Contract No.: AG-3198-C-06-0008
MPR Reference No.: 6280-008

MATHEMATICA
Policy Research, Inc.

NSLP/SBP Access, Participation, Eligibility, and Certification Study

Econometric Model for Updating Estimates of Erroneous Payments in the NSLP and SBP

June 2008

Quinn Moore
Philip Gleason
Michael Ponza

Submitted to:
U.S. Department of Agriculture
Food and Nutrition Service
Office of Analysis, Nutrition, and Evaluation
3101 Park Center Drive
Alexandria, VA 22302

Contracting Officer's Representative:
Dr. John Endahl
(703) 305-2122

Submitted by:
Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Michael Ponza

Principal Investigators: Philip Gleason
Michael Ponza

ACKNOWLEDGMENTS

We would like to thank several people who helped develop the model and prepare this report. John Endahl, of the Food and Nutrition Service (FNS's) Office of Analysis, Nutrition, and Evaluation (OANE), Susan Fouts at FNS's Child Nutrition Division, Phillip S. Kott, USDA, National Agricultural Statistics Service, Research and Development Division, and Chris R. Bollinger, Department of Economics, Gatton College of Business and Economics, University of Kentucky, reviewed the econometric model and provided comments which improved methodology and the final specification. From Mathematica, Elaine Hill, Laura Guy, Patricia Seunarine, Thomas McCall, Karin Zeller, Bonnie Hart, Amy Zambrowski, Carol Razafindrakoto, and Mathew Jacobus provided support for data file construction and statistical programming. James Ohls provided quality assurance. Jill Miller, Bill Garrett, and Cindy George provided production and editorial support.

CONTENTS

Chapter **Page**

TABLES

I. INTRODUCTION

As of the 2005–2006 school year (SY 2005–06), erroneous payments due to certification error for free or reduced-price school lunches provided under the National School Lunch Program (NSLP) equaled $759 million and amounted to 9.4 percent of all federal reimbursements for these meals (Ponza et al. 2007a). Erroneous payments due to certification error for the School Breakfast Program (SBP) equaled $177 million, or 9.2 percent of reimbursements paid for all SBP meals served. Certification error occurs when students who have applied for free or reduced-price meal benefits are not certified for the level of benefits which they are eligible or are denied benefits to which they are eligible. These findings emerged from the analysis of a nationally representative sample of students who applied or were directly certified for free or reduced-price meals as part of the Access, Participation, Eligibility, and Certification (APEC) study, conducted for the Food and Nutrition Service (FNS) by Mathematica Policy Research, Inc. (MPR).

FNS is using the information on erroneous payments due to certification errors from the APEC study to meet its reporting requirements to the Office of Management and Budget (OMB) and Congress under the Improper Payments Information Act of 2002 (IPIA). Under the IPIA, FNS is required to report on erroneous payments in its food and nutrition programs, including the NSLP and the SBP, that may be susceptible to significant erroneous payments (erroneous payments exceeding $10 million and 2.5 percent of program payments). The APEC study found that improper payments made in the NSLP and the SBP during SY 2005–06 are significant. As a result, FNS will need to report annually on the amount of estimated improper payments in the NSLP and the SBP, along with activities and progress in reducing them.

In order to enable FNS to update its erroneous payments rate estimates for the NSLP and the SBP during the next few years without having to conduct another full round of primary data collection, the APEC study included development of an econometric model that can be used to produce estimates of erroneous payments in future years based on more readily available district-level data. This report describes that model, discusses the process for using the model to predict future erroneous payment amounts and rates, and assesses the model's performance relative to the main APEC study findings for SY 2005 - 06.

A. ERRONEOUS PAYMENTS IN THE NSLP AND THE SBP SY 2005–06

The APEC study generated estimates of three different amounts and rates of erroneous payments resulting from certification errors for both the NSLP and the SBP—overpayments, underpayments, and gross total erroneous payments. Overpayments are defined as reimbursements made by the U.S. Department of Agriculture (USDA) to school districts for free or reduced-price meals served to students whose household circumstances indicated that they received a higher level of benefits than that for which they were eligible. Underpayments are reimbursements the USDA did not provide to districts for meals served to students who were certified for reduced-price benefits although they were eligible for free meals, or who applied for meal benefits but were not approved despite being from households whose circumstances indicated that they were eligible for either free or reduced-price meal benefits. The total erroneous payments measure is simply the gross sum of overpayments and underpayments. The remainder of this section describes the study design and methods used to estimate erroneous payments.

1. Sample Design

A multistage-clustered sample design was used to derive estimates of erroneous payments. We selected representative samples of school districts, schools, and free or reduced-price meal applicants participating in the NSLP and SBP in the contiguous United States and District of Columbia during SY 2005-2006. School districts that participate in the NSLP and/or SBP were selected first. Within each of the selected school districts, we selected a sample of public and private schools, and then selected students at the sampled schools who either were certified for free or reduced-price meals or had applied for but were denied these benefits (see Ponza et al. 2007b, Appendix A, for detailed description of samples and sample selection procedures).

The main study samples include the following:

1. **87 school food authorities (SFAs) that administer the meal programs, of which 78 are public and 9 are private.** In selecting the school districts, we explicitly stratified the sample by whether the districts were large enough to be selected with certainty. The non-certainty stratum was then stratified on whether the districts were expected to have schools using provision 2 or 3 (P23); in addition, we implicitly stratified on region, poverty, and SBP participation.[1] The original design called for a final sample of 100 primary sampling unit (PSU) equivalents. Because of resource limitations, the final sample target was reduced to 80 PSU equivalents. After the initial sample was selected and divided into main and replacement selections, we chose a subsample so that the main public school district sample consisted of 87 school districts with the expectation that with refusals, the participating sample would

[1]Provision 2 and Provision 3 are designed to reduce application burden and to simplify meal counting and claiming procedures. These provisions allow schools to operate under special application and meal counting rules. When schools use provisions 2 or 3, all students receive free meals without applying for meal benefits or being directly certified in a current school year. Under Provision 2, schools operate a "base year," in which they serve all meals at no charge but use standard program procedures to establish individual students' free or reduced-price meal eligibility and count meals by eligibility category. In future "non-base years," Provision 2 schools then may continue to serve all meals at no charge and take only a daily aggregate count of meals served for up to three additional years, during which they claim reimbursement based on the percentage of free, reduced-price, and paid meals served during the base year. Provision 3 schools serve all meals free for up to four years, and reimbursement is based on the total dollar reimbursement the school received during the "base year," which is the most recent year in which applications were taken and meals were counted and claimed by category. The reimbursement is adjusted each year for inflation and enrollment. Schools are most likely to find it in their financial interest to use Provision 2 or Provision 3 if they serve high-poverty populations and typically serve a large proportion of their meals free of charge.

include 80 public school district equivalents. The final sample consists of 80 public school district equivalents (78 unique public SFAs) and 9 private SFAs.

2. **266 schools, of which 256 are public and 10 are private.** Within each school district that was sampled and agreed to participate in the study, we selected a sample of schools. The number of schools selected from each district depended on whether the district represented more than one PSU equivalent and whether any schools in the district participated in P23. In study districts not using P23, we sampled six schools (three main selections and three replacements). We selected larger samples from P23 districts in order to obtain large enough samples of base-year and non–base year P23 schools. Schools were divided into school-level strata (elementary schools versus middle and high schools). The school sample included both public and private schools. Private schools were sampled from among those located within the boundaries of a sampled public school district, based on the ZIP code of the private school's location. We oversampled elementary schools because they comprise a greater share of free and reduced-price reimbursements. In addition, P23 schools were oversampled to support comparative analysis of P23 and non-P23 schools.

3. **6,776 students certified for free and reduced-priced meals, and 1,038 students who applied for and were denied benefits** (information about this sample of students was collected from their applications for free or reduced-price meal benefits).

4. **A subsample of 2,950 students certified for free and reduced-price meals and 453 denied applicants for whom we also conducted an in-person household survey.** We selected samples of students certified for free and reduced-price meals and students who applied but were denied meal benefits for the application record abstraction and household survey data collection. Students were sampled from lists provided by school districts (or sometimes from schools) participating in the study. We used two sets of records for sampling students: (1) lists of students certified for free or reduced-price meals and (2) lists of denied applicants (if no list was available, we sampled from the denied applications themselves). The household interview sample was a subsample of the record abstraction sample, resulting in a large sample of certified students and students denied meal benefits with application records and household survey data. We selected the samples of students certified for free or reduced-price meals throughout the year, selecting a majority early in the school year when most students apply and are certified. In contrast, we selected denied applicants only during the initial months of the school year.[2]

[2]For each study school that had a meal program that was either non-P23 or P23 base year, we selected the following for the household survey: (1) 20 approved free/reduced-price (F/RP) applicant students (10 main, 10 replacements) and (2) 4 denied applicants (2 main, 2 replacements) in the beginning of the school year. Then later in the school year, we selected 6 newly certified students (2 main, 4 replacements). Regarding the F/RP certified sample selected at the beginning of the school year, our target was to complete 9 to 10 F/RP household surveys. Field staff members were told to release the 10 main selections for the certified students, and if they encountered nonresponse, release up to 2 replacements, and then contact MPR field coordinators before going any deeper into the replacement sample. Similar procedures were followed for the denied applicants and newly certified students.

2. Data Sources

We collected data on these samples from several sources: surveys of households and SFA directors, administrative data from schools and districts, and observational data collected during visits to sampled schools. The data sources provided information that allowed us to measure both certification error and erroneous payments among individual students and non-certification error in the processes schools and districts use to claim reimbursements from state agencies.

Data collected on the subsample of certified students and denied applicants were used to construct estimates of erroneous payments for students across the full school year. The key sources of data collected on these students included the following:

- ***Household Surveys.*** Field interviewers completed in-person interviews with the parent or guardian of 2,950 students certified for free or reduced-price meal benefits and 453 students who applied for but were denied meal benefits. The survey collected information on household composition and size as well as detailed information on the sources of income of family members. It also collected data on participation of sampled children in the school food programs for a target week. Parents were asked to show interviewers pay stubs or other documentation to verify the sources of income and income amounts reported in the interview. These data were used along with data abstracted from the district's master eligibility roster to determine whether the students were certified improperly or improperly denied benefits to which they were eligible.
- ***Administrative Records Data Abstraction.*** We collected administrative records data from SFA directors on the samples of certified students and denied applicants that were used to estimate the number of meals received by misclassified students throughout the school year while in various certification status/eligibility status categories. These data included (1) students' enrollment start and stop dates and any changes in certification status (and dates of these changes) during the school year for 2,950 certified students and 453 denied applicants in the research sample and (2) students' monthly meal-program participation during the school year for those students attending schools that record and retain meal-program participation at the individual student level (for a total of 2,500 of the 3,403 certified students and denied applicants in the research sample).

(continued)

There was no replacement sample for the P23 application abstraction–only samples. At P23 schools that were base year, we selected an additional sample of 16 F/RP certified and 4 denied applicants (independently from the 20 F/RP and 4 denied applicants sampled for the household survey component). At P23 schools that were non–base year, we selected 16 certified and 4 denied applicants from the base year (not the current school year).

3. Estimation Methods and Results

The estimates of amounts and rates of erroneous payments due to certification error were based on all certified students (including directly certified students) and denied applicants. Certification error was determined by comparing sampled students' certification status as determined by the district with their actual free or reduced-price meal eligibility status. We determined students' certification status using data from the master eligibility lists provided by districts (free, reduced-price, paid). Students' free or reduced-price meal eligibility status was measured based on information collected during the in-person household survey on students' household income, household size, and receipt of Food Stamp Program (FSP), Temporary Assistance to Needy Families (TANF), or Food Distribution Program on Indian Reservation (FDPIR) benefits. This information reflected students' household circumstances at about the time the households submitted their applications for free or reduced-price meal benefits.

We estimated separate erroneous payments amounts and rates for the NSLP and SBP. These estimates are for NSLP/SBP participating schools in the 48 contiguous states and District of Columbia during SY 2005–2006 (the estimates exclude Alaska, Hawaii, the U.S. territories, schools operated by the Department of Defense (DOD), and Residential Child Care Institutions (RCCIs). We used a three-step procedure to derive national estimates of erroneous payments: (1) we estimated erroneous payments attributable to non-Provision 2 or 3 (NP23) and Provision 2 or 3 (P23) base-year schools (that is, excluding P23 schools in non-base years);[3] (2) we imputed estimates of erroneous payments for P23 schools in non-base years; and (3) we combined the

[3]We initially excluded non-base year P23 schools from the first step because these schools did not conduct a certification process in the data collection year and required a different estimation method. Overall, 4.5 percent of reimbursements for free or reduced-price NSLP lunches and 11.3 percent of reimbursements for free or reduced-price SBP breakfasts were at non-base year P23 schools.

two estimates into an overall estimate of erroneous payments covering all schools in the 48 contiguous states and District of Columbia for SY 2005-06.

In the case of NP23 schools and P23 base year schools, erroneous payments were determined by the certification and eligibility status of each student in the study who was certified for free or reduced-price meal benefits or applied for and was denied benefits during the study school year—that is, whether the student was certified in error or erroneously denied benefits and the number of meals he or she received over the course of the school year while incorrectly certified.[4] Table I.1 shows possible "per-meal" erroneous payment amounts for certification errors for students who participated in the NSLP; Table I.2 provides analogous information for the SBP.

We performed the following steps to determine the amounts and rates of erroneous payments in the NSLP for NP23 and P23 base-year schools:

For Each Sampled Student

- Determine the overpayment or underpayment for each school meal received by the student in a given month according to Table I.1.
- Multiply this estimate of per-meal erroneous payments by the number of school lunches received in the month to determine the total NSLP erroneous payments for that student in the month.
- Sum these totals across all months of the school year to determine the total erroneous payments for the student throughout the school year for lunches received through the NSLP.

For the Full Sample of Students

- Across all students in the sample, calculate the weighted sum of annual NSLP erroneous payments to determine total erroneous payments for students in the 48 contiguous states and District of Columbia.

[4]The sample also includes directly certified students and other students certified for free meals without submitting an application.

TABLE I.1

TOTAL UNDERPAYMENTS AND OVERPAYMENTS PER MEAL IN THE NSLP, SY 2005–2006
(in Dollars)

Student's Certification Status	Student's Eligibility Status	Total Payments[a]	Underpayments[b]	Overpayments[b]
Less than 60 Percent[c]				
Free	Free	2.5127	0.00	0.00
Free	Reduced-price	2.5127	0.00	0.40
Free	Paid	2.5127	0.00	2.10
Reduced-price	Free	2.1127	0.40	0.00
Reduced-price	Reduced-price	2.1127	0.00	0.00
Reduced-price	Paid	2.1127	0.00	1.70
Denied	Free	0.4127	2.10	0.00
Denied	Reduced-price	0.4127	1.70	0.00
Denied	Paid	0.4127	0.00	0.00
60 Percent or More[d]				
Free	Free	2.5327	0.00	0.00
Free	Reduced-price	2.5327	0.00	0.40
Free	Paid	2.5327	0.00	2.10
Reduced-price	Free	2.1327	0.40	0.00
Reduced-price	Reduced-price	2.1327	0.00	0.00
Reduced-price	Paid	2.1327	0.00	1.70
Denied	Free	0.4327	2.10	0.00
Denied	Reduced-price	0.4327	1.70	0.00
Denied	Paid	0.4327	0.00	0.00

Source: FNS program data.

[a]In the NSLP, the "paid" rate is established in Section 4 of the National School Lunch Act (NSLA); the Section 4 rate is paid for all lunches served. Section 11 of the NSLA establishes additional reimbursement ("special assistance payment") for lunches served to students certified eligible for free and reduced-price meals. The Section 11 payment is paid in addition to the Section 4 payment for those meals served to children certified eligible for free or reduced-price meals. Total reimbursement per lunch therefore equals cash reimbursement from Section 11 and Section 4 and the per-meal value of commodities. For example, for regular free meals (those not receiving the additional two cents subsidy for providing a high percentage of free and reduced-price lunches), the total amount reimbursed per free lunch equals $2.5127, and is comprised of $0.22 (Section 4 paid rate), $2.10 (Section 11 extra subsidy for free lunches), and $0.1927 (value of commodities per lunch).

[b]Erroneous payments under the NSLP refer to the reimbursement amount in error under Section 11. That is, erroneous payments only involve the extra subsidy for free or reduced-price lunches above the Section 4 paid rate.

[c]These reimbursement rates apply to school districts that claimed less than 60 percent of total lunches at the free and reduced-price rate in the second preceding school year.

[d]School districts that claimed 60 percent or more of total lunches at the free or reduced-price rate in the second preceding school year receive an extra two cents for each lunch claimed.

TABLE I.2

TOTAL UNDERPAYMENTS AND OVERPAYMENTS PER MEAL IN THE SBP, SY 2005–2006
(in Dollars)

Student's Certification Status	Student's Eligibility Status	Total Payments[a]	Underpayments[b]	Overpayments[b]
SBP, Non-Severe-Needs School[c]				
Free	Free	1.27	0.00	0.00
Free	Reduced-price	1.27	0.00	0.30
Free	Paid	1.27	0.00	1.04
Reduced-price	Free	0.97	0.30	0.00
Reduced-price	Reduced-price	0.97	0.00	0.00
Reduced-price	Paid	0.97	0.00	0.74
Denied	Free	0.23	1.04	0.00
Denied	Reduced-price	0.23	0.74	0.00
Denied	Paid	0.23	0.00	0.00
SBP, Severe-Needs School[d]				
Free	Free	1.51	0.00	0.00
Free	Reduced-price	1.51	0.00	0.30
Free	Paid	1.51	0.00	1.28
Reduced-price	Free	1.21	0.30	0.00
Reduced-price	Reduced-price	1.21	0.00	0.00
Reduced-price	Paid	1.21	0.00	0.98
Denied	Free	0.23	1.28	0.00
Denied	Reduced-price	0.23	0.98	0.00
Denied	Paid	0.23	0.00	0.00

Source: FNS program data

[a]In the SBP, payment rates for paid, reduced-price, and free meals are established in Section 4 of the Child Nutrition Act of 1966 (CNA). SBP breakfasts receive a cash subsidy only. The SBP does not receive commodities.

