Volume Title: Topics in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90298-6

Volume URL: http://www.nber.org/books/wise92-1

Conference Date: April 5-7, 1990

Publication Date: January 1992

Chapter Title: Incentive Regulation of Nursing Homes: Specification Tests of the Markov Model

Chapter Author: Edward C. Norton

Chapter URL: http://www.nber.org/chapters/c7105

Chapter pages in book: (p. 275 - 304)

9 Incentive Regulation of Nursing Homes

Specification Tests of the Markov Model

Edward C. Norton

In Norton (1990), I presented evidence that changing the reimbursement system to nursing homes to account for performance had a positive effect both on quality and on controlling costs. The analysis used a simple Markov model to estimate transition probabilities between states of health in the nursing home. A comparison of the probabilities for the control group (no incentives) and the experimental group (positive incentives) found them to be different. People in the experimental group stayed for a shorter time and had better outcomes.

The simple Markov model maintains several strong assumptions. For example, it assumes that the transition probabilities are constant over time, independent of past states, and the same for all people. If any of these assumptions are false, the conclusions of the previous paper may be ill founded. This paper extends the analysis to more general models and in doing so subjects the simple Markov model to a series of rigorous specification tests. Most of the specification tests are done on data from the control group nursing homes only so as not to mix effects of the experiment with those of the assumptions. Each section of the paper tests one of the assumptions listed below:

1. *First order.* The probability of being in state j next period depends only on the current state, not on past states.
2. *Homogeneity.* The probabilities are independent of personal characteristics, such as age, sex, race, and marital status.
3. *Stationarity.* The probabilities are constant over time.

Edward C. Norton is a postdoctoral fellow in the Project on the Economics of Aging at the National Bureau of Economic Research and the Bradley Fellow at Harvard University Medical School, Department of Health Care Policy.

The author wishes to thank Rebecca Demsetz, Jerry Hausman, Paul Joskow, Daniel McFadden, Carolyn H. Norton, Mitchell Petersen, James Poterba, Sherwin Rosen, Douglas Staiger, members of the Public Finance lunch, and the NBER Conference on the Economics of Aging for helpful comments. Grant support from the National Science Foundation and the Bradley Foundation is gratefully acknowledged.

4. *Duration dependence.* The probabilities are independent of how long a person has been in the nursing home.
5. *Learning effect.* Nursing homes in the experimental group instantly switched to optimize under the new reimbursement system with no learning period.
6. *Markov assumption.* $P(T) = P(1)^T$.
7. *Measurement error.* Reporting errors by nurses have no effect on the estimated transition probabilities.

This paper contains a summary of the experiment done by the National Center for Health Services Research (NCHSR) and the data used in the analysis. Then there is a brief review of the results in Norton (1990). The remainder of the paper is a rigorous extension of the previous analysis.

9.1 Study and Data

This section contains a brief summary of the methodology of the NCHSR experiment and a description of the data collected. For a more complete discussion, see Weissert et al. (1983).

Thirty-six proprietary, Medi-Cal certified, skilled nursing facilities in San Diego participated in the study. All these nursing homes had at least thirty beds. Only four eligible nursing homes in San Diego declined to participate. NCHSR hired Applied Management Sciences Inc. of Silver Spring, Maryland, to collect the data and to supervise the team of registered nurses that did the fieldwork. After a six-month baseline data-gathering period, the nursing homes were split from a matched sample into two groups of eighteen. The experimental group received *all three* incentive payments, while the control group was paid only a nominal amount to cover the additional cost of bookkeeping. A total of 11,389 residents were tracked during the two-and-a-half-year study. Out of these residents, 58 percent were covered by Medicaid, and the incentives applied only to these people. Table 9.1 shows a time line of how the study was conducted, and table 9.2 gives summary statistics about the data.

9.1.1 Resident Classification System

The hired registered nurses visited each resident periodically to assess their health and to determine whether they achieved certain goals. New residents were assessed within two weeks of admission, and most reassessments were made at three-month intervals. When a person left a nursing home, the date and reason were recorded. Nurses classified residents as being in one of five states of health. Classification depended primarily on how much help was needed in activities of daily living (ADL). These objective measures have been used widely as the best measure of health status of the elderly (see Katz and Akpom 1976; and Börsch-Supan, Kotlikoff, and Morris 1988). There are six ADLs: bathing, dressing, eating, using the toilet, transferring, and walk-

Table 9.1 **Time Line**

November 1980 to April 1981	Collected baseline data, no incentives
30 April 1981	Homes randomly assigned to experimental or control groups
May 1981 to April 1982	New admissions eligible for admission incentives for one year
30 April 1982	End of study if admitted prior to May 1981
May 1982 to April 1983	Reassessed if admitted after 1 May 1981

Table 9.2 **NCHSR Nursing Home Data**

Subgroup	Average Age	% Women	% White	% Married
Admitted before study:				
Control ($N = 718$)	80	70	90	15
Experiment ($N = 637$)	79	73	91	16
Admitted during study:				
Control ($N = 1,417$)	80	72	91	16
Experiment ($N = 1,080$)	80	74	89	17

Table 9.3 **Classification of the Five States of Health**

Type	Dischargeable within 90 days?	ADL Index	% of Sample at Admission
A	Yes	Usually ADL ≤ 2	23
B	No	$1 \leq$ ADL ≤ 4	22
C	No	ADL $= 5$	31
D	No	ADL $= 6$	18
E	No	ADL ≥ 4 and required special nursing services	5

ing. A person who needs assistance in four of these categories has an ADL index of four. Classification also depended on how soon someone was likely to be discharged and whether special nursing care was needed. Type A people are the "healthiest" and type E the "sickest." Table 9.3 summarizes the classification scheme.

9.1.2 Incentive Payments

Admission

Medicaid reimbursement in California was a flat prospective rate ($36 in 1981), with the result that nursing homes were reluctant to admit people who required more than average care. Nursing homes in the experimental group received a per diem bonus when they admitted type D and E residents. The size of the bonuses reflected wages needed to pay for increased nursing care coverage. The per diem bonuses for types A, B, C, and D were 0, −2.5, 0, and 5, respectively. The rate for E ranged from 3 to 28, depending on the amount of special care needed.

Outcome

Some residents needed special rehabilitation to improve their functional or health status. This requires a large fixed cost to the nursing home that is not reimbursed under flat rate reimbursement and is therefore discouraged. Experimental group nursing homes received a lump sum bonus if selected residents improved their health status (corresponding to E → D, C → B or D → B, and B → A). These goals were designed to reduce ADL dependence and eliminate the need for special nursing services. Nursing homes were paid only if the goal was met within ninety days. It is important to remember that, in order to be eligible for a goal, a resident first had to be nominated by his or her nursing home. Nursing homes nominated residents whom they felt would benefit from costly rehabilitation services, and the hired NCHSR nurses had to approve each nomination. In the experimental group nursing homes, 150 nominations were approved. The hired nurses nominated residents in the control group nursing homes. Nursing homes received an amount equal to the estimated wages needed to pay for extra nursing help, ranging from $126 to $370.

Discharge

People well enough to be discharged are also the least expensive to care for. Nursing homes prefer to keep these people, although they cannot legally prevent anyone from leaving. To encourage appropriate discharges, nursing homes received a lump sum bonus if certain residents were discharged from the nursing home promptly *and* the resident stayed out of the nursing home for ninety days. Payment was designed to offset the cost of a vacant bed and the administrative costs of discharge. Type A residents were not eligible for discharge bonuses since they were expected to be discharged soon anyway. Payment ranged from $230 down to $60, with more paid for a timelier discharge. Like outcome incentives, experimental group nursing homes had to nominate residents and have their choices approved (113 were approved).

