Chapter Two: Modeling Program Participation and Error

This analysis' modeling approach focused on the underlying research questions, the available program data, and the intrinsic nature of food stamp administrative procedures. The research builds upon previous work by Abt Associates on the effects of recertification and monthly reporting on food stamp error rates.[7] Importantly, however, the current research goes beyond the earlier work, in the following respects:

- By incorporating food stamp participation, as well as food stamp error, as an outcome of administrative procedures;
- By relaxing the previous modeling assumption that the observed error rates reflect a system already at equilibrium in any given year; and
- By undertaking state-by-state estimates, as well as national estimates, of the model.

The foundation for this work, however, remains the concept that the food stamp administrative process can be represented as a discrete-time, multi-status probability model.[8] The basic unit of observation is the household, and the basic unit of time is the month. At the start of any given month, each household is considered as occupying one of several possible groups regarding their participation in the program and the accuracy of their payment.

The simplest form of such a model is the discrete-time Markov chain, a stochastic process in which the conditional distribution of each unit's future status is dependent on its present status, but is independent of any prior history. The system is thus considered "memoryless." There is a very substantial body of literature on Markov chains, and the attached bibliography identifies a very small subset of the published work. Appendix E shows a formal specification of discrete-time Markov chains.

Under a Markov model, the period-by-period operations of a real-world system are described in terms of a matrix of transition probabilities. The transition probabilities indicate the pattern of changes in the status of items—in this instance, the status of households regarding their food stamp participation and the correctness of their benefit—as time advances from one period to the next—in this context, from one month to the next. The estimated transition

---

7 See Mills (1988).

8 To avoid confusion in the current research, we use the word "status" rather than "state" in referring to the condition occupied by a household with respect to the Food Stamp Program. We reserve the use of "state" in referring to the 50 states and the District of Columbia.

matrix can then be used to simulate the effect on the system's performance of operational changes. The particular application of Markov modeling here is to examine the effect of the more frequent food stamp recertification on the size of the active food stamp caseload and the error rate among active cases.

Illustrative Models

In describing below the methodology of a discrete-time Markov chain, we first introduce a basic two-group model and then move progressively to the five-group model used in the analysis.

Two-Group Model

The basic two-group model can be used to examine changes over time in the percentage of households that are food stamp participants. We refer to this percentage as the "aggregate participation rate," expressed as p_t in any given month t. In this simple model, the household population can be viewed as divided into two subgroups: food stamp nonparticipants and food stamp participants.[9] The condition of the system in any given month can be described by p_t.

From one month to the next, any given household can either shift from one group to the other or can remain within its group. We can describe the monthly dynamics of this system using a two-by-two transition matrix, as shown in Exhibit 3. *The entries of the transition matrix indicate the probability that a household of given status in the current month (as indicated by the associated row) will have the same status or a different status in the next month (as indicated by the associated column).* The entries in each row are conditional probabilities that sum to 1, fully describing the transitions that can occur for a household, given its current-month status.

Exhibit 3: Transition Matrix for Two-Group Model

		Next-month status:		
Current-month status:		**Nonparticipating**		**Participating**
Nonparticipating		1 – a		a
Participating		b		1 - b

9 In this study, we use the term "aggregate participation rate" to refer here to the percentage of *all* households that participate in the program. This is to distinguish it from the alternative measure (referred to here as the "conditional participation rate") that indicates the percentage of *program-eligible* households that participate.

The parameter “a” indicates the “case opening rate”—the probability that a household not receiving food stamps in one month will participate in the next month. Correspondingly, the parameter “b” represents the “case closure rate”—the probability that a household receiving food stamps in one month will not participate in the following month. This modeling approach assumes that each parameter does not depend on the length of time that a household has been in its current group. It also assumes that each parameter remains constant over time.

The system’s condition will change predictably from one month to the next. In particular, the percentage of households participating in the next month, p_{t+1}, can be expressed as a function of p_t, a, and b as follows:

(Eq. 1) $p_{t+1} = [(1-p_t)\, a] + [p_t\,(1-b)]$.

Depending on the values of a and b, the system will approach a long-term equilibrium.[10] From Eq. 1, one can derive the steady-state value of the aggregate participation rate, p*, by solving for $p_{t+1} = p_t$. Doing so, one finds that:

(Eq. 2) $p^* = a/(a+b)$

The value of p* is thus a function of the transition probabilities only and does not depend on the starting value of p_t.