[b]For the SBP, erroneous payments refer to the difference between the reimbursement rate for paid meals and the rates for free and reduced-price meals (including the additional payments for severe-needs free and reduced-price meals, as appropriate). For example, the total amount reimbursed per free breakfast in a non-severe needs school equals $1.27, and is comprised totally of the Section 4 rate.

[c]These reimbursement rates apply to school districts that claimed less than 40 percent of their total lunches at the free and reduced-price rate in the second preceding school year.

[d]School districts that claimed 40 percent or more of total lunches at the free or reduced-price rate in the second preceding school year may receive extra severe-needs reimbursement of up 24 cents per meal for all free and reduced-price breakfasts claimed.

- Derive an estimate of total reimbursements for all NSLP lunches provided to students in the 48 contiguous states and District of Columbia.
- Divide total erroneous payments by total reimbursements to determine the rate of erroneous payments in the NSLP.

We used an analogous methodology to determine the separate rates of overpayments and underpayments. We followed the same procedures to estimate amounts and rates of SBP erroneous payments; the estimates for the SBP take into account whether the student attended a severe-needs versus a non-severe-needs school, as the extra subsidy for free and reduced-priced meals is different in the two types of schools.

In P23 schools not in their base year, there is no certification process during the school year, as reimbursements are determined largely by the results of the certification process conducted during the base year. Erroneous payments due to certification error in these schools in their non-base years are caused by errors made during the base year certification process. To determine a national measure of erroneous payments that would include the non-base year P23 schools, we imputed the rates of erroneous payments in P23 non-base year schools. The imputation was based on rates of erroneous payments in P23 base year schools for which we did collect certification and eligibility data and which matched the non-base year schools in some important respects.[5]

Table I.3 presents the full set of erroneous payments estimates for SY 2005-06. For both the NSLP and SBP, approximately 9 percent of total reimbursements were erroneous due to certification errors. There were an estimated $759 million in erroneous NSLP reimbursements due to certification error, or 9.4 percent of the roughly $8.06 billion in cash reimbursements and

[5]See Ponza et al. 2007a and 2007b for detailed discussion of methods used to estimate erroneous payments for SY 2005 – 06.

TABLE I.3

ERRONEOUS PAYMENTS DUE TO CERTIFICATION ERROR IN THE NSLP AND SBP, SY 2005–06

	NSLP	SBP
Total Reimbursements (millions of dollars)	8,060	1,938
Overpayments (millions of dollars)	573	137
	(50)	(17)
Underpayments (millions of dollars)	186	40
	(15)	(6)
Total Erroneous Payments (millions of dollars)	759	177
	(54)	(18)
Erroneous Payments as Percentage of All Reimbursements		
Overpayments	7.11	7.07
	(0.62)	(0.91)
Underpayments	2.31	2.08
	(0.19)	(0.29)
Total erroneous payments	9.42	9.15
	(0.67)	(0.94)

Source: APEC study, weighted data.

Note: Standard errors in parentheses.

The estimates include erroneous payments at all schools participating in the NSLP and/or SBP in the contiguous U.S. and District of Columbia, including provision 2 or 3 non-base year schools. They are based on all students who applied for free or reduced-price meals (including denied applicants) and directly certified students. For the NSLP, Section 11 of the NSLA establishes reimbursement above the Section 4 paid rate. Erroneous payments under the NSLP refer to the reimbursement amounts in error under Section 11 of the NSLP. For the SBP, erroneous payments refer to the difference between the reimbursement rate for paid meals and the rates for free and reduced-price meals (including the additional payments for severe-needs free and reduced-price meals, as appropriate). The denominator in the erroneous payment rate calculation refers to reimbursements for all meals (free, reduced-price, and paid). For the NSLP, total reimbursements equal total cash reimbursement from Section 11 and Section 4 and the value of commodities (valued at a per-meal rate). In the SBP, payment rates for paid, reduced-price, and free meals are established in Section 4 of the CNA. SBP breakfasts receive a cash subsidy only. The SBP does not receive commodities. Total reimbursements for the SBP therefore equal total cash reimbursement from Section 4.

commodities provided to school districts for all NSLP lunches served in the contiguous United States.[6] Erroneous SBP reimbursements totaled $177 million, or 9.1 percent of the $1.94 billion in cash reimbursements paid for all SBP breakfasts served.[7] Within total payments due to certification error, overpayments were much more common than underpayments. More than three-quarters of erroneous payments due to certification error in both the NSLP and SBP were overpayments. The estimated overpayment rate was 7.1 percent and the underpayment rate was 2.3 percent for the NSLP. Similarly, the estimated overpayment rate was 7.1 percent for the SBP and the underpayment rate was 2.1 percent.

B. OVERVIEW OF APPROACH FOR UPDATING ESTIMATES OF ERRONEOUS PAYMENTS IN THE NSLP AND THE SBP

The estimates of amounts and rates of erroneous payments due to certification error for SY 2005 – 06 described in the previous section are based on nationally-representative primary data collected from the APEC study. Because improper payments made in the NSLP and the SBP during SY 2005–06 are significant, FNS will need to report annually on the amount of estimated improper payments in the NSLP and the SBP. The overall strategy for updating the national estimates of erroneous payments due to certification error in future years involves the following steps. We first disaggregated overall findings on the SY 2005–06 error rates into a set of district-level measures of error rates for the nationally representative sample of just under 90 districts

[6]The $8.05 billion refers to total cash reimbursements (Section 4 and Section 11 payments under the NSLA) and value of commodities for all reimbursable NSLP lunches provided to students attending schools in the contiguous United States during FY 2006 (including Provision 2 or 3 schools in non-base years). It excludes Alaska, Hawaii, the U.S. territories, and schools operated by the Department of Defense as well as Residential Child Care Institutions (RCCIs).

[7]The $1.95 billion refers to total cash reimbursements (Section 4 payments under the CNA) for all reimbursable SBP breakfasts provided to students attending schools in the contiguous United States during FY 2006 (including Provision 2 or 3 schools in non-base years). It excludes Alaska, Hawaii, the U.S. territories, and schools operated by the Department of Defense as well as RCCIs.

selected for the APEC study. We then estimated a series of regression models capturing the relationship between the characteristics of these districts and their estimated rates of certification error. In future years, estimated coefficients from these models can be used in conjunction with updated values of district characteristics to predict certification error in these years in any district for which data on the district characteristics are available. These certification error rates can then be translated into amounts and rates of erroneous payments in each district. By doing this for a national set of districts and then aggregating the district-level estimates, a national measure of predicted erroneous payments can be calculated for future years.

C. ORGANIZATION OF REPORT

The remainder of the report is organized as follows. The first part of Chapter II describes the data used in the analysis, both for the measures of certification error used as dependent variables in the econometric models and for the district characteristics that serve as explanatory variables in these models. The chapter then goes on to describe the two parts of our methodology: (1) estimating the econometric models and (2) using the results of these models to predict future erroneous payments. The results of the estimation of the econometric model are presented in Chapter III. We illustrate how to calculate predicted erroneous payments by using 2005–2006 data on district characteristics for all districts nationally to generate a predicted level of erroneous payments in this year. We compare this value that is predicted based on coefficients from the econometric model and district characteristics—but without using information collected as part of the APEC study on actually erroneous payments—to the sample-based estimates from the APEC study shown in Table I.3. Chapter III also discusses the strengths and limitations of the model.

II. DATA AND METHODS

The effort to predict future amounts and rates of erroneous payments in the school meal programs involves two major steps. The first is to estimate an econometric model of the relationship between a district's characteristics and its school meal program certification error rates. This econometric model is estimated using data from the Access, Participation, Eligibility, and Certification (APEC) study on the nationally representative sample of districts used in that study. The second step involves using the results of that econometric model along with data on the characteristics of a national set of school districts to predict future erroneous payments. This chapter describes the sources of data used in this empirical work and presents the details of both the econometric model and the process for estimating future erroneous payments.

A. DATA SOURCES AND MEASURES

The data required for the econometric model include the model's dependent variables (district-level measures of certification error rates) and independent variables (a set of district characteristics).

The error rate measures used as dependent variables in the econometric model are based on primary data collected in the APEC study on a nationally representative sample of certified students and denied applicants. These data sources included household survey data and administrative data from school districts on students' meal program participation and changes in eligibility and enrollment statuses during the school year.

The independent variables we considered for the model include indicators of the administrative features of the NSLP and the SBP in the district, other characteristics of the district, and demographic characteristics of students and families in the districts. The district's verification outcomes were included as key explanatory variables, since they are highly

predictive of error rates in the district. Although we needed data on all these measures only for the districts included in the APEC sample for purposes of estimating the econometric model, the data used to create the model's independent variables also had to be available for all districts nationally in the sample year and in future years in order to use the model to predict future erroneous payments. The independent variables we considered were constructed from data from several sources: (1) district-level administrative data on district characteristics and verification outcomes from the School Food Authorities (SFA) Verification Summary Reports (Form FNS-742), (2) district-level administrative data on the number of meals by type from the FNS state agencies, (3) public school district–level characteristics data from Common Core Data (CDD) and the Decennial Census, (4) private school–level data from the Private School Survey (PSS), (5) county-level data on unemployment rates from the U.S. Department of Labor's Local Area Unemployment Statistics (LAUS), and (6) APEC study data on direct certification methods and other SFA characteristics.

In the remainder of this section, we describe each of these data sources and the relevant data items we used from each source.

1. Dependent Variables: Misclassification Error Measures Based on Survey and Administrative Data on Sampled Students in the APEC Study

The models use as their dependent variables district-level rates of 12 possible categories of misclassification error for SY 2005–06. In each case, the measure describes the proportion of meals in a particular meal-price category that had a particular type of certification error. To create these measures, we aggregated meal-level estimates of certification error across all meals served to sampled students in each of the districts in the APEC study national sample.

Definitions of Misclassification Error Measures. We created the following 12 district-level error rate measures for the NSLP and the SBP:

1. ***%CF-RPE-L.*** The percentage of all free school lunches in the district that should have been classified as reduced-price (certified free, reduced-price eligible).
2. ***%CF-PE-L.*** The percentage of all free school lunches in the district that should have been classified as paid (certified free, paid eligible).
3. ***%CRP-PE-L.*** The percentage of all reduced-price school lunches in the district that should have been classified as paid (certified reduced price, paid eligible).
4. ***%CRP-FE-L.*** The percentage of all reduced-price school lunches in the district that should have been classified as free (certified reduced price, free eligible).
5. ***%NC-RPE-L.*** The percentage of all paid school lunches (that is, served to students not certified for free or reduced-price meals) in the district that should have been classified as reduced price (not certified, reduced price eligible).
6. ***%NC-FE-L.*** The percentage of all paid school lunches (that is, served to students not certified for free or reduced-price meals) in the district that should have been classified as free (not certified, free eligible).
7. ***%CF-RPE-B.*** The percentage of all free school breakfasts in the district that should have been classified as reduced price (certified free, reduced-price eligible).
8. ***%CF-PE-B.*** The percentage of all free school breakfasts in the district that should have been classified as paid (certified free, paid eligible).
9. ***%CRP-PE-B.*** The percentage of all reduced-price school breakfasts in the district that should have been classified as paid (certified reduced price, paid eligible).
10. ***%CRP-FE-B.*** The percentage of all reduced-price school breakfasts in the district that should have been classified as free (certified reduced price, free eligible).
11. ***%NC-RPE-B.*** The percentage of all paid school breakfasts (that is, served to students not certified for free or reduced-price meals) in the district that should have been classified as reduced price (not certified, reduced-price eligible).
12. ***%NC-FE-B.*** The percentage of all paid school breakfasts (that is, served to students not certified for free or reduced-price meals) in the district that should have been classified as free (not certified, free eligible).

Methods Used to Construct Error Rate Measures. The error rate variables were created using data from the APEC study student-level data set (see Section I.B.2). Each district-level error rate measure represents the proportion of meals in a particular meal-price category that had

a particular type of certification error. To create these measures, we first determined the eligibility status of each of the students in our sample of certified students and denied applicants within each district. We determined students' free or reduced-price eligibility status based primarily on information collected during the in-person household survey on students' household income, household size, and receipt of FSP/TANF/FDPIR benefits.[1] This information reflected students' household circumstances at about the time the households submitted applications for free or reduced-price meals. For students who became certified without submitting an application (for example, directly certified students), the information collected on the household survey reflected household circumstances at the beginning of the school year. We classified these students as certified accurately regardless of the information in the household survey as long as the required documentation for the type of certification (that is, documentation from the FSP/TANF/FDPIR administering agency) was available from the district. Students were classified by MPR as eligible for free meals at the time their application was certified (or the beginning of the school year if they became certified without an application) if they met any of the following conditions:

[1]It is well known that measures of income based on survey responses are typically measured with error. The APEC study implemented several approaches that were expected to minimize problems associated with measuring income from a household survey that should result in more accurate household income measures, but does not eliminate all measurement error. For example, the study used computerized-assisted personal interviewing, requested documentation of income sources and amounts and reconciled reported sources and amounts against documented ones; the reference month for the interview coincided with the month covered by the application and the interview was conducted no more than two months later (for more discussion, see pages 40 – 41 (Ponza et al., 2007a). To get a sense of the effects that measurement error might have on results, for the main study we did examine the sensitivity of certification error estimates to changes in the meal benefit eligibility thresholds. We found that increasing the thresholds for free and reduced-price eligibility both by 5 percentage points reduced certification error by 6.5 percent; increasing the thresholds by 10 percentage points, reduced the certification error rate by 15 percent. Ultimately, there is not much we can do regarding measurement error in survey reports of income sources and amounts but acknowledge that the potential for error exists, since the best we can do with the modeling effort is to replicate the APEC study estimates (which are imperfect themselves).

- Their household income was less than or equal to 130 percent of the federal poverty level
- They were receiving FSP, TANF, or FDPIR benefits
- They had been directly certified correctly or otherwise certified correctly for free meals without an application[2]

We classified students as eligible for reduced-price meals if they were not eligible for free meals but their income was less than or equal to 185 percent of the federal poverty level. An additional eligibility requirement for either free or reduced-price meals was that for students certified by application, the district could locate the application in their files. If the district did not have an application on file, the student was classified as not eligible for free or reduced-price meals, as specified in FNS rules.

Next, we determined if the student was erroneously certified or inappropriately denied benefits in which he or she was eligible. This determination was made by comparing the student's certification status on the district's master eligibility list to our independent assessment of the student's eligibility status based on information provided on the household survey. Certification error occurs when these two measures of status differ.[3] Table II.1 summarizes the possible combinations of eligibility and certification status among students who have applied for meal benefits or have been directly certified. In the table, the columns indicate students' certification status (free, reduced-price, or denied) and the rows indicate the level of benefits for

[2]In addition to directly certified students, this also includes other categories of students certified for free meals without having to submit an application, such as homeless children, runaway children, children of migrant workers, and students displaced by Hurricane Katrina (or other natural disasters).

[3]If a district does not have an application or direct certification document on file for a certified student, then according to FNS rules, the student should not have been certified, even if the student was correctly certified (certification status matches eligibility status). We treat this as a certification error. However, if an application is found that does not include the required signature or Social Security Number (SSN not required for categorically eligible applications), then the application is considered an administrative error; but if the certification status is correct based on the household circumstances reported on the household survey, then it is not considered a certification error.

TABLE II.1

POSSIBLE COMBINATIONS OF STUDENT'S CERTIFICATION AND ELIGIBILITY STATUS

	Certification Status		
Eligibility Status	Free	Reduced-Price	Denied
Free	A	B	C
Reduced-Price	D	E	F
Paid	G	H	I

which the students are eligible (free, reduced-price, or paid). An example of certification error would be if our assessment, based on information in the household survey, was that the student should have been certified for reduced-price meal benefits, but the student was recorded as certified for free meals on the master eligibility list (Cell D). Certification error can come from household reporting error, district administrative error, or a combination of both types of error. Students in cells along the diagonal (A, E, and I) were certified accurately or were appropriately denied benefits. Students in cells below the diagonal (D, G, and H) were overcertified and students in cells above the diagonal (B, C, and F) were undercertified.

Once we determined whether a student was erroneously certified or inappropriately denied benefits, we were able to construct the district-level error rate measures. To construct the first four of six error rate measures for the NSLP and the SBP, we first restricted the sample to those students enrolled in a sampled district in the appropriate certification category. For example, we based the first dependent variable, ***%CF-RPE-L***, on only the sample members certified for free meals.

Within each district, we then aggregated the number of school meals provided to all students in that certification category as well as the number of meals provided to students in that category

with a particular type of certification error.[4] For the first dependent variable, for example, we calculated the number of school lunches provided to students certified for free meals who were eligible for reduced-price meals (Cell D). We estimated the error rate by dividing the number of free lunches with that particular error by the total number of lunches provided to students in that certification category.

Calculating the final two dependent variables in each set required an additional step, since the denominator in each of these rates includes school meals provided to all non-certified students in the district, but the APEC sample included only non-certified students who had applied for but been denied free or reduced-price meal benefits (that is, denied applicants).[5] Thus, we initially calculated the proportion of meals served to denied applicants that were in a particular error category, but then we adjusted this figure by multiplying it by the ratio of the number of paid meals provided to denied applicants to the number of paid meals provided to all non-certified students in the district. We obtained information on the numerator (the number of meals served to denied applicants) by using information on meals consumed by the denied applicants in our sample. Information on the denominator (the number of meals served to all non-certified students) came from the SFA survey.

[4]In aggregating meals provided to a particular group of sample members in a district, we used the appropriate sample weights so that the weighted sum reflected the estimated number of meals provided to all enrolled students in that particular group within the district.

[5]It is more straightforward to calculate the percentages of meals served to denied applicants that should have been classified as free or reduced price. However, these rates can not be used as dependent variables in the model because the number of paid meals served to denied applicants at all schools nationally is not available. But the number of paid meals served to all non-certified students nationally is available, so we constructed the dependent variables in terms of meals served to all non-certified students as opposed to meals served to denied applicants.

2. Independent Variables

The independent variables we considered were constructed from data from several sources, described in the remainder of this section.

a. Measures Constructed from SFA Verification Summary Reports (Form FNS-742)

Beginning in SY 2004–05, FNS has required school districts to report yearly results of verification and other data on characteristics of the district to their state agencies (see Office of Analysis, Nutrition, and Evaluation, Food and Nutrition Service, U.S. Department of Agriculture, November 2005). State agencies then submit the data for their school districts to FNS central office in a single consolidated electronic file. We obtained FNS-742 data for SY 2005–06 from FNS for all the SFAs in the country that complied with the reporting requirements.