9.2 Markov Model

An implicit assumption in the NCHSR experiment is that the way nursing homes are paid affects their effort, which in turn affects residents' health. This section briefly describes a model of how reimbursement affects health. A first-order Markov model is described completely by a set of probabilities (see Amemiya 1985, chap. 11). Let P_{ij} be the probability that a person goes from state i to state j in one period. There are nine states in the model, five states of health within in the nursing home and four states to go to outside the nursing home. State of health is based on an objective index of a person's ability to perform basic tasks of living, known as the Activities of Daily Living Index.[1] When residents leave a nursing home, their state of health determines where they go next: home, intermediate care facility, hospital, or death. Although health is a random variable, it does not depend solely on biology but depends also on the nursing home's effort, e, which in turn is a function of the reimbursement system, I. Let $P_{ij} = P[e(I)]_{ij}$. Presumably, larger incentives trigger greater effort and better health.

A simple test of the plan's effectiveness is to estimate P for both the control and the experimental groups and to compare the transition matrices. One would expect that, if the program is effective, the probability of improving one's health will be greater in the experimental group nursing homes. Also, using the Markov model allows one to compare estimates of the length and cost per spell in the nursing home. If the results show no differences between the groups, it could be because the incentives were too small to induce much more effort. It could also be that increased effort from the current level would not change a resident's fortunes, and it would therefore not be worthwhile to increase effort.

The P matrices were estimated by maximum likelihood. Estimation controlled for censored observations and the fact that the time between observations was not constant. The transition probabilities were estimated for a two-week time period.

The results can be summarized as follows. Incentive regulation of nursing homes had beneficial effects on both quality and cost of care. People in experimental group nursing homes were more likely to go home or to a lower-level nursing home and less likely to be hospitalized or to die than people in control group nursing homes. The admission incentives induced nursing homes to admit more people with severe disabilities. The most striking difference between the experimental and the control groups is that both the mean and the median length of stay are much shorter in the experimental group. The incentives do seem to cause the nursing homes to discharge residents more quickly. Were this program implemented, the cost savings would not come directly from shorter stays since high occupancy rates mean a nearly constant Medicaid population in nursing homes. Instead, the more rapid turnover rate would transfer patients out of hospitals and save hospital costs. The administrative and incentive costs of the NCHSR program are negligible compared to the potential savings.

[1] There are five categories in this study based on the six ADLs (eating, bathing, transferring into and out of bed, using the toilet, walking, and dressing). ADLs are good determinants of ability to function alone (see Katz and Akpom 1976). The five categories are described in section 9.3.

9.3 First Order versus Second Order

The assumption that the Markov model is first order means that the transition probabilities depend only on current health. Any information from the past should not help predict the future. A more general model would incorporate information about whether a person has been improving, has been getting worse, or has remained stable. The first-order assumption can be tested against the alternative that the model is *second order.* If the model is second order, then the probabilities depend on both health now and health last period.

A second-order model can be written as a first-order model with many more initial states. Specifically for the nursing home data, the first-order model has five possible states of health and the second-order model thirty. More precisely, let P_{ijk} be the probability that a person will be type k next period, given that he or she is type j now and was type i last period. The index j ranges over the five categories A–E. The indices i and k range over those five and also the state of being out of the nursing home. If the model is first order, then the following probabilities in the second-order model should be the same:

$$P_{Ajk} = P_{Bjk} = P_{Cjk} = P_{Djk} = P_{Ejk} = P_{Ojk} \forall j, k.$$

Anderson and Goodman (1957) give a test for the null hypothesis that a model is first order against the alternative that it is second order. The test uses probability estimates from the expanded second-order matrix. Stationarity and homogeneity are maintained hypotheses. Anderson and Goodman show that the likelihood ratio criterion for testing the null hypothesis for a current state j is

$$\chi_j^2 = \sum_{i,k} n_{ij}\left(\hat{P}_{ijk} - \hat{P}_{jk}\right)^2/\hat{P}_{jk},$$

where $n_{ij} = \sum_k n_{ijk}$, n_{ijk} = the number of observed transitions $i \rightarrow j \rightarrow k$, $\hat{P}_{ijk} = n_{ijk}/n_{ij}$, and $\hat{P}_{jk} = \sum_i n_{ijk}/\sum_{i,k} n_{ijk}$.

The tests use data from the control group nursing homes for all people admitted during either the study or the baseline periods. The four absorbing states outside the nursing home were combined into a single state, GONE. This is different than having not yet entered the nursing home. The rows marked "O" in table 9.4 indicate out of nursing home prior to admission. These are the people who are currently in the nursing home but who had not yet been admitted last period. Only people who were observed at least twice were included.[2]

For this test, the model assumes that the time between observations is always three months. In fact, 87 percent of all observations occurred three

[2] Leaving the nursing home (GONE) counted as being observed, but being right censored did not.

Table 9.4 **Test of First Order versus Second Order**

		Future State						
Previous	Current	A	B	C	D	E	GONE	*N*
A	A	*38*	28	2	0	0	32	133
B	A	17	*45*	1	0	0	36	69
C	A	11	28	0	0	0	61	18
D	A	67	0	0	0	0	33	3
E	A	0	50	0	0	0	50	4
O	A	22	15	7	1	1	*53*	374
Average		25	22	5	1	1	47	
A	B	*16*	50	9	1	0	25	122
B	B	8	*59*	8	1	0	24	567
C	B	5	42	*37*	1	0	14	201
D	B	11	47	11	0	0	32	19
E	B	0	57	0	0	0	43	7
O	B	7	44	18	3	1	*27*	383
Average		8	51	16	1	0	24	
A	C	*4*	26	56	4	4	7	27
B	C	3	*26*	48	4	1	18	172
C	C	0	11	*62*	7	0	19	627
D	C	0	7	53	*17*	1	22	101
E	C	0	19	31	0	6	44	16
O	C	2	18	46	5	1	*27*	546
Average		1	15	54	7	1	22	
A	D	0	0	20	40	40	0	5
B	D	0	0	24	59	0	18	17
C	D	0	0	*28*	33	0	40	83
D	D	0	0	13	*54*	4	28	230
E	D	0	0	20	40	0	40	5
O	D	1	6	19	34	3	37	314
Average		0	3	18	42	3	34	
A	E	0	0	75	0	25	0	4
B	E	100	0	0	0	0	0	1
C	E	0	0	23	8	8	62	13
D	E	0	0	6	*18*	24	53	17
E	E	0	5	7	0	*44*	44	41
O	E	3	6	8	2	24	*57*	90
Average		2	4	10	4	28	52	

Note: Numbers are probabilities. The elements on the main diagonal that are the largest in their column (if the number of observations in the row ≥ 20) are set in italic.

months after the previous observations. Furthermore, 89 percent of all first assessments were taken within three months of admission. Finally, most people who left a nursing home did so within three months of their last assessment. The few observations that did not fit were left in the analysis. I decided that it was better to include observations with timing problems than to use only a partial history for some people. The results are robust against leaving these observations out of the sample.

Surprisingly, going from a first- to a second-order model simplifies the issue of timing. A large fraction of observations are over a short time interval because they are of transitions between admission and first assessment. The second-order model conditions first on the current state, then on the past state. The exact time between admission and the first assessment is not important, as long as it is less than three months. In contrast, the first-order model conditions only on the current state. Parameter estimates would be biased if either observations from admission to a first assessment were left out or if it were assumed that the time interval was exactly three months.

The results of the chi-squared tests are shown in table 9.5A both by group and overall. The null hypothesis is soundly rejected, not only overall, but also group by group. Therefore, a person's recent history affects the transition probabilities. It is possible that the model is of a higher order than second, but there are not enough data to test this.