Three-Group Model

One can elaborate on this basic two-group model by allowing a participating household to be either a correct case or an error case. In any given month, households can thus occupy one of three groups: nonparticipating households (group 1), correctly paid cases (group 2), and incorrectly paid cases (group 3). In this model, the condition of the system in month t is described by the three percentages (summing to one) that indicate the proportion of all households in each group: nonparticipating (n_t), correct (c_t), and error (e_t). The system can be regarded as having two key outcome statistics. One indicator, as before, is the aggregate participation rate, p_t. In the notation of the three-group model, $p_t = c_t + e_t = 1 - n_t$. The other indicator is the case error rate, r_t, the percentage of active cases that are in error in month t. In the terms of the model, $r_t = e_t/(c_t + e_t) = e_t/p_t$.

For this three-group model, the month-to-month transitions are fully described in the three-by-three matrix shown in Exhibit 4. As with the matrix for the two-group model, the entries

---

[10] There are some combinations of a and b for which no equilibrium is reached (for instance, if a=1 and b=1), but such scenarios are implausible ones that do not reflect the real-world monthly dynamics.

in each row sum to one. We have introduced subscripts for each cell entry (p_{ij}) to indicate the specific transition from group i in the current month to group j in the next month.

Exhibit 4: Transition Matrix for Three-Group Model

		Next-month status:				
Current-month status:		**Nonparticipating**		**Correct**		**Error**
Nonparticipating		p_{11}		p_{12}		p_{13}
Correct		p_{21}		p_{22}		p_{23}
Error		p_{31}		p_{32}		p_{33}

This three-by-three matrix provides a simplified representation of the monthly dynamics that underlie the case error rate. In any given month, the opening rate (indicated by "a" in the two-group model) is now the sum of p_{12} and p_{13}. Note that the error rate among initially certified cases is $p_{13}/(p_{12}+p_{13})$. The probability that a correct case becomes in error the following month is p_{23}. The probability that an error case becomes correct the following month is p_{32}. The probability that an error case leaves the active caseload is p_{31}.[11] This model does not enable one to examine the separate roles of interim action and recertification in controlling errors among active cases.

Five-Group Model

To address the issue of more frequent recertification and its effects on both participation and error in food stamps, the model must be further elaborated. We have done this by subdividing food stamp cases according to whether or not their current certification is about to expire (i.e., "ongoing" versus "expiring"), in addition to whether or not their benefit is in error. The five-by-five transition matrix is shown in Exhibit 5, with row totals again summing to one.

This formulation allows the dynamics of participation and error to be expressed separately for initial certifications (in the first row), interim actions (in the second and third rows), and recertifications (in the fourth and fifth rows). More frequent recertifications affect participation and error by subjecting cases to the conditional probabilities for expiring cases (in the fourth and fifth rows) rather than the conditional probabilities for ongoing cases (in the second and third rows).

[11] As indicated in Chapter Three, one can derive the equilibrium value of the case error rate (r*), as e*/p*, where p* is the equilibrium value for the share of all households that are food stamp participants (as defined above) and e* is the equilibrium value for the share of all households that are error cases.

Exhibit 5: Transition Matrix for Five-Group Model

Current-month status:	Next-month status: Non-participating	Ongoing correct	Ongoing error	Expiring correct	Expiring error
Nonparticipating	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}
Ongoing correct	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}
Ongoing error	p_{31}	p_{32}	p_{33}	p_{34}	p_{35}
Expiring correct	p_{41}	p_{42}	p_{43}	p_{44}	p_{45}
Expiring error	p_{51}	p_{52}	p_{53}	p_{54}	p_{55}

The condition of the system in any given month is now described by the five percentages indicating the shares of the total household population that are nonparticipating households (group 1), ongoing correct cases (group 2), ongoing error cases (group 3), expiring correct cases (group 4), or expiring error cases (group 5). As these categories are mutually exclusive and collectively exhaustive, the corresponding shares of the population will always sum to one.

Modeling Approach Adopted for This Study

With the preceding discussion as background, we now develop more formally the five-group model used in this study.