The FNS-742 data on SFA characteristics and verification outcomes for the APEC study's SFA sample were used to develop the econometric model. In addition, these data were used to form the group of districts needed to predict erroneous payments nationally (and to be used in future years to predict erroneous payments). Ideally, this group of districts would include every public or private SFA in the country that offers one of the USDA school meal programs. The FNS-742 data set comes close to this goal, since the great majority of such districts were required to complete the FNS-742 form and submit information that would be included in the FNS-742 data set. In reality, however, the data set excludes those districts that failed to comply with these reporting requirements. The 17,115 SFAs in the SY 2005–06 FNS-742 data file represent approximately 95 percent of the child nutrition state agencies and 82 percent of the SFAs. The number of students enrolled in these SFAs equaled 45,908,055, representing

approximately 92 percent of the total enrolled in schools operating the NSLP and/or the SBP nationwide that year.[6]

We eliminated districts that did not fall within our study population (all public and private SFAs in the contiguous 48 states and the District of Columbia). Therefore, we removed (1) districts in Alaska, Hawaii, and U.S. territories; (2) schools operated by the Department of Defense; and (3) single-entity residential child care institutions (RCCIs) that do not have day students.[7] This resulted in a final edited data file of 17,038 SFAs.[8]

---

[6]Some districts are not included in the FNS-742 data used for future predictions. Some are appropriately excluded (that is, districts which do not claim any meals for reimbursement at the free or reduced-price rate are not required to complete the FNS-742), while other districts failed to submit reports when they should have. We examined whether excluded districts differed from those included in FNS-742. We investigated the numbers and percentages of free-certified, reduced-price certified, and non-certified students in the population and in the FNS-742. The percentage of students in each category is very similar in the population and in the FNS-742 data. For example, nationally 33.7 percent of students in schools offering the NSLP and/or SBP are certified for free meals, whereas the estimate is 33.4 based on districts included in the FNS-742 data. We then calculated the percentage of each of these types of students that are in the FNS-742 datafile: it is about 85 percent for each category. We then took out the numbers of students in each certification category for the districts that are not included in the FNS-742 data. The percentage of students in each category is very similar to those in the FNS-742. On this important measure, then, the districts that are excluded from the FNS-742 datafile appear similar to those that are included. More generally, we believe that the missing data problem in FNS-742 is not a major issue for three reasons: (1) the predicted amounts and rates of erroneous payments for the study year using the districts in FNS-742 are quite close to those from the APEC study for SY 2005-06, which is based on nationally representative data; (2) The SFAs in the SY 2005 – 06 FNS-742 datafile represent high percentages of the child nutrition state agencies, SFAs, and the total enrolled students in schools operating the NSLP and/or SBP nationwide that year; and (3) we expect overtime that district response to FNS-742 will improve, as districts become more familiar with the reporting process and with improvements in the reporting process (e.g., online reporting).

[7]Because the APEC study sample of schools did not include RCCIs, we wanted to eliminate them from the FNS-742 data file. However, it is impossible to eliminate all RCCIs because of reporting conventions used in the FNS-742 summary reports. The instructions for the FNS-742 form tell states not to include on the report RCCIs with no day students, since these institutions are exempt from verification. However, most states have added this reporting into their automated information systems, which include all SFAs, and the report they generate probably includes these RCCIs. Single-entity RCCIs with no day students could be identified (enrollment equals zero) and were eliminated from the data file. (In addition, we dropped three additional districts that had missing enrollments and appeared to be single-entity no day student RCCIs.) If an RCCI (particularly one that is part of a public school district with regular schools) has day students who are subject to verification, it is included on the FNS-742 report, since the FNS does not make any distinction between this type of RCCI and regular schools. We were unable to eliminate these RCCIs from the data file.

[8]Eight SFAs in the APEC study's SFA sample of 87 districts did not submit FNS-742 summary reports in SY 2005–06. However, because most of the data elements in the FNS-742 data file were also included in the APEC study's SFA survey, and these eight SFAs completed the survey, we were able to use the data from the SFA survey and include them in the final edited data file.

Constructed Variables. The data from FNS-742 verification summary reports were used to create the following variables:

- Type of SFA (public or private)
- Type of application used (individual student, household, or both)
- Number of schools operating the NSLP and/or the SBP
- Number of P23 non–base-year schools
- Number of enrolled students with access to the NSLP and/or the SBP
- Number of students in P23 non–base-year schools
- Percentage of students certified for free meals, by certification method:
  - percentage certified for free meals, not subject to verification[9]
  - percentage certified for free meals based on income/household size information submitted on applications
  - percentage certified for free meals based on categorical eligibility from information on their applications
  - percentage of students certified as eligible for free meals in P23 non–base-year schools
- Percentage of students certified for reduced-price meals:
  - percentage of students certified for reduced-price meals
  - percentage of students certified as eligible for reduced-price meals in P23 non–base-year schools
- Number of applications selected for verification
- Percentage of total applications selected for verification
- Verification sampling method (focused/error-prone, random, or all applications sampled)
- Percentage of applications verified by certification category (free-categorical approved, free-income approved, and reduced-price–income approved)

[9]Form FNS-742 does not specifically collect information on the number of students directly certified, a group of specific interest to the FNS. It does, however, collect information on the number of students not subject to verification (NSV), a group that includes directly certified students as well as income-eligible Head Start students, pre-kindergarten Even Start students, and other groups of students not subject to verification requirements, such as homeless children, children of migrant workers, and runaways.

- Verification results for each certification category:
  - Percentage of verified applications in which household responded, no change
  - Percentage of verified applications in which household responded, benefits reduced or terminated
  - Percentage of verified applications in which household responded, benefits increased
  - Percentage of verified applications in which household did not respond, benefits terminated
  - Percentage of verified applications for which outcome is missing

Data Edits and Imputations. We first examined the raw variables in the FNS-742 data file for missing data and recording errors (inappropriate codes, extreme values) and imputed values when data were missing on key items. We repeated the process for constructed variables, examining unusually low or high values, and imputed values when data were missing. To summarize:

- ***Resolving Cases with Missing Data***. Some districts left some key data elements blank. Generally, missing data are not prevalent for most variables in the FNS-742 data file (usually less than 0.5 percent). However, for some variables, missing data occurred for as many as 5 percent of SFAs in the data file. We did the following in cases with missing data:
  - *Obtained the Information from Another Source.* For some variables, we looked up the entity on the U.S. Department of Education's National Center for Education Statistics (NCES) website, used information from the Common Core of Data, or looked at the district's website to determine what the missing element should be.
  - *Performed Miscellaneous Imputations.* When it was not possible to obtain the missing information from other sources, we used an imputation approach tailored for the situation. If possible, we used other available information for that district to make a reasonable inference about the value of the missing variable. If this was not possible, we imputed the missing value with the variable's mean in these districts with the available values. Sometimes a combination of these two approaches was used.
- ***Checked Low and High Values for Reasonableness.*** For example, districts with low numbers or percentages of free or reduced-price certified students often included small private schools attended by largely non-poor students and were not altered. Extreme outliers (e.g., percentages exceeding 100 percent) were either set equal to 100 percent or assigned mean value replacement as judged to be appropriate.

b. Measures Constructed from Administrative Data from State Agencies on School Meals

We requested information on the number of meals by type (free, reduced price, paid, and total) separately for the NSLP and the SBP by public school district and private schools from state agencies administering the school meal programs for SY 2005–06. State agencies were asked to exclude meals from RCCIs and other institutions that were not schools and to exclude other school nutrition programs such as the milk or snack programs. We also asked state agencies to distinguish meals in the breakfast program by whether they were served in severe-needs schools, since these meals receive different levels of reimbursement.[10]

For developing the model specification, these data were used to create the following independent variables:[11]

- Whether district offers SBP
- Percentage of NSLP meals by type (free, reduced price, and paid)
- Percentage of SBP meals by type—severe and non–severe needs *not* distinguished
- Percentage of SBP meals by type—severe and non–severe needs distinguished
- Percentage of total breakfasts that are severe needs
- Percentage of total meals that are SBP breakfasts and percentage of total meals that are NSLP lunches
- Percentage of total meals by meal program and type

[10]Schools are defined as "severe need" if they claimed more than 40 percent of their lunches at the free and reduced-price rate in the preceding school year. "Severe need" schools receive additional reimbursement for each free and reduced-price breakfast claimed ($0.24 per meal in SY 2005-06).

[11]We had also planned to use these data in the final stage of the prediction process in which the model is used to obtain national estimates of erroneous payments. Under that approach, for each district, we would multiply the predicted error rate in each meal category obtained from the econometric model for our national set of districts by the total number of meals reimbursed as free, reduced price, or paid using data on the number of meals by type collected from state agencies. We would then multiply the estimated number of total meals erroneously reimbursed in each error category by the dollar value of the erroneous payment per meal in each error category. The result of this computation would produce an estimate of the total erroneous payments in each category for each district. These would then be summed across districts to obtain the national estimates of erroneous payments. This second use of the state agency data was considered but ultimately dropped as we decided to use an alternative approach that instead relies on more easily obtainable data from the FNS-742 and the FNS national data files.

Data Edits and Imputations. We received the data in various formats and modes, usually in Excel spreadsheets but sometimes hard copy that was then scanned and converted to SAS. States sometimes provided data at the school level within the districts, so these were aggregated to the district level.

Thirty states did not distinguish severe-needs versus non–severe-needs breakfasts. For those states, we used administrative data from the FNS national data file to calculate the percentage of breakfasts by type by severe-needs status and applied those state-specific adjustment factors to district breakfast totals to derive the number of breakfasts by type and severe-needs status for individual districts within states.

For some districts, the states did not provide the relevant SFA-level breakdown of the data. In those cases, we used state-level estimates of the numbers of meals reimbursed in the different categories (from the FNS national data file), along with information on the distribution of the number of certified students (in each category) across each district in the state (from FNS-742 data), to estimate the number of meals served in each category at each district.

We compared sums of meals by type with meal totals separately for the NSLP and the SBP by district. These comparisons identified errors (e.g., paid meals entered in total meals) that were corrected. Low and high values of individual variables were checked for reasonableness. Extreme outliers (e.g., percentages exceeding 100 percent) were either set equal to 100 percent or assigned mean value replacement as judged to be appropriate.

Comparisons of the state totals from this data file with state totals from the FNS national data file identified other problems. Because this data file excluded RCCIs but the FNS national administrative data includes RCCIs, we expected lunch and breakfast totals from state agency data we collected to be about 95 percent of the state totals in the national database. The majority of states achieved that threshold. However, there was a set of states in which the state's number

of meals (summed across the state's districts) was well below the FNS total. Cases in which the state total was less than 95 percent of the FNS total included Arkansas (87 percent), Connecticut (86 percent), Delaware (94 percent), Louisiana (86 percent), Maryland (84 percent), Texas (91 percent), Utah (75 percent), and Washington state (80 percent). In some of these states, all the districts were included but the number of meals was missing or appeared low for several districts. When data were missing, we assigned mean value replacement. When all the districts in a state were included but numbers of meals were low, we adjusted district totals by type upward by a constant factor so that when they were summed, they would equal 95 percent of the state totals in national data.

c. Measures Constructed from Common Core of Data

The Common Core of Data (CCD) is the U.S. Department of Education's primary statistical database of public elementary and secondary schools and districts. The data set, updated annually through surveys sent to state education agencies, contains demographic and administrative information on all public schools and districts in the United States. Information from the merged CCD files (Local Education Agency Universe Survey, Local Education Agency Finance Survey, Public School Universe Survey) was used to construct the following variables, which were considered as independent variables:[12]

- Total enrollment
- Number of schools
- Enrollment by race/ethnicity/gender/grade
- Grade span of district

[12]Enrollment by race, ethnicity, gender, and grade; percentage certified for free and reduced-price lunch; and Title 1 status are reported at the school level in the CCD so were aggregated to the district level.

- Location of district (for example, large city, mid-size city, large town, small town)
- Percentage certified for free and reduced-price meals
- Percentage of schools that are Title 1–eligible
- Number of SFA administrators and support staff
- Number of teachers, school administrators, and support services staff
- Spending on food services, food service salaries, and administrative support services

Edits and Imputation. We examined the variables constructed from the data file for missing data and recording errors (e.g., inappropriate codes or extreme values) and imputed values when data on key items were missing. In most cases, missing values were assigned with mean value replacement. Because the CCD is collected only for public schools and districts, private SFAs originally had missing values for all CCD-based measures. When data such as enrollment, grade span, and race/gender composition were available, these missing values were filled based on information from the Private School Survey, described next.

d. Private School Survey

The Private School Survey (PSS) is a national data set of private schools collected by the NCES. It includes information on religious orientation, level of school, total enrollment, and enrollment by gender. We used the PSS as a source of information about private schools that participate in the NSLP or the SBP. We also linked each participating private school to the public school district in which it is located to obtain relevant public school district–level information (such as district-level income and poverty data from the census, discussed in the next section). The included PSS variables had no missing or extreme values for the nine private APEC study districts. Thus, no missing value imputations or data edits were necessary.

e. The Decennial Census and Small Area Income and Poverty Estimates

Both a district's median income and poverty rate may be important predictors of erroneous payments. Poorer districts may have fewer resources to devote to certification procedures. In additional, poorer families may be more or less likely than wealthier families to report erroneous information on their applications. To measure income and poverty rates, we used annual estimates of county-level income and poverty rates from the Census Bureau's Small Area Income and Poverty Estimates (SAIPE). The SAIPE uses both Current Population Survey (CPS) and Decennial Census data to estimate district-level income and poverty rates in non-census years. A limitation of the SAIPE is that it is less reliable than census data, because it is based on projections rather than on direct estimates from the very large census samples. As noted earlier, private SFAs were linked to the public school districts in which they are located. Because we were able to match all sample SFAs to the census county and SAIPE data, no missing value imputations were necessary.

f. Local Area Unemployment Statistics (LAUS)

Erroneous payment rates in a district may also be correlated with its unemployment rate, which, like poverty rate and income, reflects the resources available to the district to determine certification status as well as the financial circumstances of applicants. The LAUS data provide monthly estimates of unemployment rates at the county level, which can be linked to public school districts. The Bureau of Labor Statistics produces the estimates in conjunction with state employment security agencies. The estimates for counties are based on a variety of data sources, including the CPS, Current Employment Statistics, the Decennial Census, and state unemployment insurance systems and are updated each month.

g. Additional Data Items Which Could be Used if Collected by FNS

In addition to the data items already discussed, other administrative data, not currently collected by FNS, could enhance estimates of erroneous payments in future years. In particular, information on the number of students directly certified, as well as additional information reflecting the districts' administration of school meal programs, may be predictive of certification error. The variables listed here were constructed using data from the APEC study SFA survey:

- Whether the district uses direct certification
- Type of direct certification approach used (no match, district-level match, state-level match, other method)
- Whether district uses a food management company to run its meal programs
- Whether the district uses computer software to process and determine eligibility for school meal benefits
- Whether the district tracks individual student participation in the meal program

B. METHODS

This section first describes the econometric model used to estimate the relationship between district characteristics and rates of certification error using data from the APEC study on the nationally representative sample of districts used in that study. We then describe the approach to using the estimates from this model along with data on the characteristics of a national sample of districts to predict erroneous payment amounts and rates resulting from certification errors in future years.

1. Econometric Model

The overall econometric model consists of a series of 12 equations estimated using district-level data. Each equation has a different dependent variable representing a particular rate of

certification error in the district, with 6 equations reflecting certification error rates among NSLP meals served and 6 equations reflecting certification error among SBP meals. The independent variables are characteristics of the district, which may differ from equation to equation.

The basic form of the model is:

(1) $E_{kj} = X_{kj}\beta_k + u_{kj}$, where k = 1, ... , 12

with E_{kj} = rate of certification error type k in district j

X_{kj} = characteristic of district j included in equation k

β_k = relationship between characteristic X_{kj} and error rate E_{kj}

u_{kj} = term representing unobserved effects on the error rate E_{kj}

We estimated the six NSLP and six SBP equations using ordinary least squares (OLS) estimation techniques.[13] We tested the sensitivity of the results to various alternative specifications of the model. For example, we assessed the assumption of a linear specification by estimating a Tobit model.[14] The Tobit model accounts for the fact that the dependent variable (certification error rates) cannot take on values less than zero or greater than one. The OLS estimation potentially yields estimates that are biased if a large proportion of the values of the dependent value are equal to zero or one. We also conducted various other tests of the sensitivity of the results to changes in the model specification. In particular, we examined interactions between independent variables and nonlinear functions of key independent variables. Results of selected specification tests are described in Chapter III.

[13]We also estimated a version of the model using weighted least squares to account for the multistage, clustered sample design. Results from this specification are similar to those presented here.

[14]A drawback of the Tobit model is that it makes strong assumptions about the distribution of the regression error term. We use this specification as a robustness check only.

Dependent Variables. The dependent variables of the 12 equations that make up the econometric model—half of which reflect overpayments and half underpayments—were described earlier but are repeated here for convenience:[15]

%CF-RPE-L: Percentage of lunches certified free, reduced-price eligible

%CF-PE-L: Percentage of lunches certified free, paid eligible

%CRP-PE-L: Percentage of lunches certified reduced price, paid eligible

%CRP-FE-L: Percentage of lunches certified reduced price, free eligible

%NC-RPE-L: Percentage of lunches not certified, reduced-price eligible

%NC-FE-L: Percentage of lunches not certified, free eligible

%CF-RPE-B: Percentage of breakfasts certified free, reduced-price eligible

%CF-PE-B: Percentage of breakfasts certified free, paid eligible

%CRP-PE-B: Percentage of breakfasts certified reduced price, paid eligible

%CRP-FE-B: Percentage of breakfasts certified reduced price, free eligible

%NC-RPE-B: Percentage of breakfasts not certified, reduced-price eligible

%NC-FE-B: Percentage of breakfasts not certified, free eligible

Two additional aspects of the dependent variables are worth noting. First, each dependent variable is an estimate, based on samples of students within the APEC study districts. Thus, each dependent variable is subject to measurement error (resulting from the sampling error of these estimates). Further, since the size of the APEC study samples varies somewhat from district to district for a given certification category, the underlying variability in this measurement error will differ from district to district. In other words, the variance of the

[15]We considered an alternative specification with fewer dependent variables that measured overall rates of erroneous payments, distinguishing only between overpayments and underpayments. We settled on this more disaggregated set of dependent variables because the independent variables included in the model were better able to explain variation in erroneous payments when the dependent variables were disaggregated according to the type of error.

disturbance term in Equation (1) will vary from observation to observation, a condition known as heteroskedasticity. We adjusted for this heteroskedasticity by estimating robust standard errors of the coefficients in the regression equations.