We can learn more by looking at the estimated second-order transition matrix in table 9.4 An interesting pattern can be seen in the blocks of $\hat{P}_{ijk}$. The elements on each block's main diagonal are particularly large relative to the other numbers in their column (when $n_{ij} \geq 20$, with one exception). If the first-order assumption were true, then the probabilities in each column (for a given block) would all be about the same. In other words, a person who is

Table 9.5 **Test of First Order versus Second Order**

A. Test of Null Hypothesis That Model Is First Order	
A	χ^2 = 66.08 rejects null at .000014 level
B	χ^2 = 137.40 rejects null at .000000 level
C	χ^2 = 109.10 rejects null at .000000 level
D	χ^2 = 68.81 rejects null at .000006 level
E	χ^2 = 77.64 rejects null at .000000 level
Total	χ^2 = 459.03 rejects null at .000000 level
B. Renormalized Test of Null Hypothesis That Model Is First Order	
A	χ^2 = 29.59 rejects null at .24 level
B	χ^2 = 39.78 rejects null at .031 level
C	χ^2 = 48.92 rejects null at .0029 level
D	χ^2 = 37.21 rejects null at .055 level
E	χ^2 = 60.96 rejects null at .000077 level
Total	χ^2 = 216.46 rejects null at .000000 level

now type j would be type k next period with a probability that is independent of last period's state. However, the pattern in this table shows that, if a person is type j now, the probability that she will become type k next period increases if we also know that she was type k last period. Of course, people are still most likely to stay in the state they are in, and if they change states, they are most likely to go the adjacent state. This pattern can also be expressed as

$$\Pr(k \text{ future} \mid j \text{ now}) < \Pr(k \text{ future} \mid j \text{ now}, k \text{ before}).$$

The chi-squared test can be taken one step further to see whether it still rejects when controlling for the effect outlined above. To do this, renormalize the numbers in table 9.4 so that the $k \rightarrow j \rightarrow k$ pattern is eliminated but all other features of the data are preserved. If the test on renormalized data does not reject, then this is the only interesting pattern to be found. The matrix in table 9.6 is a renormalized version of table 9.4, and each block has the following properties:

1. The weighted average of any column equals the element on the main diagonal.
2. Each row sum is one.
3. By construction, the main diagonal terms contribute nothing to the likelihood ratio test.

Nonetheless, the results of the renormalized chi-squared test indicate that more than half the variation in the original test is due to the fact that people tend to return to their previous state (see table 9.5B). Although three of the group tests still reject, it is clear that the $k \rightarrow j \rightarrow k$ pattern is the primary reason that the original first-order test rejected.

The test of first order against the alternative of second order was thoroughly rejected, which implies that past information is important for predicting the future health of nursing home residents. In addition, investigating the numbers highlighted the surprising fact that people who get worse in one period do not continue to decline but instead tend to rebound to their former state. A second-order model has the advantage of being more general but the disadvantage of being large to the point of being unwieldy. Furthermore, the rebounding effect has been noticed in other studies of longitudinal data and found to be an artifact of measurement error (see Poterba and Summers 1986). This possibility is explored below.

9.4 Homogeneity

Although nursing homes admit people from widely varying backgrounds, the simple Markov model does not control for heterogeneity. As a first cut, it is far simpler to assume that transition probabilities are constant. However, preliminary work showed that Markov matrices for subgroups chosen by age, sex, and marital status differ significantly. This section tests for heterogeneity

Table 9.6 Test of First Order versus Second Order

		Future State						
Previous	Current	A	B	C	D	E	GONE	*N*
A	A	25	34	2	0	0	39	133
B	A	24	22	1	0	0	52	69
C	A	10	26	6	0	0	57	18
D	A	67	0	0	1	0	33	3
E	A	0	50	0	0	1	50	4
O	A	26	18	8	1	1	45	374
Average		25	22	6	1	1	45	
A	B	8	54	10	1	0	27	122
B	B	10	51	10	1	0	29	567
C	B	7	59	13	1	0	20	201
D	B	11	46	11	2	0	31	19
E	B	0	57	0	0	0	43	7
O	B	7	44	18	3	1	27	383
Average		8	51	13	2	0	27	
A	C	1	27	57	4	4	7	27
B	C	4	13	56	5	1	21	172
C	C	0	13	56	8	0	22	627
D	C	0	8	59	7	1	25	101
E	C	0	20	33	0	0	47	16
O	C	2	19	50	5	1	22	546
Average		1	13	56	7	0	22	
A	D	0	0	20	40	40	0	5
B	D	0	3	23	57	0	17	17
C	D	0	0	19	37	0	45	83
D	D	0	0	19	35	6	40	230
E	D	0	0	19	38	4	38	5
O	D	1	6	18	33	3	39	314
Average		0	3	19	35	4	39	
A	E	2	0	73	0	24	0	4
B	E	95	5	0	0	0	0	1
C	E	0	0	10	9	9	72	13
D	E	0	0	7	2	28	63	17
E	E	0	7	10	0	23	61	41
O	E	3	6	8	2	24	58	90
Average		2	5	10	2	23	58	

Note: Numbers are probabilities.

in two different ways, both of which control for the three characteristics outlined above and also for race. The first method adds terms to each cell of the Markov matrix, and the second is an ordered logit model.

The obvious correction for heterogeneity is to add parameters to each cell in the transition matrix. The transition probabilities then depend on a constant and on personal characteristics (X includes age, sex, race, marital status):

$$\begin{aligned}\hat{P}_{ij} &= \exp(\beta_{ij} + X\gamma)\\ &= \exp(\beta_{ij} + \gamma_1\text{AGE} + \gamma_2\text{SEX} + \gamma_3\text{RACE} + \gamma_4\text{MARRIED})\\ &= \exp(\beta_{ij})\exp(\gamma_1\text{AGE})\exp(\gamma_2\text{SEX})\exp(\gamma_3\text{RACE})\exp(\gamma_4\text{MARRIED}).\end{aligned}$$

Notice that the effects are assumed to be multiplicative, not additive. If a characteristic has no effect, then its corresponding γ_i should be zero; thus, $\exp(\gamma_i X) = 1$. AGE is defined to be true age minus eighty, divided by ten.[3] The other characteristics are dummy variables: SEX equals one if male, RACE equals one if nonwhite, and MARRIED (marital status) equals one if currently married.

Unfortunately, there are too few observations in most cells to be able to parameterize fully. Instead, the matrix was reduced to five rows by five columns by combining states. Also, some parameters were constrained to be equal across cells. Table 9.7 depicts how this was done. A Greek letter (except β) denotes a four-element vector of parameters. There were only four parameters per characteristic, far fewer than one per cell. Note that the main diagonal equals one minus the row sum, and two cells have only constants.

The results are shown in tables 9.8 and 9.9. Only AGE and SEX have significant coefficients, so it appears that heterogeneity is quite weak. Looking at probability matrices for different types of people, we see a few interesting patterns. Older people are less likely to go home, as are women. Nonwhites are the least likely to die or go to a hospital, while married people are the most likely.

Another way to check for heterogeneity is to run an ordered logit model.[4] This has the advantage of controlling for many individual characteristics, but it does not have the special timing structure of the Markov model. One ordered logit was run for each of the five states of health in the nursing home. The dependent variable was the set of possible outcomes (collapsed from nine states into only five; see table 9.10). The outcomes were ranked; worse outcomes, like death, had a higher number.

The α's reported in table 9.11 are the cutoff values for the different categories. They are strictly monotonic, with higher thresholds for worse states of health. Thus, people who are type A have the highest α's and are the least likely to go to a bad state. This is consistent with the results of the Markov model.