At the start of any given month, we regard each U.S. household as belonging to one of the following five groups, according to whether the household is participating in the Food Stamp Program and (if so) whether the household's food stamp payment is correct and whether the household is in the final month of its current food stamp certification period:

1. households not participating in the Food Stamp Program ("nonparticipating");
2. correctly paid food stamp cases, not in their final certification month ("ongoing correct");
3. incorrectly paid food stamp cases, not in their final certification month ("ongoing error");
4. correctly paid food stamp cases, in their final certification month ("expiring correct"); and
5. incorrectly paid food stamp cases, in their final certification month ("expiring error").

From one month to the next, each household may experience a transition from one group to another or may remain in its group. ***One can express the pattern of month-to-month changes in status in a five-by-five transition matrix. The entries in this matrix are transition probabilities of the form p_{ij}, indicating the probability that a household in group i in one month will enter group j in the following month (where i may equal j).***

In developing the initial estimates reported here, we have made several simplifying assumptions that rule out some month-to-month transitions. The corresponding entries in the transition matrix are thus set to zero by definition.

- We assume that a nonparticipating household (Group 1) cannot transition immediately to being an expiring case (i.e., in Group 4 or 5 as an active case in its final certification month). That is, we regard the shortest certification length as two months. (For the 0.5 percent of cases found in the data to have one-month certification periods, we treat these as two-month periods.)

- We also assume that active cases in their final certification month (Group 4 or 5) cannot remain in that group for more than one month at a time. That is, a case that becomes due for certification would not then proceed to being overdue for recertification.[12]

These assumptions are made to avoid having cells in the basic transition matrix that are occupied by a trivially small number of cases. In such instances, the simplifying assumptions serve to combine these negligible categories of cases into adjacent cells of the matrix, with negligible effect on the resulting estimates.

The purpose of adopting this general framework is that it enables the case error rate to be modeled in terms of three distinct administrative processes: initial certification, interim action, and recertification. In the terms of the model, the case error rate can be expressed as the number of households in Groups 3 and 5 divided by the number of households in Groups 2 through 5.

This model focuses on the *combined* or *total* case error rate, the percentage of active cases that are either ineligible, eligible but overpaid, or eligible but underpaid, including both agency-related and client-related error. The methodological approach could easily be used to construct separate models for the overpayment case error rate (including both the ineligible and those eligible but overpaid) and the underpayment case error rate (those eligible but underpaid). Alternatively, one could focus solely on agency-related error or client-related error. We examine the total measure here for several reasons. First, as noted above, federal standards for food stamp error are based on the combination of overpayment and

---

[12] The assumption also reflects a recognized limitation of the QC data. As will be explained later, the QC data do not accurately identify all cases that are overdue for recertification.

underpayment errors among active cases. The combined measure is thus the focus of executive and legislative attention. Second, to the extent that errors of both overpayment and underpayment are generated jointly from the same administrative process, it would be inappropriate for the analysis to separate them artificially, implying that one is independent of the other or that one carries more significance than the other.[13]

The decision to focus on the total *case* error rate, rather than the total *dollar* error rate, is a pragmatic one. The total case error rate, defined as the ratio of active cases in error to total active cases, is a proportional outcome bounded between zero and one for any set of cases. All active cases contribute equally to the denominator; each case contributes either zero or one to the numerator. In contrast, the total dollar error rate, the ratio of error dollars among active cases to total dollars paid to these cases, can potentially range higher than one for any set of cases. Moreover, cases contribute unequally to the denominator, depending on the magnitude of their monthly benefit, which is itself a function of the error amount. Their contribution to the numerator is also a continuous variable, as small as zero (for correct cases) and as large as the full benefit payment (for ineligible cases).

Arithmetically, the total dollar error rate equals the product of the total case error rate and the ratio of the average monthly dollar error amount to the average monthly benefit payment to an active case. A state's total case error rate is highly correlated with its total dollar rate. For Fiscal Year 2001, the correlation between these two measures was 0.923, implying that more than 85 percent (0.853) of the variation in the dollar error rate can be explained by the case error rate.[14]

---

[13] Similarly, separate models could be constructed for agency-related error and client-related error. The distinction between the two error types can be somewhat arbitrary, however, and one needs to establish rules for classifying multiple-error cases that contain both agency and client errors.

[14] The correlation coefficient of 0.923 is based on state-reported values of the total case error rate and the official values of the total dollar error rate in fiscal year 2001, for the 50 states and the District of Columbia.