Second, the values of the dependent variables reflect the sampled districts' certification error rates only in non–provision 2 or 3 (non-P23) schools or base-year P23 schools. There is no certification process at non–base-year P23 schools—that is, students do not apply for free or reduced-price meals during the school year. As a result, the APEC study student sample did not include students at these schools, and we could not directly estimate certification error rates for these students.[16] Thus, the method for predicting national amounts and rates of erroneous payments involves first using this econometric model to predict these values in non-P23 and base-year P23 schools, and then adjusting these values to account for erroneous payments at non–base-year P23 schools in a subsequent step. The details of this process are described later.

Approach Used to Select Independent Variables. In selecting the independent variables for the model, we sought factors likely to be highly correlated with certification error rates. As described in Ponza et al. (2007a), the two possible sources of certification error are administrative error and household reporting error. Administrative error occurs when school districts make mistakes in (1) processing applications or direct certification documents, (2) determining eligibility, (3) recording certification status information on the application, or (4) transmitting the status recorded on the application or direct certification documents onto the master eligibility list. Administrative error is likely to be most heavily influenced by administrative features of the school meal program in the district and other administrative characteristics of the district. Household reporting error occurs when households misreport

[16]Instead, the APEC study accounted for certification error at non–base-year P23 schools using an imputation process. See Ponza et al. (2007b) for details.

family size, income, or qualifying program participation on free/reduced-price meal benefit applications. It may be influenced both by administrative features of the programs (such as the type of verification procedures used) and by demographic characteristics of students and families in the district. Therefore, the explanatory variables we considered included indicators of the administrative features of the NSLP and the SBP in the district, other characteristics of the district, and demographic characteristics of students and families in the district.

Variables representing verification results were also included as independent variables in the model, as we believed that they would likely be highly predictive of districts' certification error rates. For example, one might expect that districts with high underlying error rates would yield verification samples of approved applications more likely to be found to be in error by the verification process. If true, this would imply that districts in which larger proportions of applications whose benefits were reduced or terminated by the verification process would also have higher certification error rates.[17] This relationship is complicated, however, by the fact that not all families selected for verification respond to the request for documentation; thus, the results among cases in which the families do respond are not necessarily representative of the full verification sample.

An alternative explanation for a relationship between verification results and underlying certification error rates is that the verification results proxy for hard-to-measure aspects of districts' administration of the NSLP and the SBP. It is possible that districts in which a large proportion of verified applications have their benefits reduced or terminated are simply districts that do a more careful job of protecting the integrity of the meal programs generally. These

[17]This would be particularly likely to be true in districts that selected random, as opposed to focused, verification samples. Given the potential for different verification results in districts that selected random versus focused samples, we interacted the variables showing verification results with a variable indicating the type of verification sample.

districts may more carefully review free and reduced-price meal applications and give families more information and assistance in completing these applications, which would result in lower certification error rates. At the same time, they may work harder to get families selected for verification to respond to the request for documentation and carry out the verification audit more carefully, resulting in higher proportions of verified cases having their benefits reduced or terminated. Under this scenario, districts in which larger proportions of applicants have their benefits reduced or terminated by the verification process would have lower certification error rates.

Thus, the two alternative explanations for expecting a relationship between verification results and certification error rates have very different implications as to the expected nature of this relationship. Each explanation, however, implies that the econometric model should include variables representing verification results, leaving the issue of which (if either) of the two explanations dominates an open, empirical question.

Key Additional Considerations. Aside from including in the model independent variables that theory suggests should be predictive of certification error rates, three additional considerations influenced our strategy for selecting a specification for the econometric model: (1) the limited number of degrees of freedom in the model, (2) the need to focus on policy-sensitive variables, and (3) the practical need to end up with a model that will be easy for FNS to use to predict future erroneous payments. We next describe each of these considerations.

The degrees of freedom for each of the equations to be estimated as part of the econometric model are limited by the fact that the APEC sample included just 87 districts. With a limited number of degrees of freedom, the number of independent variables whose relationship with certification error rates can be estimated with a reasonable degree of precision is also limited. Thus, we needed to be economical in selecting independent variables for the model. Partly for

this reason, we decided to allow each of the 12 equations in the model to have a unique set of independent variables. In other words, each equation includes the independent variables that best predict that equation's dependent variable.

Second, this need to be economical in defining the specification of each equation led us to focus especially on variables representing factors that could potentially proxy for or be influenced by the districts' efforts to improve the integrity of the NSLP or the SBP. If, for example, districts were making policy changes resulting in more accurate certification, we wanted independent variables whose values would change from year to year to reflect the underlying policy changes. These policy-sensitive variables would be more valuable in our model than variables such as student demographic characteristics that might be correlated with districts' error rates but would be unlikely to change much from year to year in response to changing meal program policies.

Finally, our strategy for selecting the model's independent variables was influenced by the fact that the results of the model are designed to be used in future years to predict erroneous payments. This future effort will involve assembling a data set that includes values of each independent variable over all districts nationally that offer the NSLP or the SBP. Thus, we had to select independent variables that would be available in future years and could be incorporated into a single data file with relative ease and at modest cost.

This consideration led us to use the FNS-742 data set described earlier as a starting point. Three features of this data set are particularly important for the modeling effort. First, aside from a relatively small set of districts that failed to provide data, the FNS-742 file contains information on the full population of public and private districts nationally that participate in the meal programs. Second, the data will be collected in future years and available to the FNS

relatively quickly at the conclusion of each school year. Third, the key independent variables in the model—verification procedures and results—are available in the FNS-742 data set.

In addition to the variables available from FNS-742 data, we also considered variables constructed from the other data sources listed in Section II.A.2. Although most of these data sources will be available in future years, they must be merged with the FNS-742 data to be used in the modeling effort. Because the system for identifying districts in the FNS-742 data set differs from that used by these other data sets, this data merge is not straightforward. Thus, any improvements in the model's predictive power arising from the inclusion of independent variables from these other data sources must be weighed against future costs of creating a merged data file that can be used to predict erroneous payments if such variables are included.[18]

Variable Selection. Given the considerations described above, our strategy for selecting a set of variables and specification for each of the 12 equations in the model consisted of the following elements:

- We first came up with a list of candidate variables to consider as potential independent variables in the model. This list included variables from FNS-742 as well as variables from the other data sources we acquired.
- Based on theoretical considerations as well as some preliminary empirical work, we selected from FNS-742 data a set of independent variables we defined as core variables in the model.
- We used an automated procedure for selecting an additional set of variables to be included in the equation as independent variables, with independent variables constructed using FNS-742 data included in the set of candidate variables. We then repeated this process, including as candidate variables independent variables from all available data sets.

[18]Obtaining data on number of meals by type for the NSLP and the SBP at the district level was especially time intensive and costly to process and was burdensome to respondents. These data must be obtained from the individual state agencies administering child nutrition programs. We sent letters to these agencies requesting the data and needed to follow up by telephone to clarify our request. The data received from state agencies came in various formats and modes, usually in electronic files, but sometimes in hard copy, which then needed to be scanned and converted to an electronic file.

The candidate variables we considered along with their data sources are listed in Table II.2. In addition to the basic version of these variables, we also considered squared terms of some of the key variables and interactions between selected variables. From among these candidate variables, the core variables represent verification procedures and results, variables representing the proportion of students in the district in various certification categories, and district enrollment. The variables representing verification results were defined a bit differently for the different equations in the model.[19] In equations in which the dependent variable represents an error rate among meals served to students certified for free meals, the core independent variables show verification results among the free meal applications that were verified. In equations in which the dependent variable is a reduced-price meal error rate, the core independent variables show verification results among the reduced-price applications that were verified. Finally, the core independent variables in the equations in which the dependent variable is an error rate among denied applicants show verification results among all applications that were verified.

After selecting the core variables, an automated process was used to select additional variables for the model. These additional variables were selected in a stepwise fashion based on correlations of all variables in the set being considered with each dependent variable, controlling for the core variables (that is, with the residual from the regression of each dependent variable on the core variables). The variables that explained the greatest proportion of the variation of this residual were included as additional independent variables in the model.[20] Using this process, we developed four specifications:

[19]We experimented with a number of other specifications of verification results variables, such as aggregating verification results to one percentage of all verified applications that changed or did not respond. We chose the selected specification based on the strength of the relationship with the dependent variables.

[20]Some exceptions were made when the automated process selected two variables that were based on similar content. For example, if both the number of applications verified and the percentage of applications verified were selected, we replaced the independent variable that had the weaker relationship with the dependent variable with another unrelated independent variable.

TABLE II.2

INDEPENDENT VARIABLES CONSIDERED FOR INCLUSION IN THE ECONOMETRIC MODEL, BY DATA SOURCE

Data Source	Variables	Additional Information
FNS-742	Type of SFA	Public or private
	Type of application used	Individual student, household, or both
	Number of schools operating the NSLP and/or the SBP	
	Number of P23 non–base-year schools	
	Number of enrolled students with access to the NSLP and/or the SBP	
	Percentage of students certified as free eligible by certification method	Calculated separately for free approved based on income/household size information, categorical eligibility, not subject to verification (e.g., direct certification), and in P23 non–base-year schools
	Percentage of students certified as reduced-price eligible	Overall and in P23 non–base-year schools only
	Applications selected for verification	Number
	Verification sampling method	Focused/error prone, random, or all applications
	Percentage of verified applications in which household responded, no change, by certification category	Percentages of free and reduced-price verified applications were calculated
	Percentage of verified applications in which household responded, benefits reduced or terminated, by certification category	Percentages of free and reduced-price verified applications were calculated
	Percentage of verified reduced-price applications in which household responded, benefits increased	
	Percentage of verified applications in which household did not respond, benefits terminated, by certification category	Percentages of free and reduced-price verified applications were calculated
	Percentage of verified applications for which outcome is missing, by certification category	Percentages of free and reduced-price verified applications were calculated
State Agency Administrative Data	Whether offers the SBP	
	Percentage of NSLP meals by type	Free, reduced price, and paid
	Percentage of SBP meals by type	Free, reduced price, and paid; severe and non-severe needs *are not* distinguished
	Percentage of SBP meals by type	Free, reduced price, and paid; severe and non-severe needs *are* distinguished
	Percentage of total meals which are severe needs	
	Percentages of total meals from the SBP and the NSLP	
	Percentage of total meals by meal program and type	
Common Core of Data	Total enrollment	
	Number of schools	
	Average school size	Number of students per school
	Enrollment, by race/ethnicity/gender/grade	Number and percentage
	Grade span of district	

Table II.2 *(continued)*

Data Source	Variables	Additional Information
	Urbanicity	Urban, suburban, town, rural
	Percentage certified for free and reduced-price meals	
	SFA administrators	Number and number per student
	SFA support staff	Number and number per student
	Teachers	Number, number per school, and students per teacher
	School administrators	Number and number per student
	School support staff	Number and number per student
	Spending on food services	Amount and amount per student
	Spending on food service salaries	Amount and amount per student
	Spending on administrative support services	Amount and amount per student
Private School Survey	Total enrollment	
	Number of schools	
	Average school size	Number of students per school
	Enrollment, by race/ethnicity/gender/grade	Number and percentage
	Grade span	
	Urbanicity	Urban, suburban, town, rural
	Teachers	Number, number per school, and students per teacher
Census and SAIPE	Local median household income	
	Local poverty rate among school age population	School age defined as ages 5 to 17
	SFAs expected number of students in poverty	Constructed by multiplying the poverty rate by the CCD report of SFA enrollment
	Ratio of free and reduced-price–certified students to the expected number of students in poverty	Constructed by dividing the CCD report of free and reduced-price students by the number expected to be in poverty
APEC Study	Whether district uses direct certification	
	Type of direct certification	No match, district-level match, state-level match, other method
	Whether uses a food management company to run its meal programs	
	Whether uses computer software to process and determine school meal benefit eligibility	
	Whether tracks individual student participation in the meal program	

1. Core variables only
2. Core variables plus three additional FNS-742 variables
3. Core variables plus six additional FNS-742 variables
4. Core variables plus three additional FNS-742 variables plus three additional variables from any data source

Ultimately, we compared the model results from the various specifications previously described to determine a preferred specification. The cost of using model results to predict erroneous payments in future years with the specification that includes only the core variables is about the same as the cost of doing so with the specification that includes the core variables plus three additional FNS-742 variables. Thus, we based our judgment about these two specifications on whether the three additional variables improved the goodness of fit of the model compared to the model that did not include these variables. In general, we found that these additional variables did explain a significant amount of variation; thus, we included them in our preferred specification. We then compared this specification to a specification that included three additional independent variables selected from among the full set of candidate variables, by examining model goodness of fit. Although we generally found higher goodness of fit for the specification that included all candidate variables, the improvement seemed relatively small compared to the cost of using a model that includes multiple data sources in future years. Thus, our final specification included the core variables and three additional FNS-742 variables. The specific variables included are listed in Tables II.3 and II.4.[21]

[21]Typically, in prediction models, one would use a randomly chosen sample from the full data set, to develop the model and use the larger remaining sample to re-estimate the primary models. This prevents over-fitting within sample. With the small number of districts in the APEC study sample this was not possible.

TABLE II.3

INDEPENDENT VARIABLES INCLUDED IN MODELS USED IN ESTIMATING NSLP ERRONEOUS PAYMENTS

	Overpayments			Underpayments		
	%CF-RPE-L	%CF-PE-L	%CRP-PE-L	%CRP-FE-L	%NC-FE-L	%NC-RPE-L
Verification Variables						
(1)	Used focused verification sample	Used focused verification sample	Used focused verification sample	Used focused verification sample	Used focused verification sample	Used focused verification sample
(2)	Percentage of verified free applications that had benefits reduced or terminated in verification	Percentage of verified free applications that had benefits reduced or terminated in verification	Percentage of verified reduced-price applications that had benefits reduced or terminated in verification	Percentage of verified reduced-price applications that had benefits reduced or terminated in verification	Percentage of all verified applications that had benefits changed in verification	Percentage of all verified applications that had benefits changed in verification
(3)	Interaction of (1) and (2)	Interaction of (1) and (2)	Interaction of (1) and (2)	Interaction of (1) and (2)	Interaction of (1) and (2)	Interaction of (1) and (2)
(4)			Percentage of verified reduced-price applications that had benefits increased in verification	Percentage of verified reduced-price applications that had benefits increased in verification		
(5)			Interaction of (1) and (4)	Interaction of (1) and (4)		
(6)	Percentage of verified free applications that did not respond in verification	Percentage of verified free applications that did not respond in verification	Percentage of verified reduced-price applications that did not respond in verification	Percent of verified reduced-price applications that did not respond in verification	Percent of all verified applications that did not respond in verification	Percent of all verified applications that did not respond in verification
(7)	Interaction of (1) and (6)	Interaction of (1) and (6)	Interaction of (1) and (6)	Interaction of (1) and (6)	Interaction of (1) and (6)	Interaction of (1) and (6)
(8)			Indicator that reduced-price verification results were missing	Indicator that reduced-price verification results were missing		

Table II.3 *(continued)*

	Overpayments			Underpayments		
	%CF-RPE-L	%CF-PE-L	%CRP-PE-L	%CRP-FE-L	%NC-FE-L	%NC-RPE-L
Certification Variables						
(9)	Percentage of students certified without an application	Percentage of students certified without an application			Percentage of students certified without an application	Percentage of students certified without an application
(10)	Percentage of students certified categorically	Percentage of students certified categorically			Percentage of students certified categorically	Percentage of students certified categorically
District Characteristics						
(11)	Enrollment	Enrollment	Enrollment	Enrollment	Enrollment	Enrollment
(12)	Percentage of students certified for free meals	Percentage of students certified for free meals	Percentage of students certified for free meals	Percentage of students certified for free meals	Percentage of students certified for free meals	Percentage of students certified for free meals
(13)	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals
Additional Variables						
(14)	Privately operated	Number of applications certified	Privately operated	Privately operated	Privately operated	Number of applications categorically certified as free eligible
(15)	Uses individual student application	Uses individual student application	Uses individual student application	Uses individual student application	Uses individual student application	Number of schools operating the NSLP and/or the SBP
(16)	Number of students certified as free eligible not subject to verification	Number of applications categorically certified as free eligible	Number of students certified as free eligible not subject to verification	Number of students certified as free eligible based on income	Number of students certified as free eligible not subject to verification	Number of applications verified

TABLE II.4

INDEPENDENT VARIABLES INCLUDED IN MODELS USED IN ESTIMATING SBP ERRONEOUS PAYMENTS

	Overpayments			Underpayments		
	%CF-RPE-B	%CF-PE-B	%CRP-PE-B	%CRP-FE-B	%NC-FE-B	%NC-RPE-B
Verification Variables						
(1)	Used focused verification sample	Used focused verification sample	Used focused verification sample	Used focused verification sample	Used focused verification sample	Used focused verification sample
(2)	Percentage of verified free applications that had benefits reduced or terminated in verification	Percentage of verified free applications that had benefits reduced or terminated in verification	Percentage of verified reduced-price applications that had benefits reduced or terminated in verification	Percentage of verified reduced-price applications that had benefits reduced or terminated in verification	Percentage of all verified applications that had benefits changed in verification	Percentage of all verified applications that had benefits changed in verification
(3)	Interaction of (1) and (2)	Interaction of (1) and (2)	Interaction of (1) and (2)	Interaction of (1) and (2)	Interaction of (1) and (2)	Interaction of (1) and (2)
(4)			Percentage of verified reduced-price applications that had benefits increased in verification	Percentage of verified reduced-price applications that had benefits increased in verification		
(5)			Interaction of (1) and (4)	Interaction of (1) and (4)		
(6)	Percentage of verified free applications that did not respond in verification	Percentage of verified free applications that did not respond in verification	Percentage of verified reduced-price applications that did not respond in verification	Percentage of verified reduced-price applications that did not respond in verification	Percentage of all verified applications that did not respond in verification	Percentage of all verified applications that did not respond in verification
(7)	Interaction of (1) and (6)	Interaction of (1) and (6)	Interaction of (1) and (6)	Interaction of (1) and (6)	Interaction of (1) and (6)	Interaction of (1) and (6)
(8)			Indicator that reduced-price verification results were missing	Indicator that reduced-price verification results were missing		

Table II.4 *(continued)*

	Overpayments			Underpayments		
	%CF-RPE-B	%CF-PE-B	%CRP-PE-B	%CRP-FE-B	%NC-FE-B	%NC-RPE-B
Certification Variables						
(9)	Percentage of students certified without an application	Percentage of students certified without an application			Percentage of students certified without an application	Percentage of students certified without an application
(10)	Percentage of students certified categorically	Percentage of students certified categorically			Percentage of students certified categorically	Percentage of students certified categorically
District Characteristics						
(11)	Enrollment	Enrollment	Enrollment	Enrollment	Enrollment	Enrollment
(12)	Percentage of students certified for free meals	Percentage of students certified for free meals	Percentage of students certified for free meals	Percentage of students certified for free meals	Percentage of students certified for free meals	Percentage of students certified for free meals
(13)	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals	Percentage of students certified for reduced-price meals
Additional Variables						
(14)	Number of applications certified	Number of students certified as free eligible based on income	Number of applications categorically certified as free eligible	Number of applications categorically certified as free eligible	Number of applications categorically certified as free eligible	Number of students categorically certified as free eligible
(15)	Uses individual student application	Uses individual student application	Number of applications certified	Uses individual student application	Number of students certified as free eligible not subject to verification	Uses individual student application
(16)	Number of applications categorically certified as free eligible	Privately operated	Uses individual student application	Number of students certified as free eligible based on income	Number of schools operating the NSLP and/or the SBP	Number of applications certified

2. Predicting Future Erroneous Payments

After we estimated the econometric model of error rates using data on the APEC study sample of districts, the estimated parameters from this model could be used to generate national estimates of overpayments, underpayments, and overall erroneous payments in future years. To illustrate this process, we have "predicted" erroneous payments in the study year (SY 2005–06) using the steps described later while ignoring direct information on erroneous payments in sampled districts collected as part of the APEC study. In other words, we have made this prediction under the premise that we have only the estimates from the econometric model and data on district characteristics for all districts nationally from the FNS-742 data set. We can then compare the predicted amounts and rates of erroneous payments using this process with the estimated erroneous payments amounts and rates from the APEC study. If the predicted and estimated amounts and rates of erroneous payments are close to each other, we will have more confidence in the ability of the model to predict erroneous payments in future years.