[3] Eighty was chosen because it is the average age of a nursing home resident.

[4] For an explanation of ordered logit models, see Maddala (1983).

Table 9.7 **Parameterization for Test for Homogeneity**

	Home & ICF	Hosp. & Death	A	B, C	D, E
Home & ICF	1	0	0	0	0
Hosp. & Death	0	1	0	0	0
A	$\exp(\beta_{11}+X\mu)$	$\exp(\beta_{12}+X\gamma)$	(1 − row sum)	$\exp(\beta_{13}+X\theta)$	$\exp(\beta_{14})$
B, C	$\exp(\beta_{21}+X\mu)$	$\exp(\beta_{22}+X\gamma)$	$\exp(\beta_{23}+X\delta)$	(1 − row sum)	$\exp(\beta_{24}+X\theta)$
D, E	$\exp(\beta_{31}+X\mu)$	$\exp(\beta_{32}+X\gamma)$	$\exp(\beta_{33})$	$\exp(\beta_{34}+X\delta)$	(1 − row sum)

Note: ICF = intermediate care facility. Hosp. = hospitalization.

Table 9.8 **Test of Homogeneity: Maximum Likelihood Estimation of Parameters for Three-Month Matrix with Heterogeneity Correction**

Parameter	From	To	Coefficient	SD
β_{11}	A	Home, ICF	−2.46	.19
β_{12}	A	Hosp., Death	−2.87	.23
β_{13}	A	B, C	−1.16	.11
β_{14}	A	D, E	−3.90	.46
β_{21}	B, C	Home, ICF	−4.22	.18
β_{22}	B, C	Hosp., Death	−3.09	.11
β_{23}	B, C	A	−3.26	.13
β_{24}	B, C	D, E	−3.11	.12
β_{31}	D, E	Home, ICF	−4.32	.35
β_{32}	D, E	Hosp., Death	−2.29	.13
β_{33}	D, W	A	−5.14	.63
β_{34}	D, E	B, C	−1.43	.10
μ_1 AGE	A–E	Home, ICF	−.332	.067
γ_1 AGE	A–E	Hosp., Death	−.003	.069
δ_1 AGE	B–E	Better in NH	−.048	.074
θ_1 AGE	A–D	Worse in NH	.137	.072
μ_2 SEX	A–E	Home, ICF	.52	.23
γ_2 SEX	A–E	Hosp., Death	.34	.17
δ_2 SEX	B–E	Better in NH	.04	.19
θ_2 SEX	A–D	Worse in NH	.08	.19
μ_3 RACE	A–E	Home, ICF	.16	.35
γ_3 RACE	A–E	Hosp., Death	−.41	.28
δ_3 RACE	B–E	Better in NH	−.19	.27
θ_3 RACE	A–D	Worse in NH	−.09	.32
μ_4 MARRIED	A–E	Home, ICF	.13	.31
γ_4 MARRIED	A–E	Hosp., Death	.37	.20
δ_4 MARRIED	B–E	Better in NH	−.03	.23
θ_4 MARRIED	A–D	Worse in NH	.14	.29

No. of transitions = 2,512
−Log (likelihood) = 1,983.05

Note: Sample is all people in control group nursing homes admitted after 1 May 1981. AGE = (true age − 80)/10. SEX = 1 if male, 0 else. RACE = 1 if nonwhite, 0 else. MARRIED 1 if married now, 0 else. Hosp. = hospitalization. ICF = intermediate care facility. NH = nursing home.

Table 9.9 **Test of Homogeneity: Estimated Three-Month Markov Transition Matrix**

	Next Period				
This Period	Home, ICF	Hosp., Death	A	B, C	D, E
	Base case: 80-Year-Old Single White Woman				
A	8.5	5.7	52.3	31.5	2.0
B, C	1.5	4.6	3.8	85.6	4.5
D, E	1.3	10.2	.6	23.8	64.1
	AGE: 90-Year-Old Single White Woman				
A	6.1	5.7	56.2	30.0	2.0
B, C	1.1	4.5	3.7	85.6	5.1
D, E	0.9	10.1	.6	27.4	41.0
	SEX: 80-Year-Old Single White Man				
A	14.3	8.0	42.9	32.8	2.0
B, C	2.5	6.4	4.0	82.3	4.8
D, E	2.2	14.3	.6	25.7	57.2
	RACE: 80-Year-Old Single Nonwhite Woman				
A	10.0	3.8	58.2	26.0	2.0
B, C	1.7	3.0	3.2	88.0	4.1
D, E	1.6	6.8	.6	21.7	69.3
	MARRIED: 80-Year-Old Married White Woman				
A	9.7	8.2	49.4	30.7	2.0
B, C	1.7	6.6	3.7	82.9	5.1
D, E	1.5	14.8	.6	27.4	55.7

Note: ICF = intermediate care facility. Hosp. = hospitalization.

Table 9.10 **Ordered Logit Test for Homogeneity**

States
Go home, or go to ICF
Get better, but stay in NH[a]
Stay the same
Get worse but stay in NH[b]
Go to hospital or die

Note: ICF = intermediate care facility. NH = nursing home.

[a] Does not apply to state A.

[b] Does not apply to state E.

Table 9.11 **Test of Homogeneity: Results of Ordered Logit Models**

Variable	A	B	C	D	E
α_1[a]	2.72	−1.33	−4.46	−7.60	−9.4
	(.57)	(.40)	(.38)	(.64)	(1.1)
α_2[a]	4.14	−.69	−2.64	−4.80	−6.7
	(.59)	(.40)	(.39)	(.61)	(1.2)
α_3[a]	5.75	1.72	−.01	−2.62	−4.2
	(.58)	(.41)	(.40)	(.60)	(1.2)
α_4[a]		2.82	.49	−2.38	
		(.40)	(.39)	(.60)	
TIME (days)	.435	−.051	−.543	−.834	−1.42
	(.048)	(.028)	(.029)	(.038)	(.17)
TIME^2	−.0145	.0067	.0294	.0422	.076
	(.0021)	(.0011)	(.0012)	(.0013)	(.011)
NH type (treatment = 1)	−.35	−.305	−.029	.14	.24
	(.16)	(.100)	(.089)	(.13)	(.23)
AGE (true age)	.0230	.0117	.0111	−.0045	−.0009
	(.0067)	(.0048)	(.0042)	(.0066)	(.012)
SEX (male = 1)	−.01	−.01	.08	.05	−.82
	(.17)	(.11)	(.11)	(.17)	(.28)
RACE (nonwhite = 1)	−.31	.01	−.38	−.11	.50
	(.28)	(.18)	(.14)	(.21)	(.40)
MARRIED (married = 1)	.23	.40	.21	.33	−.07
	(.24)	(.15)	(.12)	(.19)	(.32)
−Log (likelihood)	972.91	2,284.69	2,662.65	1,103.85	334.57
% predicted correctly	53	53	60	60	64
N	779	1,737	2,241	996	370

Note: Standard deviations are in parentheses. NH = nursing home.

[a]Dummy cutoff variables.

The coefficients for personal characteristics vary greatly in their effect. For example, a positive coefficient on AGE[5] means that older people are more likely to be less healthy at their next assessment. SEX and RACE seem to have no effect. AGE and MARRIED have significant and positive coefficients in three models. However, by far the most significant coefficients are those for TIME (since admission) and TIME^2. Although it would be nice to be able to control for age and marital status in the Markov model, the coefficients for TIME since admission (and TIME^2) suggest that duration should be an essential part of any model of transitions in and out of nursing homes.

