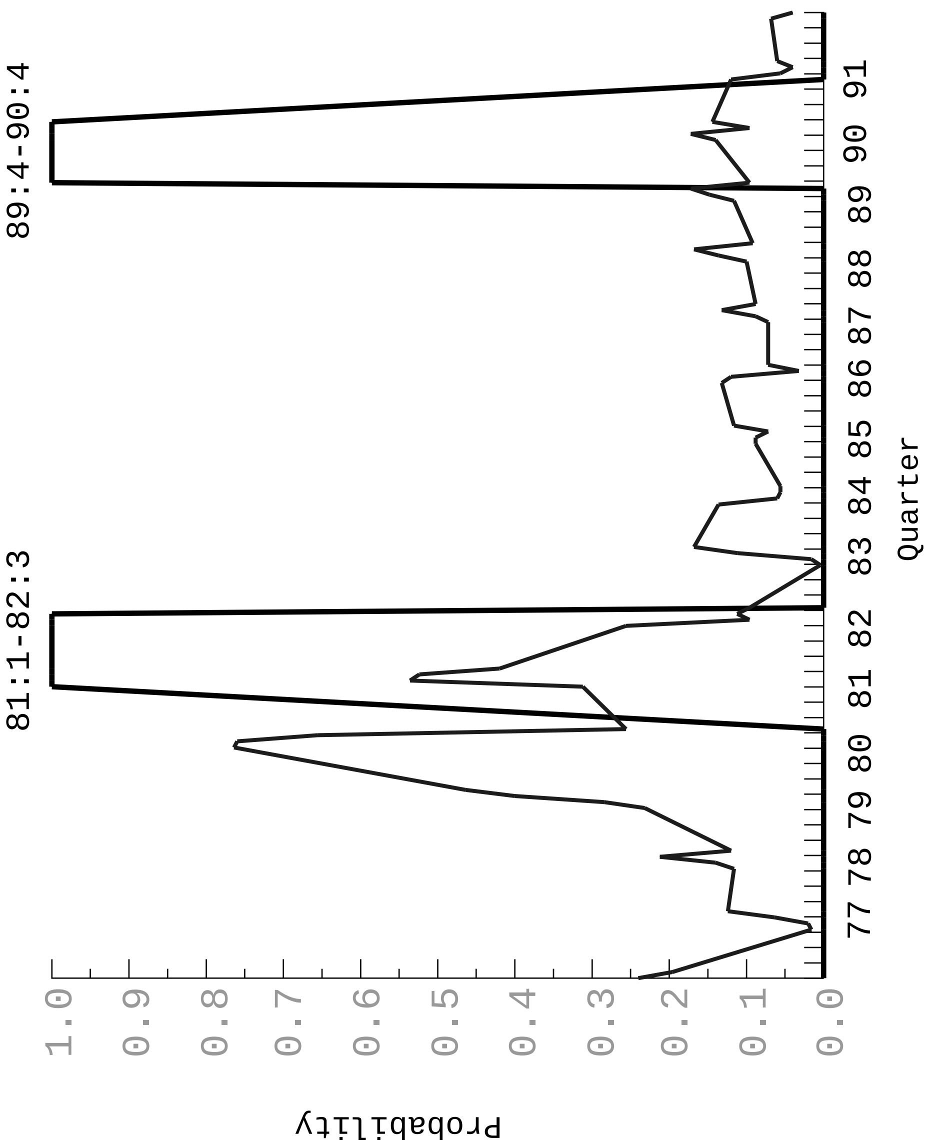

Figure 8.1 Estimated Probabilities for Event A for

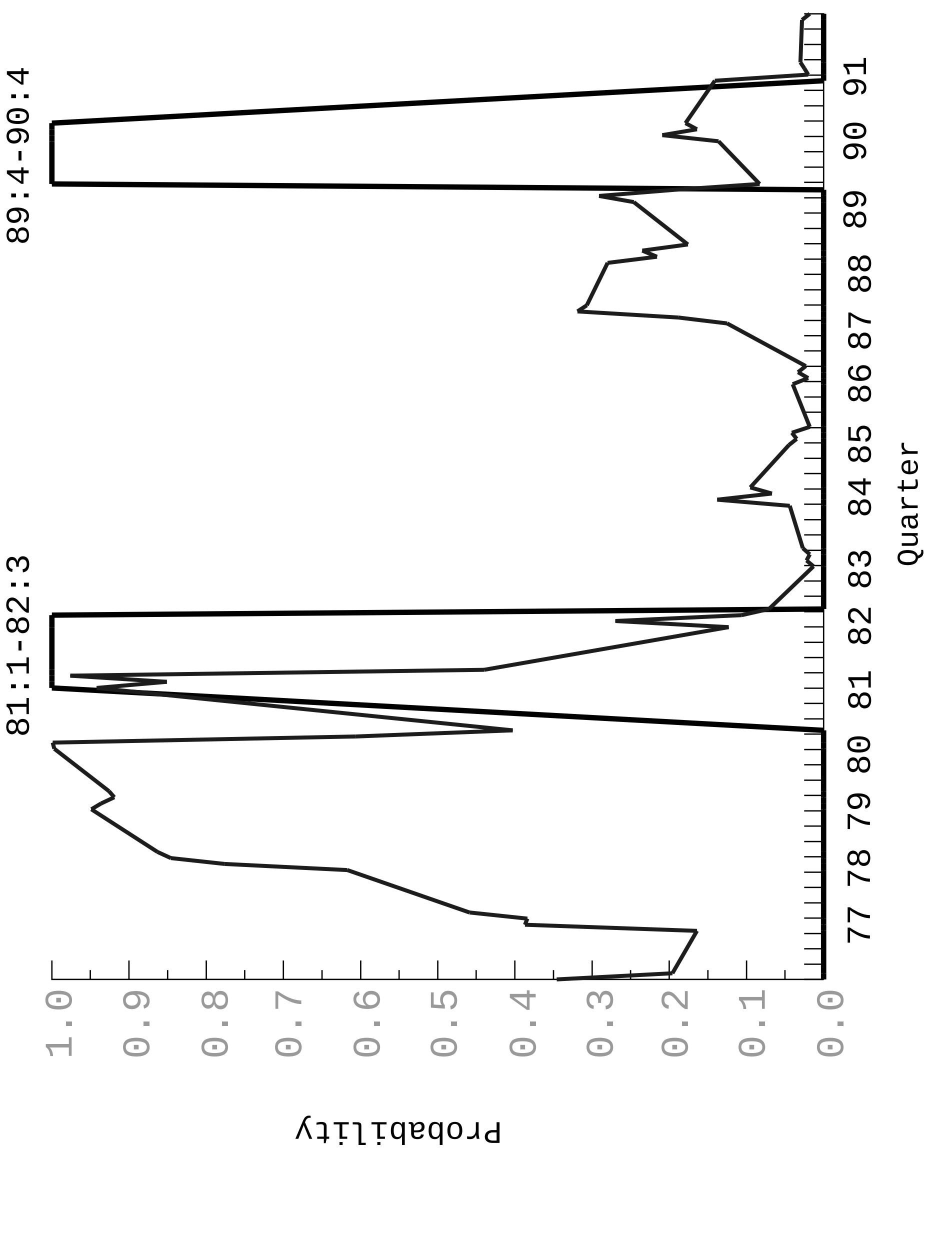

Figure 8.2 Estimated Probabilities for Event A fo