As described earlier, the econometric model is represented by 12 equations (6 for the NSLP and 6 for the SBP) that all have the form shown in Equation (1), reproduced here for convenience, with the estimated parameters represented by β_k:

(1) $E_{kj} = X_{kj}\beta_k + u_{kj}$, where k = 1, ... , 12

Given this framework, the process for predicting erroneous payments based on these parameters and data on the characteristics of school districts nationally is given by the following seven steps:

Step One: Gather the Data. The first step is to collect data for all districts nationally on all of the district characteristics represented by the variables in the vector X_{kj} for each of the 12 equations. In the preferred specification of the model, all the variables come from the FNS-742

data set, so collecting the data to be used in the model amounts to collecting and cleaning the FNS-742 data. The process for doing this data cleaning and imputation of missing values is described in Section II.A.2.

We treat the FNS-742 data as if there were an observation for every district that offers the NSLP and/or the SBP nationally. Although this is true in theory, in practice some districts fail to report FNS-742 data and are not represented in the data. The model's erroneous payment estimates are based on the assumption that rates of erroneous payments in these districts are similar to rates in districts that are represented in the data.[22]

Step Two: Generate Predicted NSLP and SBP Meal Error Rates in Each District Nationally. The next step involves multiplying the values of the variables in X_{kj} by the values of the associated coefficients from the econometric model to generate a predicted value of each type of NSLP and SBP meal error rate for each district in the FNS-742 data. In particular, there will be 12 predicted error rates for each of these districts, calculated as: $\widehat{\%\text{CF-RPE-L}}_j = X_{1j}\hat{\beta}_1$, ... , $\widehat{\%\text{NC-FE-B}}_j = X_{12j}\hat{\beta}_{12}$. Each will represent the predicted percentage of a particular type of meal (free, reduced price, or paid) served in the district that has a particular type of error (for example, a free meal that should have been reimbursed at the paid rate).

[22]As discussed earlier in Section II.A.2, in order to assess the similarity of districts that are not represented in the FNS-742 to those that are, we generated estimates of the percentage of students in each school meal certification category for both types of districts. Student counts for districts not represented in the FNS-742 were generated by subtracting totals of students by certification status in the FNS-742 from national totals available from the FNS national database, the CCD, and the PSS. We found similar percentages of students by certification status in the two groups of districts. In FNS-742 districts, 34 percent of students were certified for free meals, 8 percent were certified for reduced price meals, and 59 percent were not certified for school meal benefits. In districts not represented in the FNS-742, these percentages were 35 percent, 7 percent, and 58 percent, respectively. Although it is possible that these two types of districts differ in other important ways, it is reassuring that they are similar in one important dimension.

Step Three: Impute the Number of Each Category of Meals Served (Free, Reduced Price, or Paid) in Each District. The FNS-742 data set does not contain information on the number of free, reduced-price, or paid meals served in each district. It does, however, contain information on the number of enrolled *students* in each meal-price category in each district. In order to estimate the number of meals of each type served in a given district, we assumed that the district served the same percentage of the total meals served in the state, by eligibility category, as its percentage of the total number of students enrolled in the state, by eligibility category. For example, if 2 percent of all of Pennsylvania's free-certified students were enrolled in District X, we assumed that 2 percent of all the free meals served in Pennsylvania were served in District X. Following this example, imputing the number of free school lunches served in District X requires the following:

- Dividing the number of free-certified students in District X by the sum of free-certified students in all of Pennsylvania's districts who were present in the FNS-742 data. This fraction represents the proportion of Pennsylvania's free-certified students in District X.
- Obtaining data on the total number of free lunches served in Pennsylvania from the FNS national data file (Version 8.2 Public Use).
- Multiplying the proportion of Pennsylvania's free-certified students in District X by the total number of free lunches served in Pennsylvania. The resulting product represents the imputed number of free lunches served in District X.[23]

The process for imputing the number of school breakfasts served by District X in each category is analogous but includes one additional step. In particular, it is necessary to estimate the number of free and reduced-price breakfasts served in severe-needs schools, since the federal

[23]This procedure implicitly assumes that districts from a given state included in the FNS-742 database are representative of all districts in that state, including those missing from the database. This is because for a given state, students in districts not included in the FNS-742 database are distributed across all of the state's districts that are included in the database in proportion to their size.

reimbursement level for these meals is different from that of free and reduced-price meals served in non–severe-needs schools. Since the FNS-742 database does not include information on severe-needs status, we assumed that the district served the same percentage of the total severe need breakfasts served in the state, by eligibility category, as its percentage of the total number of students enrolled in the state, by eligibility category. For example, if the number of students certified for free meals in a district represents 2 percent of all free-certified students in the state, this assumption implies that the district also has 2 percent of the state's free meals served in severe-needs schools.

The end result of the set of calculations described in this step is a set of values for the following variables for each district nationally:

#CF-L_j :	Number of free lunches served in district j
#CRP-L_j :	Number of reduced-price lunches served in district j
#NC-L_j :	Number of paid lunches served in district j
#CF-B_j :	Number of free breakfasts served in all schools in district j
#CF-SNB_j :	Number of free breakfasts served in severe-needs school in district j
#CRP-B_j :	Number of reduced-price breakfasts served in all schools in district j
#CRP-SNB_j :	Number of reduced-price breakfasts in severe-needs school in district j
#NC-B_j :	Number of paid breakfasts served in all schools in district j

Step Four: Calculate the Total Number of Meals in Each Error Category for Each District. This calculation involves multiplying the district's error rate in each error category by its imputed number of free, reduced-price, or paid meals, as appropriate. For the NSLP, these calculations are:

- Number of free lunches erroneously served to reduced-price–eligible students in district j = (Number of free lunches served in j) * (imputed % of free lunches served to reduced-price–eligible students)

or,

$\#\text{CF-RPE-L}_j = \#\text{CF-L}_j * \widehat{\%\text{CF-RPE-L}_j}$

- $\#\text{CF-PE-L}_j = \#\text{CF-L}_j * \widehat{\%\text{CF-PE-L}_j}$
- $\#\text{CRP-PE-L}_j = \#\text{CRP-L}_j * \widehat{\%\text{CRP-PE-L}_j}$
- $\#\text{CRP-FE-L}_j = \#\text{CRP-L}_j * \widehat{\%\text{CRP-FE-L}_j}$
- $\#\text{NC-RPE-L}_j = \#\text{NC-L}_j * \widehat{\%\text{NC-RPE-L}_j}$
- $\#\text{NC-FE-L}_j = \#\text{NC-L}_j * \widehat{\%\text{NC-FE-L}_j}$

The calculations for the SBP are analogous to those of the NSLP, except that we must also calculate estimates of meals erroneously served in severe needs schools because the dollar amount associated with some error types is different in severe needs schools than in other schools. We do not have separate error rate estimates for severe-needs schools versus other schools; thus, when it is necessary, we assume that the error rates at the two sets of schools in a given district are the same. We also need to distinguish between erroneously served paid meals in severe-needs schools and other schools, since the amount of the erroneous payment will depend on whether the school is a severe-needs school. To do this, we multiplied the number of paid breakfasts served in the district by the fraction of free (or reduced-price) meals served to students in severe-needs schools in the district. These calculations for the SBP are:

- Number of free breakfasts erroneously served to reduced-price–eligible students in all schools in district j = (Number of free breakfasts served in j) * (imputed % of free breakfasts served to reduced-price–eligible students)

  or,

  $\#\text{CF-RPE-B}_j = \#\text{CF-B}_j * \widehat{\%\text{CF-RPE-B}_j}$

- $\#\text{CF-PE-B}_j = \#\text{CF-B}_j * \widehat{\%\text{CF-PE-B}_j}$

- Number of free breakfasts erroneously served to paid-eligible students in severe-needs schools in district j = (Number of free breakfasts served in severe-needs schools in j) * (imputed % of free breakfasts served to paid-eligible students)

  or,

  $\#\text{CF-PE-SNB}_j = \#\text{CF-SNB}_j * \widehat{\%\text{CF-PE-B}_j}$

- $\#\text{CRP-PE-B}_j = \#\text{CRP-B}_j * \widehat{\%\text{CRP-PE-B}_j}$
- $\#\text{CRP-PE-SNB}_j = \#\text{CRP-SNB}_j * \widehat{\%\text{CRP-PE-B}_j}$
- $\#\text{CRP-FE-B}_j = \#\text{CRP-B}_j * \widehat{\%\text{CRP-FE-B}_j}$
- $\#\text{NC-RPE-B}_j = \#\text{NC-B}_j * \widehat{\%\text{NC-RPE-B}_j}$
- $\#\text{NC-RPE-SNB}_j = \#\text{NC-B}_j * \left(\frac{\#\text{CRP-SNB}_j}{\#\text{CRP-B}_j} \right) * \widehat{\%\text{NC-RPE-B}_j}$
- $\#\text{NC-FE-B}_j = \#\text{NC-B}_j * \widehat{\%\text{NC-FE-B}_j}$
- $\#\text{NC-FE-SNB}_j = \#\text{NC-B}_j * \left(\frac{\#\text{CF-SNB}_j}{\#\text{CF-B}_j} \right) * \widehat{\%\text{NC-FE-B}_j}$

Step Five: Calculate the Total Dollars Erroneously Reimbursed in Each Meal Category as Well as the Total Dollars Reimbursed Overall for Lunch and Breakfast. To calculate dollars erroneously reimbursed, we multiply the total number of meals in each error category by the dollar value of erroneous payments per meal in that category. Per meal erroneous payments and total payments for each category for the NSLP and the SBP are shown in Tables I.1 and I.2 (see Chapter I). As noted in Table I.1, districts in which more than 60 percent of lunches are reimbursed at the free or reduced-price rates receive an extra reimbursement of $0.02 for each lunch served. This additional reimbursement does not affect our calculations of erroneous payments (since it applies to all meal types), but it does affect the calculation of total reimbursements. To account for these "60 percent districts," we created a binary indicator (FRP60-L_j) showing whether at least 60 percent of the district's lunches are

reimbursed at the free or reduced-price level.[24] We use this variable and other previously constructed variables to calculate erroneous payments and total reimbursements for the NSLP as follows:

- Total dollars of erroneous payments (overpayments) for free lunches served to reduced-price–eligible students in district j = (Number of free lunches served to reduced-price–eligible students in j) * (per meal erroneous payment for free lunches served to reduced-price–eligible students)

  or,

  $\$CF\text{-}RPE\text{-}L_j = \#CF\text{-}RPE\text{-}L_j * \$0.40$

- $\$CF\text{-}PE\text{-}L_j = \#CF\text{-}PE\text{-}L_j * \$2.10$
- $\$CRP\text{-}PE\text{-}L_j = \#CRP\text{-}PE\text{-}L_j * \$1.70$
- $\$CRP\text{-}FE\text{-}L_j = \#CRP\text{-}FE\text{-}L_j * \$0.40$
- $\$NC\text{-}RPE\text{-}L_j = \#NC\text{-}RPE\text{-}L_j * \$1.70$
- $\$NC\text{-}FE\text{-}L_j = \#NC\text{-}FE\text{-}L_j * \$2.10$
- Total dollars of NSLP reimbursements in a district are based on the number of lunches of each type served in the district:

  $FRP60\text{-}L_j = 1$ if $(\#CF\text{-}L_j + \#CRP\text{-}L_j) / (\#CF\text{-}L_j + \#CRP\text{-}L_j + \#NC\text{-}L_j) \geq 0.60$

  $= 0$ otherwise

  $\$TR\text{-}L_j = (\#CF\text{-}L_j * \$2.5127) + (\#CRP\text{-}L_j * \$2.1127) + (\#NC\text{-}L_j * \$0.4127)$
  $+ FRP60\text{-}L_j * (\#CF\text{-}L_j + \#CRP\text{-}L_j + \#NC\text{-}L_j) * \$0.02$

In the case of the SBP, the amount of erroneous payments for a given meal depends in part on whether the meal is served in a severe-needs schools. In particular, a larger per meal erroneous payment (an overpayment) is associated with free or reduced-price meals served to

[24]Our estimates of whether a district receives an additional reimbursement of $0.02 for each lunch served is based on the current year's percentage of free and reduced-price reimbursements. In actuality, a district's 60 percent status is based on the previous year's percentage of free and reduced-price reimbursements. This simplification is unlikely to be important since districts' distributions of free and reduced-price reimbursements tend to be stable from year to year.

students who should be getting paid meals at severe-needs schools versus the same situation at non–severe-needs schools. Conversely, there is a larger underpayment for meals consumed by students at severe-needs schools who applied for and are eligible for free or reduced-price meals, but were not certified for any benefits, versus the same situation at other schools. Finally, total SBP reimbursements will differ according to whether meals are served at severe-needs schools:

- Total dollars of erroneous payments (overpayments) for free breakfasts served to reduced-price–eligible students in district j = (Number of free breakfasts served to reduced-price–eligible students in j) * (per meal erroneous payment for free breakfasts served to reduced-price–eligible students)

  or,

  $\$CF\text{-}RPE\text{-}B_j = \#CF\text{-}RPE\text{-}B_j * \$0.30$

- $\$CF\text{-}PE\text{-}B_j = \#CF\text{-}PE\text{-}B_j * \$1.04 + \#CF\text{-}PE\text{-}SNB_j * \$0.24$
- $\$CRP\text{-}PE\text{-}B_j = \#CRP\text{-}PE\text{-}B_j * \$0.74 + \#CRP\text{-}PE\text{-}SNB_j * \$0.24$
- $\$CRP\text{-}FE\text{-}B_j = \#CRP\text{-}FE\text{-}B_j * \$0.30$
- $\$NC\text{-}RPE\text{-}B_j = \#NC\text{-}RPE\text{-}B_j * \$0.74 + \#NC\text{-}RPE\text{-}SNB_j * 0.24$
- $\$NC\text{-}FE\text{-}B_j = \#NC\text{-}FE\text{-}B_j * \$1.04 + \#NC\text{-}FE\text{-}SNB_j * 0.24$
- Total dollars of SBP reimbursements in a district are based on the number of breakfasts of each type served in the district, as well as the number of these breakfasts served in severe-needs schools:

  $$\$TR\text{-}B_j = (\#CF\text{-}B_j * \$1.27 + \#CF\text{-}SNB_j * 0.24) + (\#CRP\text{-}B_j * \$0.97 + \#CRP\text{-}SNB_j * \$0.24) + (\#NC\text{-}B_j * \$0.23)$$

Step Six: Calculate the Preliminary Estimates of Total Reimbursements as Well as the Total Amounts and Rates of Overpayments, Underpayments, and Overall Erroneous Payments Across All Districts Nationally. Since we calculated total reimbursements as well as erroneous payments in each of the relevant error categories for each district nationally in the previous step, this step simply requires us to sum these totals across districts, after grouping the

appropriate error categories into either overpayments or underpayments. The preliminary estimates of erroneous payments rates are calculated by dividing the initial amount of erroneous payments by total reimbursements. For the NSLP, the relevant calculations are as follows:

- $\text{OP-L}_{\text{prelim}} = \sum_{j=1}^{J} \{\$\text{CF-RPE-L}_j + \$\text{CF-PE-L}_j + \$\text{CRP-PE-L}_j\}$
- $\text{UP-L}_{\text{prelim}} = \sum_{j=1}^{J} \{\$\text{CRP-FE-L}_j + \$\text{NC-RPE-L}_j + \$\text{NC-FE-L}_j\}$
- $\text{EP-L}_{\text{prelim}} = \text{OP-L}_{\text{prelim}} + \text{UP-L}_{\text{prelim}}$
- $\text{TR-L}_{\text{prelim}} = \sum_{j=1}^{J} \$\text{TR-L}_j$
- $\text{OPR-L}_{\text{prelim}} = \frac{\text{OP-L}_{\text{prelim}}}{\text{TR-L}_{\text{prelim}}} * 100$
- $\text{UPR-L}_{\text{prelim}} = \frac{\text{UP-L}_{\text{prelim}}}{\text{TR-L}_{\text{prelim}}} * 100$
- $\text{EPR-L}_{\text{prelim}} = \frac{\text{EP-L}_{\text{prelim}}}{\text{TR-L}_{\text{prelim}}} * 100$

An analogous set of calculations can be made for the SBP:

- $\text{OP-B}_{\text{prelim}} = \sum_{j=1}^{J} \{\$\text{CF-RPE-B}_j + \$\text{CF-PE-B}_j + \$\text{CRP-PE-B}_j\}$
- $\text{UP-B}_{\text{prelim}} = \sum_{j=1}^{J} \{\$\text{CRP-FE-B}_j + \$\text{NC-RPE-B}_j + \$\text{NC-FE-B}_j\}$
- $\text{EP-B}_{\text{prelim}} = \text{OP-B}_{\text{prelim}} + \text{UP-B}_{\text{prelim}}$
- $\text{TR-B}_{\text{prelim}} = \sum_{j=1}^{J} \$\text{TR-B}_j$
- $\text{OPR-B}_{\text{prelim}} = \frac{\text{OP-B}_{\text{prelim}}}{\text{TR-B}_{\text{prelim}}} * 100$

- $\text{UPR-B}_{\text{prelim}} = \frac{\text{UP-B}_{\text{prelim}}}{\text{TR-B}_{\text{prelim}}} * 100$

- $\text{EPR-B}_{\text{prelim}} = \frac{\text{EP-B}_{\text{prelim}}}{\text{TR-B}_{\text{prelim}}} * 100$

These are all preliminary estimates because the error rates on which these estimates are based are applicable only to non-P23 and P23 base-year schools, since the underlying sample upon which the econometric model was estimated excluded P23 non–base-year schools. However, we would like our estimates to reflect erroneous payments at P23 non–base-year schools as well. Our approach for incorporating these schools is described in the next step.