The conclusion seems to be that personal characteristics have only weak effects on the transition probabilities, particularly race. Including extra parameters slows down the computation substantially. The benefits of control-

[5] In this test, AGE equals true age.

ling for heterogeneity in a Markov model do not outweigh this cost. Other aspects of these data are more important to model.

9.5 Stationarity

The simple Markov model assumes that the probabilities are constant over time. Time can be measured on both a "relative" and an "absolute" scale, and the simple model requires that the probabilities are constant for both. Here, the relative scale measures time since admission. Stationarity on a relative scale refers to the assumption that the transition probabilities are independent of the length of time a person has been in the nursing home. The absolute scale refers to calendar time. In this case, stationarity means that there is no general trend over time. For example, all nursing homes are assumed to have adjusted instantly to the new system at the start of the experiment.

Since the stationarity assumption could fail in two different ways, it is tested in two different ways. The tests of stationarity, implications of failure, and corrections to the model are different for each method. The following section tests for duration dependence in a parametric duration model, and the one following tests whether the transition matrices for the experimental group change over the study period.

9.6 Duration Dependence

The simple Markov model assumes that transition probabilities are independent of how long a person has been in a nursing home. There are two reasons why this assumption may be false. Cumulative time spent living in a nursing home may affect probabilities directly. For instance, the initial move into a nursing home may be an unpleasant shock that fades as a person makes new friends and grows accustomed to the new surroundings. The second reason is that unobserved (or uncontrolled for) heterogeneity will cause duration dependence. Heckman and Singer (1984, p. 78) give the following explanation: "Intuitively, more mobility prone persons are the first to leave the population leaving the less mobile behind and hence creating the illusion of stronger negative duration dependence than actually exists."

Heckman and Singer go on to explain that ignoring heterogeneity will cause bias toward more negative duration dependence. It would therefore be incorrect to test for duration dependence without trying to control for heterogeneity. The two tests in this section control for the four observed characteristics: age, sex, race, and marital status. If the test rejects the hypothesis of no duration dependence, it could be due either to real duration dependence or to unobservable characteristics.

The first test treats time in the nursing home like another personal characteristic. A variable for time since admission is added to each cell. As before, a five by five matrix is used to estimate the three-month transition probabilities. In addition to time, both age and sex were used as explanatory variables (race and marital status were shown to be less important by the heterogeneity tests in sec. 9.1). The null hypothesis is rejected if the parameters for time are significantly different than zero.

The effects of time in this model are insignificant, as shown in table 9.12. Not one parameter differed significantly from zero. Not surprisingly, the parameters for age and sex were very close to values estimated in the section on heterogeneity. Thus, it does not seem necessary to control for duration dependence in this model.

The second test for duration dependence is a standard parametric duration model, the Weibull model, which measures both duration dependence and

Table 9.12 Test of Duration Dependence: Maximum Likelihood Estimation of Parameters for Three-Month Matrix with Time Heterogeneity Correction

Parameter	From	To	Coefficient	SD
β_{11}	A	Home, ICF	−2.38	.29
β_{12}	A	Hosp., Death	−2.88	.28
β_{13}	A	B, C	−1.17	.20
β_{14}	A	D, E	−3.89	.46
β_{21}	B, C	Home, ICF	−4.11	.35
β_{22}	B, C	Hosp., Death	−3.02	.22
β_{23}	B, C	A	−3.29	.23
β_{24}	B, C	D, E	−3.10	.23
β_{31}	D, E	Home, ICF	−4.38	.48
β_{32}	D, E	Hosp., Death	−2.23	.22
β_{33}	D, E	A	−5.15	.64
β_{34}	D, E	B, C	−1.49	.19
μ_1 AGE	A–E	Home, ICF	−.345	.069
γ_1 AGE	A–E	Hosp., Death	−.023	.067
δ_1 AGE	B–E	Better in NH	−.052	.068
θ_1 AGE	A–D	Worse in NH	.127	.071
μ_2 SEX	A–E	Home, ICF	.55	.24
γ_2 SEX	A–E	Hosp., Death	.41	.16
δ_2 SEX	B–E	Better in NH	−.02	.18
θ_2 SEX	A–D	Worse in NH	.05	.19
μ_3 TIME	A–E	Home, ICF	−.049	.060
γ_3 TIME	A–E	Hosp., Death	−.011	.042
δ_3 TIME	B–E	Better in NH	−.022	.044
θ_3 TIME	A–D	Worse in NH	−.012	.047

No. of transition = 2,512
−Log (likelihood) = 1,986.57

Note: Sample is all people in control group nursing homes admitted after 1 May 1981. AGE = (true age − 80)/10. SEX = 1 if male, 0 else. TIME = time since admission. ICF = intermediate care facility. Hosp. = hospitalization. NH = nursing home.

Table 9.13 **Test of Duration Dependence: Weibull Model**

Control/ Experiment	Type at Admission	Duration Parameter (α)	CONSTANT (β_0)	AGE (β_1)	MALE (β_2)	NONWHITE (β_3)	MARRIED (β_4)
Control	A	.628	−3.43	−.161	.27	.09	.31
		(.025)	(.15)	(.037)	(.11)	(.16)	(.15)
Control	B	.677	−4.12	−.336	.03	−.49	.34
		(.033)	(.21)	(.058)	(.16)	(.22)	(.18)
Control	C	.683	−4.34	−.076	.42	.30	.40
		(.028)	(.18)	(.047)	(.11)	(.17)	(.13)
Control	D	.577	−3.09	.049	.24	.09	.12
		(.024)	(.15)	(.056)	(.12)	(.17)	(.15)
Control	E	.701	−3.54	.056	.50	.26	.04
		(.052)	(.31)	(.088)	(.23)	(.32)	(.26)
Experiment	A	.719	−3.46	−.131	.06	.12	−.31
		(.042)	(.24)	(.076)	(.20)	(.30)	(.25)
Experiment	B	.764	−4.45	−.137	.26	−.31	−.31
		(.054)	(.34)	(.072)	(.19)	(.32)	(.30)
Experiment	C	.814	−4.90	−.006	.44	−.06	.17
		(.044)	(.27)	(.066)	(.16)	(.19)	(.17)
Experiment	D	.695	−3.86	−.037	.48	−.28	.02
		(.039)	(.24)	(.067)	(.17)	(.22)	(.18)
Experiment	E	.658	−3.51	.22	−.19	.02	.48
		(.052)	(.32)	(.11)	(.25)	(.27)	(.24)

Note: Standard deviations are in parentheses. AGE = (true age − 80)/10. SEX = 1 if male, else 0. RACE = 1 if nonwhite, else 0. MARRIED = 1 if married now, else 0. TIME = days.

heterogeneity. This test sacrifices many aspects of the Markov model, but it has the advantage of being easy to compute. The Weibull model has only two states: a person is either in or out of a nursing home. Because duration depends on a person's health, and because health usually remains constant, all residents were grouped according to their health at admission (types A–E). Also, because the model was estimated on both the control and experimental samples, this provides confirmation that the average length of stay was shorter in the experimental group.

The hazard function for the Weibull model depends on time-invariant characteristics X and a duration parameter α:

$$h(t|X) = \alpha t^{\alpha-1}\exp(X\beta)$$

X includes a constant term and variables for age, sex, race, and marital status, as defined before. The parameter α distinguishes the Weibull model from the exponential model, which has a constant hazard. If α is greater (less) than one, the model has positive (negative) duration dependence.

The results from the Weibull model tests are shown in table 9.13. The most striking result is that the estimated duration parameter, $\hat{\alpha}$, is significantly less than one in all cases. The models have *negative* duration dependence, which means that the probability of leaving the nursing home declines over time. Note that, for all types except E, $\hat{\alpha}$ is greater in the experimental group than in the control, so duration dependence is less pronounced in experimental group nursing homes. However, the constant term for experimental group nursing homes is the same or more negative. This implies that, at admission, the hazard rate is no larger in the experimental group, but over time the hazard rate declines more slowly. This partially explains the shorter length of stay found in Norton (1990).

Once again, the results on heterogeneity are mixed. Age and sex are significant in about half the cases. Surprisingly, older people have lower hazard rates than younger, and men have higher hazard rates than women. Only race is clearly insignificant.

The results from the Markov and Weibull tests are quite different. When time terms were added to the Markov model, the results were insignificant. However, the Weibull model showed strong results that there is strong negative duration dependence. It is not clear why time should be so much more significant in one model than the other.

9.7 Learning Effect

Whenever a new program is put into effect, it takes time for the participants to adjust to the new system. Too often, though, economists assume that people adjust instantly and perfectly at the start of a new program. The NCHSR nursing home experiment was a complicated system of incentives and assessments. Although the nursing homes did have a six-month period in which to learn about the new reimbursement system, they did not know whether they would be in the control or the experimental group until the day before the reimbursement system went into effect. This was good for experimental design in some ways, but it meant that nursing homes in the experimental group needed time to adjust.

In a previous paper (Norton 1990), I found that the distribution of types of health at admission changed slowly from the beginning to the conclusion of the study. The nursing homes did not adjust immediately to the new admission incentives. If they also did not adjust promptly to the outcome and discharge incentives, then the test in the previous paper may be biased toward no effect. Suppose that the experimental group took a while to hire new nurses and set new operational procedures. Then the effects of the experiment would not appear in the Markov transition matrix until after several months had passed.

To test whether there was a learning period, the data for the experimental group were split into two parts: those admitted during the first six months of the study and those admitted thereafter.[6] Two-week transition matrices were

[6] The background period lasted from 1 November 1980 to 30 April 1981. All thirty-six nursing homes were then divided randomly into one of two groups. Anyone admitted to an experimental

estimated using the continuous time model for people admitted during each period.

The results for the two groups are clearly different, as can be seen in tables 9.14 and 9.15. The second group has a much shorter average length of stay and better outcomes in general. In particular, the probability of dying decreased, and the probability of going home increased, for almost all types. Therefore there is strong evidence that it took the nursing homes in the experimental group time to adjust, and the results in the previous paper may be underestimated.

9.8 Markov Assumption

The key step in correcting the problem that the times between observations varied widely is to use the identity from Markov processes that

$$P(T) = P(1)^T,$$

where $P(T)$ is the matrix of transition probabilities over T time periods. I call this the *Markov assumption*. This allows parameters from the transition matrix of any time interval to be expressed in terms of the parameters of the shortest time interval. This assumption is false if the probabilities depend on anything other than a constant. If the model is not first order *and* homogeneous *and* stationary (in both senses), it will fail this test. Testing the Markov assumption is therefore a good summary test of specification error.

On the other hand, the test is not valid if there is selection bias on the basis of time interval. Here, selection bias means that certain types of transitions are oversampled (or undersampled) at particular frequencies because of the way in which the data were collected. For example, if all transitions within the nursing home are observed at three-month intervals but people leave at random times, then there is selection bias. The estimated matrix for any time other than three months will have positive probabilities only for leaving. A test of the Markov assumption on this data would fail simply because of the selection bias. The nursing home data have this problem since almost all observations over a short time period are of people entering or leaving a nursing home. Therefore, the test includes only transitions in the nursing home.

The test of the Markov assumption compares transition matrices of different time periods ($\hat{P}(T)$ for different T). Since the data for each matrix have the same time interval, estimation is easy. The continuous time correction is not needed, so an element $\hat{P}_{ij} = n_{ij}/\sum_j n_{ij}$. The test uses control group nursing home data for periods of two, three, and four months.

group nursing home during the following year was eligible for three types of incentives. Although residents were reassessed after 1 May 1982 and could earn bonuses for the nursing home, anyone admitted after that time was not eligible to earn bonuses.

Table 9.14 Test of Learning Effect: Early Part of Study

	Next Period								
	Home	ICF	Hosp.	Death	A	B	C	D	E
A	4.6	2.5	2.1	.6	83.5	6.1	.4	.2	.0
B	.7	1.7	1.2	.5	.9	91.6	3.2	.2	.0
C	.3	.2	1.0	.9	.1	2.9	92.6	1.8	.2
D	.1	.1	2.0	3.2	.1	.2	4.6	88.8	.9
E	.0	.5	1.8	4.5	.0	.6	1.6	.7	90.3

	Length of Stay	
	Mean	Median
A	16.6	8
B	25.6	17
C	30.0	22
D	23.3	15
E	18.5	11

	Probability That Person Starting in State Leaves Nursing Home in State			
	Home	ICF	Hosp.	Death
A	35	27	26	13
B	17	30	31	21
C	13	17	36	34
D	7	10	36	48
E	4	11	29	57

Note: ICF = intermediate care facility. Hosp. = hospitalization.

The estimated matrices show that the Markov assumption does not hold in the nursing home data (see table 9.16). The left-hand column has one-period matrices, which are not comparable. The right-hand column has these matrices raised to the sixth, fourth, and third powers; they are therefore comparable since the time interval is forty-eight weeks (almost one year). As a general misspecification test, this confirms the results in the previous sections that the simple Markov model was misspecified.

9.9 Measurement Error

Poterba and Summers (1986) showed that even a small probability of reporting error can lead to large errors in duration estimates when using a Markov model. They adjusted labor market transition probabilities using reporting errors in the Census Population Survey. Reporting errors distort the true

Table 9.15 **Test of Learning Effect: Late Part of Study**

This	Next Period								
Period	Home	ICF	Hosp.	Death	A	B	C	D	E
A	4.6	2.7	1.8	1.2	85.0	3.7	1.0	.0	.0
B	.8	1.0	1.5	.7	3.2	88.7	3.3	.0	.8
C	.3	.4	2.0	1.1	2.8	.8	92.4	.0	.2
D	.2	2.5	2.5	3.1	.0	.0	1.1	90.4	.2
E	.0	.0	3.9	2.9	.0	.0	1.5	3.4	88.3

	Length of Stay (3 months)	
	Mean	Median
A	12.8	8
B	19.5	14
C	20.3	15
D	13.1	9
E	14.9	10

	Probability That Person Starting in State Leaves Nursing Home in State			
	Home	ICF	Hosp.	Death
A	38	24	23	14
B	24	21	35	20
C	21	17	40	23
D	5	28	32	36
E	4	10	48	38

Note: ICF = intermediate care facility. Hosp. = hospitalization.

probabilities by overestimating the frequency of transitions between different states. A person who is unemployed for a long time and who misreports being employed makes it seem as if he has much shorter spells of unemployment. After adjusting for error rates of only 5 percent, Poterba and Summers found that true spells of unemployment were as much as 80 percent longer than conventional measures.