Step Seven: Adjust the Preliminary Estimates of the Erroneous Payments Rates to Account for P23 Non–Base-Year Schools. As noted earlier, the preliminary estimates of the rates of overpayments, underpayments, and total erroneous payments calculated in Step Six reflect information on error rates only at non-P23 and P23 base-year schools. In the main APEC study, Ponza et al. (2007a) initially estimated rates of erroneous payments for this same group but then used an imputation process to estimate rates of erroneous payments in P23 non–base-year schools and combined the two sets of estimates to generate the final estimated rates of erroneous payments at all schools.

In developing a method for predicting rates of erroneous payments in all schools nationally in future years, we use the assumption that the relationship between these rates at non-P23/P23 base-year schools and the same rates across all schools remain constant over time. Thus, we can use this relationship that was estimated in the APEC study to adjust the preliminary predictions of erroneous payments rates in future years calculated in Step Six to generate predictions of these rates that will apply to all schools.

For example, the final APEC estimate of the rate of overpayments in the NSLP was 7.1092 percent. This was slightly lower than the estimated rate of overpayments at non-P23/P23 base-year schools alone, which was 7.1294 percent. Thus, if the preliminary predicted rate of NSLP overpayments in non-P23/P23 base-year schools was 8.00 percent in some future year, one would expect that the predicted overpayment rate in all schools would be slightly lower than this. In particular, we could generate a predicted rate in all schools as follows: 8.00 * (7.1092 / 7.1294) = 7.9773. In other words, the ratio of the rate of the NSLP overpayments at all schools to the rate of NSLP overpayments at non-P23/P23 base-year schools from the APEC study is used to translate the predicted rate of overpayments at non-P23/P23 base-year schools in some future year to a predicted rate of overpayments at all schools in that future year. Table II.5 shows the adjustment factors we use as part of this calculation.

TABLE II.5

ADJUSTMENT FACTORS FOR CALCULATING PREDICTED RATES OF ERRONEOUS PAYMENTS AT ALL SCHOOLS

	NSLP			SBP		
Category	All Schools	Non-P23/P23 Non–Base Year	Adjustment Factor	All Schools	Non-P23/P23 Non–Base Year	Adjustment Factor
Overpayment Rate	7.1092	7.1294	0.9972	7.0691	7.1495	0.9888
Underpayment Rate	2.3077	2.1167	1.0902	2.0640	1.9027	1.0848

Source: Ponza et al. (2007a).

This results in the following calculations to determine the final predicted rates of erroneous payments:

- $OPR\text{-}L_{final} = OPR\text{-}L_{prelim} * (0.9972)$
- $UPR\text{-}L_{final} = UPR\text{-}L_{prelim} * (1.0902)$
- $EPR\text{-}L_{final} = OPR\text{-}L_{final} + UPR\text{-}L_{final}$
- $OPR\text{-}B_{final} = OPR\text{-}B_{prelim} * (0.9888)$
- $UPR\text{-}B_{final} = UPR\text{-}B_{prelim} * (1.0848)$
- $EPR\text{-}B_{final} = EPR\text{-}B_{prelim} * OPR\text{-}B_{final} + UPR\text{-}B_{final}$

III. ECONOMETRIC MODEL RESULTS

This chapter describes the econometric model we developed to estimate the amounts and rates of erroneous payments in the school meal programs in future years. In Section A, we discuss results from the ordinary least squares (OLS) version of the regression equations that form the basis of the model. Section B presents findings that result when alternate specifications of these regressions are used. In Section C, we describe the performance of the model when applied to national FNS-742 data from SY 2005-06. We discuss strengths and some limitations of the model in Section D.

A. OLS ECONOMETRIC MODEL REGRESSION RESULTS

In carrying out the imputation procedures, we estimated the 12 OLS regression equations that make up the model using the APEC study sample districts. As discussed in the previous chapter, dependent variables for these models were constructed from APEC study data. Explanatory variables, such as verification results and SFA characteristics, were based on data from the SY2005-2006 FNS-742 verification summary reports.

We focus our discussion of estimates from the econometric model on the regression equations for the percentage of free meals served to students not eligible for either free or reduced-price meals (%CF-PE-L for the NSLP, and %CF-PE-B for the SBP). Errors of this type constitute approximately 50 percent of total erroneous payments for both the NSLP and SBP (Ponza, et al. 2007a). As a result, the estimates from these two equations are most critical to the performance of the model as a whole. These estimates are presented in Table III.1. Coefficient estimates and goodness-of-fit statistics from all 12 estimated equations are presented in Appendix Table 1.

TABLE III.1

COEFFICIENT ESTIMATES FROM ESTIMATED REGRESSION EQUATIONS

		%CF-PE-L	%CF-PE-B
Verification Variables			
(1)	Selected focused verification sample	-5.392 (4.225)	-1.939 (2.773)
(2)	Percentage of free applications changed to reduced-price or paid status during verification	-0.044 (0.043)	-0.033 (0.067)
(3)	Interaction of (1) and (2)	0.184 (0.080)**	0.162 (0.089)*
	F-statistic of (2)+(3)	[6.08]**	[5.08]**
(4)	Percentage of free applications in which household did not respond to Verification	0.063 (0.054)	0.079 (0.055)
(5)	Interaction of (1) and (6)	-0.007 (0.073)	-0.067 (0.066)
	F-statistic of (4)+(5)	[1.46]	[0.09]
Certification Variables			
(6)	Percentage of certified students that were certified non-applicants	-0.006 (0.053)	0.054 (0.041)
(7)	Percentage of certified students that were certified categorically	-0.091 (0.091)	0.015 (0.081)
Core SFA Characteristics			
(8)	Enrollment (in tens of thousands)	0.003 (0.009)	0.010 (0.062)
(9)	Percentage of students certified eligible for free meals	-0.067 (0.035)*	-0.051 (0.035)
(10)	Percentage of students certified for reduced-price meals	-0.211 (0.291)	-0.034 (0.180)
Additional FNS-742 Variables			
(11)	Application is for individual student	-2.699 (2.740)	-1.852 (1.779)
(12)	SFA privately operated	14.303 (5.104)***	
(13)	Free eligibles not subject to verification (in tens of thousands)	0.075 (0.577)	
(14)	Total number of applications certified (in tens of thousands)		0.307 (0.126)**

Table III.1 *(continued)*

		%CF-PE-L	%CF-PE-B
(15)	Number of verified applications that were free eligible categorically (in thousands)		-0.863 (0.308)***
	Observations	**86**	**74**
	R-squared	**.38**	**.23**

Source: FNS-742 Verification Summary Reports and APEC study.

Note: Robust standard errors in parentheses.

*Significant at 10 percent.
**Significant at 5 percent.
***Significant at 1 percent.

In general, the regression equations explain the variance of the %CF-PE-L and %CF-PE-B dependent variables well. The R-squared value of the %CF-PE-L model is .38, while that of the %CF-PE-B model is .23 (Table III.1). The R-squared values of the models for other dependent variables are also generally high, ranging from .39 to .13 (Appendix Table 1).

The coefficient estimates of the %CF-PE-L and %CF-PE-B models are generally similar in terms of magnitude and statistical significance. The following discussion of these estimates is organized by type of explanatory variable.

Variables Related to Verification Results. As described in Chapter II, verification results may be related to a particular certification error rate, either because the verification process actually creates a rough estimate of the underlying error rate or because the process proxies for the extent to which the district's administration of the meal programs is designed to promote program integrity. Whatever the underlying reason for a relationship between verification results and certification error, it is possible that this relationship may differ depending upon whether a district selects a focused verification sample or a random sample. Thus, the model specification allows the estimated coefficients for variables related to the results of the verification process to vary by the type of method used by the SFA to select its verification sample.

In SFAs selecting focused verification samples, the estimated relationship between verification results and %CF-PE error rates is consistent with the idea that the verification process provides a measure of the underlying error rates. Districts that have higher percentages of verified applications in which benefits are reduced or terminated based on documentation provided by responding households tend to have higher %CF-PE-L and %CF-PE-B error rates. In particular, a 10-point increase in the percentage of certified free applications changed to reduced-price or paid status during verification is associated with a 1.40 percentage point increase in the predicted value of %CF-PE-L; this relationship is statistically significant at the

.05 level.[1] Similarly, a 10 percentage point increase in this same variable is associated with a 1.29 percentage point increase in the predicted value of %CF-PE-B; this relationship is also statistically significant at the .05 level. For SFAs selecting applications for verification using focused sampling, there is a small positive association between %CF-PE error rates and the percentage of certified free applications selected for verification in which the household did not respond that is not statistically significant.

For SFAs selecting verification samples by methods other than through focused sampling (mostly through random sampling), higher percentages of certified free applications changed to reduced-price or paid status during verification are associated with lower %CF-PE error rates although this relationship is not statistically significant. In particular, a 10 point increase in the percentage of certified free applications changed to reduced-price or paid status in verification is associated with a 0.44 percentage point *decrease* in the predicted value of %CF-PE-L. Similarly, a 10 point increase in the percentage of certified free applications changed to reduced-price or paid status in verification is associated with a 0.33 percentage point *decrease* in the predicted value of %CF-PE-B. Although we can not reject that these negative coefficients are equal to zero, it is possible that this variable is serving as a proxy for how rigorously the SFA checks applications in its verification process. If the rigor of an SFA's verification checks is positively associated with the rigor with which it reviews applications in the initial certification process, a negative association of changes in verification and %CF-PE error rates would follow.

[1]This marginal effect is calculated by summing two relevant coefficients and multiplying by ten. In other words, we added the coefficient on the percentage of certified free applications changed to reduced-price or paid status during verification (-0.044) and the coefficient on the interaction of this variable with whether the SFA selected a focused verification sample (0.184). We then multiplied this sum (0.140) by ten. Statistical significance of the marginal effect was determined using a Wald test of whether the sum of these two coefficients was equal to zero; *F*-statistics from these tests are reported in brackets in Table III.1.

As with the SFAs selecting verification samples through focused sampling, those using other methods to select verification samples show a positive association between %CF-PE error rates and the percentage of certified free applications selected for verification in which the household did not respond to the request for documentation that is not statistically significant.

Variables Related to the Method of Certification. Students may be certified eligible to receive free or reduced-price meals in several ways. They may be certified for free or reduced-price meals based on their family income and family composition. They may be certified for free meals based on categorical eligibility; that is, by reporting a food stamp, FDPIR or TANF case number on their application. In SFAs that utilize direct certification, students may be determined eligible for free meals without submitting an application. Under direct certification, districts use information provided by FSP-, FDPIR-, and TANF-administering agencies to establish that a student is a member of a household participating in one of these programs and is thus automatically eligible to receive free meals. It is possible that these different methods of certification may be associated with different rates of erroneous payments. Districts that have higher percentages of students that are determined to be eligible for benefits categorically may have less error for two reasons. First, it is simpler for families to report program participation on their free/reduced-price meal applications than it is to report income, thus reducing reporting error. Second, determining the eligibility of these households is simpler for program administrators, thus reducing administrative error. Similarly, districts are able to eliminate reporting error entirely for those students who are certified without an application.

The percentage of certified free applications that are approved by establishing categorical eligibility for free meals is associated with a lower %CF-PE-L error rate, although this association is not statistically significant. A 10 percentage point increase in this measure is associated with a 0.91 percentage point decline in the %CF-PE-L error rate. This suggests that in

the NSLP, being approved for free meals based on a food stamp, FDPIR, or TANF case number may be associated with less certification error than is being approved based on household income. However, the relationship between categorical certification and this type of certification error is much smaller and is not statistically significant for the SBP. The relationship between the percentage of certified students who are certified without an application (mainly through direct certification) and the %CF-PE error rates is small and not statistically significant in both the NSLP and the SBP.

Other SFA Characteristics. Of the remaining variables included in the %CF-PE equations, few of the associations with certification error rates are statistically significant. Three core SFA characteristics were included in all regression equations: SFA enrollment, the percentage of enrolled students certified for free meals, and the percentage of enrolled students certified for reduced-price meals. The percentage of students certified for free meals is significantly negatively associated with %CF-PE-L. The other two core SFA characteristics are not significantly associated with %CF-PE-L, and none of the three are significantly associated with %CF-PE-B.

Three additional SFA characteristics were included in the model based on their relatively strong correlations with %CF-PE error rates. In the %CF-PE-L equation, these variables are: (1) whether the SFA used individual student (as opposed to household) applications, (2) whether the SFA was privately operated, and (3) the number of free eligible students not subject to verification. In the %CF-PE-B equation, these variables are (1) whether the SFA used individual student applications, (2) the total number of applications certified, and (3) the number of applications that were categorically eligible for free meals. Privately operated SFAs have significantly higher rates of %CF-PE-L; none of the other additional characteristics are significantly associated with %CF-PE-L. The total number of applications certified has a

positive, statistically significant association with %CF-PE-B, while the number of verified applications that were free eligible categorically has a negative, significant relationship with %CF-PE-B.

B. ALTERNATE SPECIFICATIONS OF THE REGRESSION EQUATIONS

We explored three alternate specifications to the preferred OLS regression equations on which the econometric model is based. These alternate specifications are: (1) including only the core variables of the preferred specification—that is, excluding the three additional FNS-742 variables; (2) adding three more FNS-742 variables to the preferred specification, for a total of six additional variables; and (3) adding three more variables to the preferred specification that could have been drawn from any available dataset. In addition to exploring alternative OLS specifications, we estimated the econometric model using Tobit regression equations, which account for the fact that our dependent variables are left-censored at zero.

1. Alternate OLS Specifications

Relevant goodness of fit measures for the OLS specifications are presented in Table III.2. In particular, this table reports the R-squared value for the preferred specification and for each of the three alternate specifications for the 12 dependent variables. R-squared values are higher when the independent variables do a better job of explaining the variation in the dependent variable, and increase with an increase in the number of explanatory variables.[2]

In general, goodness of fit measures support the preferred specifications over alternate specifications that use only FNS-742 data. There is evidence that using variables from sources

[2]We also examined other goodness of fit measures, including adjusted R-squared and the Akaike Information Criterion. The pattern that emerges from these measures is similar to that shown with R-squared values.

TABLE III.2

GOODNESS OF FIT MEASURES FOR VARIOUS SPECIFICATIONS OF REGRESSION EQUATIONS RELATED TO NSLP AND SBP ERROR RATES

		NSLP					
		%CF-PE-L	%CF-RE-L	%CR-PE-L	%CR-FE-L	%NC-FE-L	%NC-RE-L
(1)	Preferred specification	0.38	0.15	0.13	0.23	0.25	0.18
(2)	Remove 3 error rate-specific FNS-742 variables from (1)	0.19	0.14	0.10	0.09	0.21	0.15
(3)	Add 3 error rate-specific FNS-742 variables to (1)	0.40	0.15	0.15	0.23	0.27	0.18
(4)	Add 3 error rate-specific FNS-742 variables from any available dataset to (1)	0.48	0.19	0.19	0.28	0.43	0.27
		SBP					
		%CF-PE-B	%CF-RE-B	%CR-PE-B	%CR-FE-B	%NC-FE-B	%NC-RE-B
(1)	Preferred specification	0.23	0.19	0.17	0.14	0.20	0.39
(2)	Remove 3 error rate-specific FNS-742 variables from (1)	0.18	0.13	0.13	0.13	0.16	0.32
(3)	Add 3 error rate-specific FNS-742 variables to (1)	0.26	0.22	0.18	0.15	0.25	0.41
(4)	Add 3 error rate-specific FNS-742 variables from any available dataset to (1)	0.49	0.28	0.21	0.19	0.26	0.41

Source: FNS-742 Verification Summary Reports and APEC study.

other than the FNS-742 improves model fit; however, we do not believe that this improvement justifies the cost of incorporating additional datasets into national estimates. The alternate OLS specifications are discussed below.

Augmenting the Core Independent Variables with Three Additional Variables from FNS-742. Across the 12 equations of the model, goodness of fit measures typically support augmenting the core variables in our preferred specification with three additional FNS-742 variables. For example, the R-squared value for the %CF-PE-L equation is .38 when these variables are included, compared to .19 when they are not. The R-squared value for the %CF-PE-L equation is .23 when the additional FNS-742 variables are included, compared to .18 when they are not.