The nursing home data may be subject to a similar problem. Nurses assessed each resident's health periodically according to somewhat subjective criteria. The decision was based primarily on a person's activity of daily living (ADL) index. Although the ADL index was designed to be an objective measure of disability in performing basic functions, in practice there is some subjectivity in assessing whether another person can do something satisfactorily. It is especially difficult to judge whether a person is type A, which includes people expected to be sent home within ninety days, regardless of current

Table 9.16 Test of Markov Assumption

	2 Months					→	1 Year (48 weeks)				
	A	B	C	D	E		A	B	C	D	E
A	60	25	15	0	0		11	46	36	7	0
B	5	75	18	2	0		8	47	38	8	0
C	2	23	69	6	0		7	43	40	10	0
D	2	2	31	65	0		6	36	43	15	0
E	0	0	25	0	75		4	29	42	7	18
	3 Months					→	1 Year (48 weeks)				
	A	B	C	D	E		A	B	C	D	E
A	52	36	9	1	1		17	44	31	6	2
B	11	67	20	2	0		13	43	36	7	2
C	2	20	70	8	1		8	34	44	12	2
D	1	4	29	61	5		6	24	44	22	5
E	7	13	25	5	50		10	32	39	10	8
	4 Months					→	1 Year (48 weeks)				
	A	B	C	D	E		A	B	C	D	E
A	27	60	6	3	3		12	46	32	8	3
B	13	57	26	3	0		10	44	37	7	2
C	2	26	66	4	3		7	36	45	8	4
D	0	10	29	54	7		4	25	41	24	7
E	0	0	14	43	43		2	16	36	34	13

ADL level. Less than half were actually discharged on time, so there was considerable error for this category.

The NCHSR study tried to minimize the problems of measurement error (see Applied Management Sciences 1986, apps. 8–10). Nurses were trained during a six-month baseline period before the study began. There were four types of tests used throughout the study to check reliability. In two of the tests, nurses assessed residents in pairs, either concurrently or successively. Also, thirty-three residents were videotaped while being assessed. These tapes were reviewed periodically to check agreement between nurses. These reliability studies give some idea of the magnitude of the reporting error problem.

This section adjusts the observed transition probabilities, given knowledge of the measurement errors, to estimate the true transition probabilities P^*. Following Poterba and Summers, we will estimate flows of people, F, then convert the flows back to probabilities. Let f^*_{ij} and f_{ij} be the true and the observed numbers of people to go from state i to state j in a single period. Define Q to be the matrix of error rates, where q_{ij} = Pr(observed state = j | true state = i). The estimated flows, F, depend on Q and F^* as follows:

$$f_{ij} = \sum_{k,l} q_{ki} f^{*}_{kl} q_{lj}.$$

In matrix notation,

$$F = Q'F^{*}Q.$$

This can be rewritten to solve for F^{*} and then converted to P^{*} by dividing by the row sum:[7]

$$F^{*} = (Q^{-1})'F(Q^{-1}).$$

The reliability tests did not estimate the matrix Q explicitly, so we must use limited information to construct a plausible Q. First, note that the measurement error is a problem only when classifying residents in the nursing home. A misreport of where someone went after discharge is more likely to be a clerical than a judgment error. Classification within the nursing home depends on three judgments: need for special care, ADL level, and whether dischargeable within ninety days. Only the last two are subject to much discretion since it should be clear whether someone is comatose, requires tube feeding, or receives decubitus ulcer care. Therefore, type E is always assumed to be correctly classified. The reliability tests found that trained nurses agreed 90 percent of the time on residents' ADL level. It is plausible to assume that each nurse reports the truth 95 percent of the time and that errors are uncorrelated. Then a person who is truly type B, C, or D (where classification depends on ADL level) would be misclassified 5 percent of the time (types B and D reported as C, type C reported as either B or D). The most important test distinguishing types A and B is the Mental Status Questionnaire (see Applied Management Sciences 1986, app. 8A, p. 7). Since this also had an agreement of 90 percent, assume that 5 percent of As and Bs each were misclassified as the other type. A plausible Q is then

$$Q = \begin{bmatrix} .95 & .05 & 0 & 0 & 0 \\ .05 & .90 & .05 & 0 & 0 \\ 0 & .025 & .95 & .025 & 0 \\ 0 & 0 & .05 & .95 & 0 \\ 0 & 0 & .10 & 0 & 1 \end{bmatrix}.$$

The matrix F was constructed from a three-month probability matrix, not directly from data on flows. It would be wrong to use the estimated two-week matrix for P since the average time between observations is three months, and this would imply misreporting at two-week intervals. Therefore, we use $P = P(2 \text{ week})^{6}$, estimated in Norton (1990). To convert from probabilities

[7] The reader may wonder why Q was not defined in the obvious way to avoid taking inverses. Meyer (1988) compares these two methods and concludes that the second depends on a subtle but implausible assumption.

to flows, each column of the probability matrix was multiplied by a weighting vector of the probability distribution over the five types at admission.

Unlike in Poterba and Summers, reporting errors seem to have little effect, as shown in table 9.17 compared to the corrected values in table 9.18. The matrices P^* and P differ by a few percentage points in many cells, and in general the off-diagonal terms are smaller in P^*. There is almost no difference, though, in the length of stay and in the probability of ending in each of the absorbing states. The differences are much smaller than between the control and the experimental groups. Even when the errors estimated in Q were increased dramatically, these basic results were unchanged.

There are several conclusions that can be drawn from these findings. Poterba and Summers probably found larger effects because all their states were liable to be misclassified, while only four of the nine nursing home states were. Furthermore, the major results, such as average length of stay and probability of ending in each absorbing state, are not very sensitive to measure-

Table 9.17 **Test of Measurement Error: Uncorrected Three-Month Markov Transition Matrix**

	Next Period								
	Home	ICF	Hosp.	Death	A	B	C	D	E
A	17.3	10.4	12.0	2.5	31.0	21.0	4.5	.8	.5
B	2.8	3.0	7.1	2.6	5.4	61.7	15.8	1.3	.2
C	1.4	.9	7.5	5.1	.9	13.7	64.4	5.5	.6
D	.9	.4	11.0	11.3	.4	2.1	17.2	54.6	2.2
E	3.2	.3	18.0	23.3	1.6	4.3	10.9	3.2	35.2

	Length of Stay	
	Mean	Median
A	3.7	2
B	5.7	4
C	5.9	4
D	4.9	3
E	3.2	2

	Probability That Person Starting in State Leaves Nursing Home in State			
	Home	ICF	Hosp.	Death
A	31	20	35	13
B	17	15	45	22
C	13	10	47	30
D	8	6	46	39
E	9	4	42	45

Note: ICF = intermediate care facility. Hosp. = hospitalization.

Table 9.18 **Test of Measurement Error: Corrected Three-Month Markov Transiton Matrix**

This Period	Next Period								
	Home	ICF	Hosp.	Death	A	B	C	D	E
A	17.3	10.4	12.0	2.5	33.9	19.3	3.3	.8	.5
B	2.8	3.0	7.1	2.6	.4	71.7	11.2	.9	.2
C	1.4	.9	7.5	5.1	.2	11.4	70.5	2.4	.6
D	.9	.4	11.0	11.3	.3	1.5	12.6	59.8	2.3
E	3.2	.3	18.0	23.3	1.5	4.3	11.0	3.0	35.2

	Length of Stay	
	Mean	Median
A	3.7	2
B	6.2	5
C	6.3	5
D	4.9	3
E	3.4	2

	Probability That Person Starting in State Leaves Nursing Home in State			
	Home	ICF	Hosp.	Death
A	31	21	35	13
B	15	15	46	23
C	12	9	48	30
D	7	5	47	41
E	9	4	42	45

Note: ICF = intermediate care facility. Hosp. = hospitalization.

ment error within the nursing home. This would be different if nurses could not tell reliably whether a person had gone home or died. These results also put the conclusions of section 9.1 in perspective. A quick calculation from table 9.7 shows that shifting 10 percent of the sample would eliminate the effect that people return to their past state. Thus, about half the rebound effect found in the first-order section is probably due to measurement error.

9.10 Conclusion

This paper tested the assumptions of the simple Markov model on nursing home data and found that several tests failed. The model is not first order but second order. In particular, people tend to rebound to the state they were in last period more than a first-order model would predict. The Weibull model shows that there is strong negative duration dependence. Dividing the data into two parts shows that the nursing homes in the experimental group took time to adjust to the reimbursement system. Finally, a test of the Markov assumption, as a general specification test, failed.