Including Additional Variables from FNS-742 in the Preferred Specification. When the preferred specification is augmented with additional variables from the FNS-742, there are very modest increases in R-squared values. For example, the R-squared value for the preferred specification of the %CF-PE-L regression is .38, compared to .40 when three additional variables are included; for the %CF-PE-B equation, these values are .22 and .26 respectively. Based on this evidence, we did not expand the number of additional FNS-742 variables included in the model.

Including Additional Variables from Any Available Dataset. Goodness of fit measures do improve when the preferred specification is augmented with variables from datasets other than the FNS-742, including variables from the Common Core data, the Private School Survey, the Decennial Census and Small Area Income and Poverty Estimates, State Agency Administrative data, and APEC study data. The gains from adding these variables tend to be larger than those from adding additional variables from the FNS-742. In the %CF-PE-L regression, for example, the R-squared value increases from .38 to .48 when these variables are

added In the %CF-PE-B regression, the R-squared value increases from .23 to .49 when these variables are added.

As discussed in the previous chapter, linking the FNS-742 to other datasets is costly and time-consuming. Although including non-FNS-742 variables improves model fit in most cases, we do not believe that these gains outweigh the additional data preparation costs that would be incurred in future years.

2. Tobit Specification

In addition to exploring specifications that change the included variables in the OLS models, we also experimented with using Tobit models to estimate the relationship between the 12 error rates and the independent variables. Results from these models are presented in Appendix Table 2. In general, the magnitude of the coefficients and the levels of statistical significance are similar in the OLS and Tobit specifications.

C. GENERATING NATIONAL ESTIMATES OF ERRONEOUS PAYMENTS

In this section we report the results of using the econometric model with national data from the FNS-742 to predict erroneous payments for SY 2005 - 06. This serves three purposes: (1) to illustrate how the model works; (2) to generate baseline estimates to which predicted amounts and rates of erroneous payments derived from the model in future years can be compared; and (3) to assess the performance of the model in predicting erroneous payments for SY 2005 - 06. For the third purpose, we can use the estimates of erroneous payments from the APEC study as a standard. Since estimates from the APEC study are subject to sampling error, the appropriate standard for comparison is a confidence interval around the APEC study estimates. As such, we compare the predicted erroneous payment amounts and rates that emerge from the econometric model to the 90 percent confidence intervals of the APEC study estimates.

After estimating the regression equations of the econometric model with the APEC study sample, the coefficient estimates from these equations were applied to national data from the FNS-742 for SY 2005-2006. This process yielded a set of estimated error rates for each SFA nationally. As shown in Table III.3, the imputed error rates for the national FNS-742 sample are similar to the estimated error rates obtained from the APEC study sample. For example, in the APEC study, 5.7 percent of total free NSLP meals were served to students who were not eligible for free or reduced-price meals. Based on imputations from the econometric model, we estimate a similar rate of 6.0 percent for the national FNS-742 sample. In the SBP, imputations suggest that 5.4 percent of free meals were received by students not eligible for free or reduced-price meals. This rate compares favorably with the analogous APEC study estimate of 5.7 percent.

Because the imputed error rates derived from the model's regression equations are expressed in terms of percentages of meals, we needed measures of the number of meals served by an SFA in order to produce national estimates of the amount or rate of erroneous payments. However, reimbursable *meal* totals are not available in the FNS-742. As discussed in the previous chapter, we imputed meal totals by reimbursement category using information that is available in the FNS-742 on the number of *students* in each meal price category in the SFA. For example, if an SFA has 0.3 percent of its state's free-certified students, we assumed that it also serves 0.3 percent of its state's total number of free meals. As shown in Table III.4, average imputed NSLP meal counts for the FNS-742 sample during the 2005-2006 school year were 143,399 free lunches, 28,047 reduced-price lunches, and 118,454 paid lunches. Average imputed SBP meal counts were 69,124 free breakfasts, 8,878 reduced-price breakfasts, and 18,048 paid breakfasts.

Using imputed error rates and meal totals, we estimated the number of meals of a given type that were in error for each SFA. Following the method described in Chapter II, we generated dollar estimates of overpayments, underpayments, and total erroneous payments for each SFA.

TABLE III.3

ERROR RATES ESTIMATED FROM APEC STUDY SAMPLE AND IMPUTED FROM ECONOMETRIC MODEL
(Percentage)

	APEC Study Findings	Imputed from Econometric Model
NSLP		
Free Meals Served to Students Not Eligible for Free or Reduced-Price Meals	5.7	6.0
Free Meals Served to Reduced-Price-Eligible Students	7.9	8.0
Reduced-Price Meals Served to Students Not Eligible for Free or Reduced-Price Meals	23.2	23.7
Reduced-Price Meals Served to Students Eligible for Free Meals	26.7	26.5
Paid Meals Served to Students Eligible for Free Meals	1.5	1.2
Paid Meals Served to Students Eligible for Reduced-Price Meals	1.3	1.2
SBP		
Free Meals Served to Students Not Eligible for Free or Reduced-Price Meals	5.7	5.4
Free Meals Served to Reduced-Price-Eligible Students	6.9	6.5
Reduced-Price Meals Served to Students Not Eligible for Free or Reduced-Price Meals	24.6	27.9
Reduced-Price Meals Served to Students Eligible for Free Meals	24.7	27.9
Paid Meals Served to Students Eligible for Free Meals	1.4	1.5
Paid Meals Served to Students Eligible for Reduced-Price Meals	1.9	1.6

Source: FNS-742 Verification Summary Reports and APEC study.

TABLE III.4

AVERAGE MEAL COUNTS IMPUTED FOR THE FNS-742 NATIONAL SAMPLE OF SFA'S

	Total Meals
NSLP	
Free Meals	143,399
Reduced-Price Meals	28,047
Paid Meals	118,434
Additional $0.02 Meals	119,110
SBP	
Free Meals	69,164
Reduced-Price Meals	8,878
Paid Meals	18,048
Severe-Needs Free Meals	61,131
Severe-Needs Reduced-Price Meals	7,364

Source: FNS-742 Verification Summary Reports.

These amounts were then summed to generate national rates of erroneous payments and adjusted to account for erroneous payments in Provision 2 or 3 schools not in their base years.

Table III.5 presents national estimates of predicted erroneous payments derived from the econometric model along with the main findings from the APEC study for SY 2005 - 06.[3] For both the NSLP and the SBP, the model predicts overpayment and total erroneous payments well but predicts underpayment rates and amounts that are less than those from the APEC study. In the NSLP, overpayments imputed from the econometric model are especially close to those estimated in the APEC study. Imputed overpayments were $578 million (7.2 percent of total reimbursements); overpayments from the APEC study were estimated to be about 1 percent lower at $573 million (7.1 percent of total reimbursements). Imputed underpayments, which were $152 million (1.89 percent of total reimbursements), were less than underpayments estimated in the APEC study, which were $186 million (2.31 percent of total reimbursements). Imputed underpayments fell below the 90 percent confidence interval around the APEC underpayment estimate, the lower bound of which was $161 million (6.09 percent of total reimbursements).

One reason that imputed findings from the model and the APEC study findings do not match as well for underpayments as they do for overpayments may be that two of the underlying underpayment rates (%NC-FE-L and %NC-RE-L) require additional adjustments after they are estimated in the APEC study. In particular, the dependent variables in these cases are adjusted

[3]Appendix Table 3 presents national erroneous payment estimates derived from the Tobit-based model. Results from this model are similar to those derived from the OLS-based model. For the NSLP, the Tobit-based model does a somewhat better job of predicting underpayments but is somewhat worse at predicting overpayments and total erroneous payments. For example, the difference between the APEC results for total erroneous payment in the NSLP and the results from the OLS-based model is about $29 million, while the difference for the Tobit-based model is about $28 million. For the SBP, the Tobit-based model generates predicted underpayments and total erroneous payments that differ from the APEC results more than do those of the OLS-based model.

TABLE III.5

COMPARISON OF NATIONAL ESTIMATES OF ERRONEOUS PAYMENTS BASED ON APEC STUDY AND ON OLS IMPUTATION MODEL

	APEC Study		OLS Imputation Model	
	Erroneous Payments (in millions of dollars)	Percentage of All Reimbursements in Error	Erroneous Payments (in millions of dollars)	Percentage of All Reimbursements in Error
NSLP				
Overpayments	573 (50) [491-655]	7.11 (0.62) [6.09-8.13]	578	7.21
Underpayments	186 (15) [161-211]	2.31 (0.19) [2.00-2.62]	152	1.89
Total Erroneous Payments	759 (54) [670-848]	9.42 (0.67) [8.32-10.52]	730	9.10
SBP				
Overpayments	137 (17) [109-165]	7.07 (0.91) [5.58-8.56]	142	7.13
Underpayments	40 (6) [30-50]	2.08 (0.29) [1.60-2.56]	25	1.26
Total Erroneous Payments	177 (18) [147-207]	9.15 (0.94) [7.61-10.69]	167	8.39

Source: FNS-742 Verification Summary Reports and APEC study.

Note: Standard errors in parentheses; 90 percent confidence interval in brackets.

so that they apply to all paid meals rather than only to those served to denied applicants.[4] This step may introduce additional error into the imputation process. In addition, it is possible that the imputation of meal totals is less accurate for paid meals than for free and reduced-price meals, which would affect underpayments, but not overpayments. The key assumption of the meal total imputations is that a district's percentage of its state's meals by certification category is the same as its percentage of its state's students by certification category. This may not be as accurate for paid meals since the propensity of students not certified for free or reduced-price meals to consume school meals may vary from district to district.[5]

Although underpayments imputed by the model are somewhat less than underpayments estimated in the APEC study, imputed total erroneous payments are more similar to the APEC findings. Total erroneous payments generated by the econometric model were $730 million (9.10 percent of total reimbursements), compared to $759 million (9.42 percent of total reimbursements) estimate from the APEC study.

As with the NSLP, imputed error rates and dollar amounts of error associated with overpayments and total erroneous payments in the SBP fall within 90 percent confidence intervals around the equivalent APEC estimates, while those associated with underpayments do not. SBP overpayments predicted by the model were $142 million (7.21 percent of all reimbursements), slightly more than the APEC study estimate of $137 million (7.07 percent of all reimbursements). Differences for underpayments were larger. Predicted underpayments

[4]The APEC study sample contains denied applicants, but does not include other students who receive meals at the paid rate (those who did not apply for free or reduced-price meals). Thus, the error rates related to paid meals must be multiplied by an estimate of the ratio of the number of meals served to denied applicants to the total number of paid meals. See Chapter II for more discussion of this point.

[5]Because free and reduced price students are all subject to the same certification requirements, it is probably reasonable to assume that these students have similar levels of need and similar propensities to receive reimbursable meals. If this is the case, a district's percentage of its state's meals by certification category will match its percentage of its state's students by certification category.

were $25 million (1.26 percent of total reimbursements), compared to $40 million (2.08 percent of total reimbursements) in the APEC study. The estimate of total erroneous payments generated by the econometric model was $167 million (8.39 percent of total reimbursements), compared to $177 million (9.15 percent of total reimbursements) estimate of total erroneous payments from the APEC study.

D. STRENGTHS AND LIMITATIONS OF THE MODEL

This chapter has summarized the performance of an econometric model that will be used by FNS for tracking erroneous payments in the NSLP and SBP in future years, based on the relationship between district characteristics and erroneous payments as measured in SY 2005-2006. The econometric model consists of a set of 12 regression equations. The dependent variable in each model is a particular district-level certification error rate for either NSLP lunches or SBP breakfasts and is drawn from the APEC study. The explanatory variables in these models are district characteristics that can be measured using FNS-742 data. After estimating the coefficients on these explanatory variables using the sample of districts included in the APEC sample (for whom we could measure the dependent variables), these coefficient values could be used in conjunction with district characteristics from the full FNS-742 dataset to generate imputed values of the 12 certification error rates for all districts nationally. The process of predicting amounts and rates of erroneous payments nationally involves a series of additional transformations of the imputed error rates along with information on the number of meals served in each district.

1. Strengths of the Approach

The district-level model developed to update estimates of erroneous payments in future years has several strengths. One strength of this approach is the high quality of the certification

error rate measures used as dependent variables in the 12 regression equations. Measuring certification error rates is challenging. Our measures are based on the certification status and household circumstances of representative samples of students from the sample districts, which were collected and recorded as part of the APEC Study (Ponza et al. 2007a).

A second strength of this approach is the quality of the model's explanatory variables—the district characteristics used as predictors of certification error rates in a district. These district characteristics include measures—such as variables indicating the results of the district's verification process—that can reasonably be expected to explain variation in the district's certification error rates. In addition, the dataset from which these variables are drawn includes observations on nearly all districts offering school meal programs nationally and will be readily available to FNS in future years in a timely way. Thus, it should be possible to use this model in the future to track trends in erroneous payments.

As part of the testing of the econometric model, we used FNS-742 data from the 2005-2006 school year along with coefficient estimates from the model to predict amounts and rates of erroneous payments nationally for that school year. The original APEC study covered the 2005-2006 school year, so we could compare the predicted erroneous payments amounts and rates from the econometric model with estimates based on the APEC study. The predicted values from the econometric model were close to the study-based estimates in most cases, suggesting that—at least in the study year—the model gave good predictions of amounts and rates of erroneous payments in the NSLP and SBP.

An analogous process can be used to predict erroneous payments in future years, though study-based estimates for those years will not be available for comparison. However, the future

predicted amounts and rates of erroneous payments can be compared against the 2005-2006 predictions to examine trends in these values.[6]

Looking beyond the study, a key aspect of the model's success will be its ability to capture changes in erroneous payments resulting from changes in the way districts conduct the certification process. Changes in the certification process or in districts' administration of the meal programs generally may come about in an effort to improve the integrity of these programs (that is, reduce certification error). Alternatively, they may come about for some other reason but, as a side effect, may influence program integrity. In either case, changes such as these are difficult to measure, especially since they may occur in unexpected ways in future years.

We believe that the explanatory variables included in the model will track future changes in erroneous payments resulting from administrative factors such as these reasonably well. The explanatory variables in the model explained a good proportion of the variation in certification error rates across districts (with R-squared values ranging from 0.13 to 0.39). In addition, they may capture directly or proxy for some of these hard-to-observe aspects of districts' program administration. Some of the explanatory variables are direct measures of these administrative features, such as the use of individual-level or household free and reduced-price meal applications or whether the district selects a random or focused verification sample. Others likely proxy for districts' efforts to improve the integrity of the certification process. In particular, the results of the verification process (for example, what proportion of applications selected for verification does the household respond to the request for documentation) may

[6]The future predicted values should be compared to the predicted values from 2005-2006 rather than the study-based estimates from that year. This is because the study-based estimates will be subject to sampling error since they are based on data from a sample of districts, whereas the predicted values for 2005-2006 and future years will not be subject to that particular form of sampling error since they will be based on information from nearly all districts nationally.

reflect how intensely the district carries out this process; this, in turn, may proxy for the way the district carries out the initial certification process.

2. Limitations of the Econometric Model

Some important limitations of the econometric model and process for predicting erroneous payments should be considered. One key caveat is that the purpose of the econometric model is to predict or track erroneous payments in the future. It is not designed to measure causal relationships between erroneous payments and district characteristics, and the results of the model should not be used as a basis for developing policies aimed at reducing erroneous payments.

Some additional limitations of the model include the following:

- ***Cross-Sectional Model.*** The econometric model is a cross-sectional model for explaining variation in erroneous payments over time. In other words, relationships between districts' characteristics and their error rates at a given point in time are used to infer how error rates will change in the future as the characteristics change. While it is not unusual for a cross-sectional model to be used to make inferences about changes over time, we would ideally like to estimate a model which considered how districts' error rates changed over time as their characteristics changed.
- ***Unexplained Variation.*** As noted above, the explanatory variables in the model explain a reasonably large amount of the variation in districts' error rates. However, there is also a substantial amount of variation that remains unexplained by the model. In other words, there are unobserved factors that cause certification error rates to be higher in some districts than in others. To the extent that changes in these unobserved factors also lead to changes in error rates (and, consequently, erroneous payments) in future years, the model will not capture these changes. However, the potential effects of these unobserved factors on the predicted amounts and rates of erroneous payments are reflected in the standard errors of the predicted values.
- ***Assumption of Stable Relationships Between Error Rates and District Characteristics.*** The strategy of using an econometric model based on estimated relationships between district characteristics and certification error rates in 2005-2006 to predict erroneous payments in the future implicitly assumes that these relationships remain constant over time. For example, if we estimated that districts with higher rates of nonresponse to the verification request had higher error rates in 2005-2006, using this estimated relationship to predict erroneous payments in 2007-2008 assumes that the nonresponse rate still had the same relationship with error rates two years

later. We believe that this is a reasonable assumption in the short run of a few years. However, the further out in the future the 2005-2006 econometric model results are used to predict erroneous payments, the less reasonable this assumption becomes.

- ***Additional Modeling Assumptions.*** In the course of using the results of the econometric model to predict amounts and rates of erroneous payments, the procedure described above makes various additional assumptions about the values of particular district characteristics. These assumptions are likely to be accurate on average, but in any given district the assumed value of a particular variable might differ substantially from its actual value. Three such assumptions are:

  - In assigning each district a number of meals of a particular type, we assumed that the proportion of a state's meals of a given meal price category (for example, free NSLP lunches) served in a given district is the same as the proportion of the state's students in that meal price category who are enrolled in that district.
  - We assumed that the distribution across districts of free and reduced-price breakfasts served in severe needs schools is the same as the distribution across districts of all free meals served in that state.
  - We assumed that the estimate from the APEC study of the relationship between erroneous payments in non-P23 schools and base year P23 schools and erroneous payments in all schools (including non-base year P23 schools) remains constant in future years.

The implication of these limitations is that in any future year, the predicted amounts and rates of erroneous payments will not be as accurate or credible as new estimates of these values from a large-scale nationally representative study similar to APEC would be. However, conducting additional APEC-style studies in the near future would be costly and time-consuming. In addition, we believe that the econometric model described in this report will provide predicted amounts and rates of erroneous payments that are reasonable estimates of their actual values. Moreover, the predicted values will allow FNS to effectively track the direction and general magnitude of changes in erroneous payments in the future, at minimal cost and in a timely manner.

REFERENCES

Office of Analysis, Nutrition, and Evaluation, Food and Nutrition Service, U.S. Department of Agriculture. "Analysis of Verification Summary Data, School Year 2004–2005 (Corrected)." Alexandria, VA: USDA, November 2005. [www.fns.usda.gov/oane/MENU/Published/CNP/FILES/CNVerification.pdf].