Some tests supported the use of the simple model. In a variety of tests, heterogeneity seemed to have weak effects. Thus, the increase in complexity by controlling for heterogeneity overshadows any gains in information. Correcting for measurement error has almost no effect on the average length of stay or on the probability of ending in the absorbing states, only small effects on transitions within the nursing home. In light of this, controlling for second-order effects does not seem worthwhile, especially since about half the rebound effect was due to measurement error. Also, the duration dependence that is so strong in Weibull models was not detected in a Markov model. Finally, the fact that the nursing homes in the experimental group did not adjust instantly means that the results of the previous paper are underestimated.

The Markov model should be viewed as a reasonable but imperfect model of transitions in nursing homes. Research in this area could benefit from trying other kinds of duration models, such as competing hazard and semiparametric. These models may have advantages in speed of computation, a more flexible form, and an emphasis on duration and outcome that are important for public policy.

References

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press.

Anderson, T. W., and Leo A. Goodman. 1957. Statistical Inference about Markov Chains. *Annals of Mathematical Statistics* 28:89–110.

Applied Management Sciences Inc. 1986. Controlled Experiment to Evaluate the Impacts of Incentive Payments on Nursing Home Admissions, Discharges, Case Mix, Care, Outcomes, and Costs: Documentation Report. Prepared for the National Center for Health Services Research and Health Care Technology Assessment. Rockville, Md.: National Technical Information Services, February. Manuscript.

Börsch-Supan, Axel, Laurence J. Kotlikoff, and John N. Morris. 1988. The Dynamics of Living Arrangements by the Elderly. NBER Working Paper no. 2787. Cambridge, Mass.: National Bureau of Economic Research.

Heckman, James J., and Burton Singer. 1984. Economic Duration Analysis. *Journal of Econometrics* 24 (1/2):63–132.

Katz, Sidney, and C. A. Akpom. 1976. Index of ADL. *Medical Care* 14 (5):116–18.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Meyer, Bruce D. 1988. Classification-Error Models and Labor-Market Dynamics. *Journal of Business and Economic Statistics* 6 (3):385–90.

Norton, Edward C. 1990. Incentive Regulation of Nursing Homes. Cambridge, Mass.: Massachusetts Institute of Technology, March. Manuscript.

Poterba, James M., and Lawrence H. Summers. 1986. Reporting Errors and Labor Market Dynamics. *Econometrica* 54 (6):1319–38.

Weissert, William G., William J. Scanlon, Thomas T. H. Wan, and Douglas E. Skinner. 1983. Care for the Chronically Ill: Nursing Home Incentive Payment Experiment. *Health Care Financing Review* 5 (2):41–49.

Comment Sherwin Rosen

This is a promising empirical analysis of whether monetary incentives affect the selection and length of stay of nursing home residents. Using data from a Medicaid experiment in San Diego, Edward C. Norton finds that incentives apparently caused shorter lengths of stay and more frequent admissions of patients in poorer health. These results are interesting but, like all experiments, must be reenforced by replication in independent samples.

The data description is somewhat confusing or incomplete. We know that Medicaid finances nearly half of nursing home patients, but those reimbursements do not nearly cover average costs. To achieve financial viability, nursing homes with larger fractions of Medicaid patients must either offer a lower quality of service to Medicaid patients or cross-subsidize them by charging non-Medicaid patients more. The experiment subsidized admissions of Medicaid patients in worse states of health and rewarded favorable outcomes in selected cases, but we are not told what proportions of *all patients* were covered by the schemes or indeed the relative frequency of outcome and discharge bonus used in the experiment. A reader wants to know how important these subsidies were to the overall operations of experimental units and whether the payments were large enough or frequent enough to have a plausible effect. Knowledge of the proportions of Medicaid and non-Medicaid patients in the sample nursing homes might also be useful in assessing the general quality of care in the sample. Quality of care is known to vary among nursing homes, but the monetary incentives in the experiment were not conditioned on nursing home characteristics. If lower-quality units were more eager to participate in the experiment, the results might be biased toward the null hypothesis.

In assessing the results, it should also be borne in mind that the experimental design did not satisfy some commonly accepted rules. In particular, there were no double-blind safeguards in assigning patients to categories and in assessing outcomes. The same skilled nursing staff was employed for both. Complete double-blind safeguards are obviously impossible in experiments of this kind, but independent assessments of initial classifications and final outcomes would have made for a better experiment. And since many of these

Sherwin Rosen is the Edwin A. and Betty L. Bergman Professor of Economics and chairman of the Department of Economics, University of Chicago, and a research associate of the National Bureau of Economic Research.

assignments and assessments were done jointly by both the house staff and the experimental skilled nursing staff, there is potential bias if the classification criteria vary according to unobserved quality of care. Were the data available, it would be very interesting to examine the experience of non-Medicaid patients in both experimental and control group nursing homes. Using "within"–nursing home differences in patient types as well as "between"–experimental and control group nursing home differences in Medicare patients might help control for differences in nursing home quality and possible differential assignment of cases among them.

Studies of labor market and geographic mobility have found that identifying some agents as "stayers" and others as "movers" is necessary to fit the data. So many incomplete spell lengths make these distinctions more difficult to test in these data. Still, it must be noted that Norton's estimates of mean and median length of stay by category are based on manipulations of the estimated transition matrix, not on direct observation, because many spells are still in progress at the end of the experiment. This is a limitation of the experiment, certainly not of Norton's methodology, but it must in some sense increase the standard error of estimated experiment effects. If in addition some of the stayer-mover logic is applicable to nursing home residents, a resident's initial state is not a sufficient statistic for probable future states. How persons arrive at that state affects the unconditional duration estimates, and those numbers (computed in tables 9.2 and 9.4) may not be accurate. One wonders why the observed assessment intervals of residents vary so much and whether they are behaviorally related to the mover-stayer (or permanent-temporary) assessment of a resident's condition at the starting times.

It can be argued that Medicare patients in nursing homes are a more homogeneous group in the above sense because they are likely to have "spent down" any other insurance or private resources in earlier hospitalization or nursing home stays. Then these residents are more likely to be "stayers" (or permanent residents), and the strict stationary, homogeneous Markov model may be a reasonable approximation. Knowledge of residents' previous history would be very useful in assessing the importance of this point. But assuming it is true, by what mechanism do these financial incentives work to reduce turnover and length of stay? Except for the patients initially assessed in state A, there are no substantial differences in the sum of ultimate hospitalization and death probabilities between treatment and controls in tables 9.2 and 9.4. The main effects on spell length occur for persons in the better initial states. Perhaps this is as it should be. We should not be subsidizing something that cannot occur. Yet insofar as these subsidies focus greater care and attention on the temporary residents, they promote a kind of adverse selection against the most difficult and costly ("permanent") residents who may be in greatest need of care. We must assess these schemes not only in terms of the monetary costs to the Medicare system but also with regard to the values of service among various classes of patients, including those for whom improvement and rehabilitation is very unlikely.

In using a flat fee reimbursement system independent of patient condition, the existing system promotes adverse selection of the easier and less costly cases. The system investigated in the experiment does not factor in the costs of classifying or the potential abuses and moral hazard problems arising if the entire system were converted to the experimental reimbursement mechanism. The hothouse environment of the experiment does not produce any data whatsoever on these latter costs, which may be substantial in any feasible system. These costs must be weighed against any efficiency gains in levels of care and lengths of stay that these financial incentives provide.

This Page Intentionally Left Blank