Ponza, Michael, Philip Gleason, Lara Hulsey, and Quinn Moore. "NSLP/SBP Access, Participation, Eligibility, and Certification Study, Erroneous Payments in the NSLP and SBP, Volume I: Findings." Final report submitted to the U.S. Department of Agriculture, Food and Nutrition Service, Office of Analysis, Nutrition, and Evaluation. Princeton, NJ: Mathematica Policy Research, Inc., October 2007a. [http://www.fns.usda.gov/oane/MENU/Published/CNP/FILES/apecvol1.pdf]

Ponza, Michael, Philip Gleason, Eric Grau, John Hall, Lara Hulsey, and Quinn Moore. "NSLP/SBP Access, Participation, Eligibility, and Certification Study, Erroneous Payments in the NSLP and SBP, Volume II: Sampling and Data Analysis Appendices." Final report submitted to the U.S. Department of Agriculture, Food and Nutrition Service, Office of Analysis, Nutrition, and Evaluation. Princeton, NJ: Mathematica Policy Research, Inc., October 2007b. [http://www.fns.usda.gov/oane/MENU/Published/CNP/FILES/apecvol2.pdf]

APPENDIX A

ADDITIONAL CHAPTER III TABLES

TABLE A.1

COEFFICIENT ESTIMATES FROM ESTIMATED REGRESSION EQUATIONS

		(1) % CF-NE-L	(2) % CF-RE-L	(3) % CR-NE-L	(4) % CR-FE-L	(5) % NC-FE-L	(6) % NC-RE-L	(7) % CF-NE-B	(8) % CF-RE-B	(9) % CR-NE-B	(10) % CR-FE-B	(11) % NC-FE-B	(12) % NC-RE-B
Verification Variables													
(1)	Selected focused verification sample	-5.392 (4.225)	0.394 (3.912)	-9.047 (13.990)	6.529 (15.379)	-1.580 (3.051)	1.977 (2.709)	-1.939 (2.773)	2.197 (4.858)	-1.088 (21.838)	-4.453 (24.709)	-0.810 (3.996)	-3.027 (6.925)
(2)	Percentage of verified applications that had benefits reduced or terminated during verification[a]	-0.044 (0.043)	0.000 (0.049)	0.554 (0.436)	-0.646 (0.505)	-0.013 (0.022)	0.007 (0.039)	-0.033 (0.067)	-0.026 (0.101)	1.464 (0.713)**	-0.870 (0.761)	0.000 (0.083)	-0.238 (0.179)
(3)	Interaction of (1) and (2)	0.184 (0.080)**	0.027 (0.080)	0.398 (0.832)	0.498 (0.907)	0.007 (0.059)	-0.060 (0.065)	0.162 (0.089)*	0.036 (0.124)	-0.863 (1.012)	1.540 (1.351)	-0.005 (0.116)	0.175 (0.201)
	F-statistic of (2)+(3)	[6.08]**	[0.18]	[1.68]	[0.04]	[0.01]	[1.13]	[5.08]**	[0.03]	[0.61]	[0.39]	[0.00]	[1.15]
(4)	Percentage of verified applications that had benefits increased during verification			0.205 (0.172)	-0.100 (0.185)					0.506 (0.664)	-0.721 (0.472)		
(5)	Interaction of (1) and (4)			-0.192 (0.315)	0.074 (0.400)					-0.820 (0.792)	1.039 (0.637)		
	F-statistic of (4)+(5)			[0.00]	[0.01]					[0.57]	[0.55]		
(6)	Percentage of applications selected for verification in which household did not respond to request for verification	0.063 (0.054)	0.099 (0.052)*	-0.316 (0.218)	0.268 (0.249)	0.013 (0.035)	0.037 (0.039)	0.079 (0.055)	0.093 (0.059)	-0.378 (0.267)	0.207 (0.382)	0.010 (0.052)	-0.043 (0.082)
(7)	Interaction of (1) and (6)	-0.007 (0.073)	-0.045 (0.070)	0.234 (0.245)	-0.069 (0.282)	0.069 (0.066)	-0.050 (0.047)	-0.067 (0.066)	-0.070 (0.082)	0.473 (0.319)	-0.266 (0.395)	0.111 (0.107)	-0.047 (0.078)
	F-statistic of (6)+(7)	[1.46]	[0.91]	[0.31]	[1.27]	[3.01]*	[0.16]	[0.09]	[0.12]	[0.19]	[0.06]	[1.68]	[5.70]**

Table A.1 *(continued)*

		(1) % CF-NE-L	(2) % CF-RE-L	(3) % CR-NE-L	(4) % CR-FE-L	(5) % NC-FE-L	(6) % NC-RE-L	(7) % CF-NE-B	(8) % CF-RE-B	(9) % CR-NE-B	(10) % CR-FE-B	(11) % NC-FE-B	(12) % NC-RE-B
(8)	Indicator that verification results were missing			-3.490 (13.227)	3.999 (12.664)					-9.489 (18.548)	5.686 (14.298)		
Certification Variables													
(9)	Percentage of certified students that were certified non-applicants	-0.006 (0.053)	-0.122 (0.058)**			-0.016 (0.055)	0.022 (0.034)	0.054 (0.041)	-0.036 (0.060)			0.0623 (0.070)	0.017 (0.049)
(10)	Percentage of certified students that were certified categorically	-0.091 (0.091)	-0.168 (0.091)*			-0.066 (0.060)	-0.012 (0.053)	0.015 (0.081)	-0.070 (0.106)			-0.131 (0.094)	0.0154 (0.121)
Core SFA Characteristics													
(11)	Enrollment (in tens of thousands)	0.003 (0.009)	-0.136 (0.068**	-0.451 (0.603)	-0.389 (0.308)	0.111 (0.132)	-0.105 (0.144)	0.010 (0.062)	-0.135 (0.070)*	1.522 (0.962)	0.099 (0.136)	-0.126 (0.427)	0.118 (0.118)
(12)	Percentage of students certified eligible for free meals	-0.067 (0.035)*	-0.026 (0.044)	0.036 (0.184)	0.113 (0.162)	0.072 (0.041)*	0.057 (0.029)*	-0.051 (0.035)	-0.056 (0.054)	-0.061 (0.212)	0.322 (0.230)	0.143 (0.054)**	0.157 (0.076)**
(13)	Percentage of students certified eligible for reduced-price meals	-0.211 (0.291)	0.351 (0.272)	-0.253 (0.993)	-0.719 (1.171)	0.046 (0.139)	0.019 (0.164)	-0.034 (0.180)	0.481 (0.365)	-0.416 (1.318)	-0.974 (1.447)	-0.126 (0.226)	-0.470 (0.342)
Additional FNS-742 Variables													
(14)	SFA privately operated	14.303 (5.104)***		-13.610 (8.314)	22.728 (9.064)**	-0.554 (1.729)			2.735 (3.116)				
(15)	Application is for individual student	-2.699 (2.740)	-1.176 (2.398)	6.139 (13.032)	25.295 (13.410)*	2.965 (4.044)		-1.852 (1.779)	-3.633 (2.029)*	17.519 (21.100)	12.765 (20.646)		-3.275 (3.152)
(16)	Free eligibles not subject to verification (in tens of thousands)	0.075 (0.577)		2.294 (3.102)		-0.985 (0.707)						-3.100 (1.641)*	
(17)	Number of verified applications that were free eligible categorically (in thousands)		0.249 (0.434)				0.051 (0.048)	-0.863 (0.308)***		-0.728 (0.538)		-0.144 (0.139)	

Table A.1 *(continued)*

		(1) % CF-NE-L	(2) % CF-RE-L	(3) % CR-NE-L	(4) % CR-FE-L	(5) % NC-FE-L	(6) % NC-RE-L	(7) % CF-NE-B	(8) % CF-RE-B	(9) % CR-NE-B	(10) % CR-FE-B	(11) % NC-FE-B	(12) % NC-RE-B
(18)	Total number of applications certified (in tens of thousands)		0.186 (0.155)					0.307 (0.126)**		-2.551 (1.811)			-0.235 (0.189)
(19)	Free eligibles approved categorically (in thousands)										-0.198 (0.934)		-0.157 (0.100)
(20)	Number of schools						-0.005 (0.014)					0.039 (0.050)	
(21)	Free eligibles approved by income (in thousands)				0.129 (0.110)				0.044 (0.033)		0.010 (0.343)		
(22)	Number of applications verified (in thousands)						2.122 (0.895)**						
(23)	Constant	11.354 (4.992)**	9.175 (5.040)*	29.770 (13.323)**	18.669 (15.119)	-0.538 (1.782)	-1.069 (1.850)	4.401 (3.458)	4.303 (5.479)	22.691 (19.912)	24.826 (25.772)	-2.055 (4.061)	5.527 (6.942)
	Observations	**86**	**86**	**86**	**86**	**83**	**83**	**74**	**74**	**71**	**71**	**72**	**72**
	R-squared	**0.38**	**0.15**	**0.13**	**0.23**	**0.25**	**0.18**	**0.23**	**0.19**	**0.17**	**0.14**	**0.20**	**0.39**

Source: FNS-742 Verification Summary Reports and APEC study.

Note: Robust standard errors in parentheses. Units in rows (11), (16), (17), (18), (19), (21), and (22) have been changed to the thousands or tens of thousands for presentation purposes; in the actual model these variables are simple counts (i.e. the units are ones).

[a]For columns (5), (6), (11), and (12), this independent variable is the percentage of all verified applications that changed to a different status during verification.

*Significant at 10 percent.
**Significant at 5 percent.
***Significant at 1 percent.

TABLE A.2

COEFFICIENT ESTIMATES FROM ESTIMATED TOBIT REGRESSION EQUATIONS

		(1) % CF-NE-L	(2) % CF-RE-L	(3) % CR-NE-L	(4) % CR-FE-L	(5) % NC-FE-L	(6) % NC-RE-L	(7) % CF-NE-B	(8) % CF-RE-B	(9) % CR-NE-B	(10) % CR-FE-B	(11) % NC-FE-B	(12) % NC-RE-B
Verification Variables													
(1)	Selected focused verification sample	-6.817 (5.165)	0.408 (4.847)	-9.709 (19.037)	4.221 (17.283)	0.897 (3.755)	3.104 (3.540)	-5.094 (4.491)	0.738 (6.167)	-4.984 (33.955)	-5.499 (37.129)	0.086 (8.409)	-2.700 (8.037)
(2)	Proportion of verified applications that had benefits reduced or terminated during verification[a]	-0.048 (0.056)	0.016 (0.056)	0.648 (0.503)	-1.082 (0.711)	0.025 (0.044)	0.023 (0.065)	-0.056 (0.091)	0.028 (0.110)	2.550 (1.028)**	-1.250 (1.158)	-0.185 (2.56)	-0.403 (0.261)
(3)	Interaction of (1) and (2)	0.220 (0.093)**	0.034 (0.096)	0.717 (1.116)	0.745 (1.120)	-0.077 (0.081)	-0.048 (0.090)	0.266 (0.123)**	0.024 (0.144)	-0.915 (1.599)	2.475 (1.824)	0.159 (0.302)	0.374 (0.291)
	F-statistic of (2)+(3)	[6.02]**	[0.39]	[1.83]	[0.15]	[0.55]	[0.17]	[6.33]**	[0.34]	[1.64]	[0.82]	[0.02]	[0.07]
(4)	Proportion of verified applications that had benefits increased during verification			0.264 (0.219)	-0.139 (0.197)					0.696 (0.929)	-0.755 (0.694)		
(5)	Interaction of (1) and (4)			-0.299 (0.423)	0.114 (0.443)					-1.250 (1.094)	1.319 (0.900)		
	F-statistic of (4)+(5)			[0.01]	[0.00]					[0.93]	[0.97]		
(6)	Proportion of applications selected for verification in which household did not respond to request for verification	0.070 (0.063)	0.127 (0.58)**	-0.311 (0.263)	0.311 (0.257)	0.031 (0.046)	0.067 (0.055)	0.102 (0.070)	0.098 (0.073)	-0.567 (0.400)	0.436 (0.478)	0.061 (0.120)	0.027 (0.105)
(7)	Interaction of (1) and (6)	-0.001 (0.090)	-0.057 (0.081)	0.221 (0.305)	-0.060 (0.296)	0.089 (0.078)	-0.072 (0.066)	-0.067 (0.091)	-0.038 (0.100)	0.846 (0.466)*	-0.519 (0.521)	0.137 (0.174)	-0.147 (0.121)
	F-statistic of (6)+(7)	[1.25]	[1.15]	[0.20]	[0.21]	[4.37]**	[0.01]	[0.30]	[0.59]	[1.01]	[0.07]	[2.26]	[3.10]*
(8)	Indicator that verification results were missing			-3.135 (15.631)	8.436 (14.504)					-17.735 (27.460)	7.224 (22.723)		
Certification Variables													
(9)	Proportion of certified students that were certified non-applicants	-0.003 (0.067)	-0.121 (0.062)*			-0.007 (0.061)	0.035 (0.044)	0.063 (0.057)	-0.055 (0.066)			0.022 (0.113)	0.047 (0.085)
(10)	Proportion of certified students that were certified categorically	-0.094 (0.107)	-0.163 (0.095)*			-0.093 (0.083)	0.020 (0.067)	0.049 (0.109)	-0.132 (0.129)			-0.171 (0.172)	0.258 (0.176)

Table A.1 *(continued)*

		(1) % CF-NE-L	(2) % CF-RE-L	(3) % CR-NE-L	(4) % CR-FE-L	(5) % NC-FE-L	(6) % NC-RE-L	(7) % CF-NE-B	(8) % CF-RE-B	(9) % CR-NE-B	(10) % CR-FE-B	(11) % NC-FE-B	(12) % NC-RE-B
Core SFA Characteristics													
(11)	Enrollment (in tens of thousands)	0.062	-0.139	-0.796	-0.296	0.236	-0.283	0.012	-0.517	1.534	0.517	0.257	0.411
		(0.103)	(0.072)*	(0.811)	(0.331)	(0.193)	(0.2267)	(0.070)	(0.236)**	(1.077)	(1.382)	(0.735)	(0.215)*
(12)	Proportion of students certified eligible for free meals	-0.096	-0.027	-0.060	0.082	0.108	0.062	-0.084	-0.060	-0.222	0.368	0.266	0.243
		(0.048)**	(0.048)	(0.228)	(0.181)	(0.048)**	(0.039)	(0.050)*	(0.0627)	(0.283)	(0.306)	(0.094)***	(0.100)**
(13)	Proportion of students certified eligible for reduced-price meals	-0.135	0.376	0.099	-0.647	0.076	0.061	-0.113	0.481	-1.768	-0.751	-0.450	-0.461
		(0.321)	(0.301)	(1.311)	(1.308)	(0.214)	(0.273)	(0.302)	(0.388)	(2.153)	(1.780)	(0.484)	(0.548)
Additional FNS-742 Variables													
(14)	SFA privately operated	14.505		-19.579	23.932	-2.391			4.294				
		(5.355)***		(11.945)	(9.992)**	(3.662)			(3.097)				
(15)	Application is for individual student	-3.176	-1.729	0.030	24.831	2.486		-3.382	-5.466	22.133	5.279		-5.158
		(3.481)	(2.909)	(16.688)	(14.666)*	(3.875)		(3.431)	(3.046)*	(25.699)	(26.379)		(5.457)
(16)	Free eligibles not subject to verification (in tens of thousands)	0.001		4.57		-2.361						-6.294	
		(0.001)		(4.192)		(1.182)**						(2.991)**	
(17)	Number of verified applications that were free eligible categorically (in thousands)		0.319				0.050	-0.846		-0.689		-0.516	
			(0.440)				(0.061)	(0.407)**		(0.609)		(0.364)	
(18)	Total number of applications certified (in tens of thousands)		0.203					0.374		-2.381			-1.178
			(0.152)					(0.159)**		(2.108)			(0.473)**
(19)	Free eligibles approved categorically (in thousands)										-0.461		-0.405
											(0.953)		(0.469)
(20)	Number of schools						-0.003					0.049	
							(0.021)					(0.081)	
(21)	Free eligibles approved by income (in thousands)				0.100				1.271		-0.033		
					(0.125)				(0.553)**		(0.346)		
(22)	Number of applications verified (in thousands)						4.421						
							(2.016)**						
(23)	Constant	10.686	7.242	25.895	17.977	-5.540	-5.634	4.391	4.633	24.118	7.715	-11.264	-4.175
		(5.147)**	(5.619)	(17.520)	(15.934)	(3.428)	(3.514)	(5.011)	(6.410)	(33.081)	(36.070)	(9.175)	(9.221)
	Observations	**86**	**86**	**86**	**86**	**83**	**83**	**74**	**74**	**71**	**71**	**72**	**72**

Table A.1 *(continued)*

Source: FNS-742 Verification Summary Reports and APEC study.

Note: Robust standard errors in parentheses. Units in rows (11), (16), (17), (18), (19), (21), and (22) have been changed to the thousands or tens of thousands for presentation purposes; in the actual model these variables are simple counts (i.e. the units are ones).

[a]For columns (5), (6), (11), and (12), this independent variable is the Proportion of all verified applications that changed to a different status during verification.

*Significant at 10 percent.
**Significant at 5 percent.
***Significant at 1 percent.

TABLE A.3

COMPARISON OF NATIONAL ESTIMATES OF ERRONEOUS PAYMENTS BASED ON APEC STUDY AND ON TOBIT IMPUTATION MODEL

	APEC Study		Tobit Imputation Model	
	Erroneous Payments (in Dollars)	Percentage of Reimbursement in Error	Erroneous Payments (in Dollars)	Percentage of Reimbursement in Error
NSLP				
Overpayments	573 (50) [491-655]	7.11 (0.62) [6.09-8.13]	556	6.94
Underpayments	186 (15) [161-211]	2.31 (0.19) [2.00-2.62]	165	2.06
Total Erroneous Payments	759 (54) [670-848]	9.42 (0.67) [8.32-10.52]	721	9.00
SBP				
Overpayments	137 (17) [109-165]	7.07 (0.91) [5.58-8.56]	128	6.42
Underpayments	40 (6) [30-50]	2.08 (0.29) [1.60-2.56]	25	1.24
Total Erroneous Payments	177 (18) [147-207]	9.15 (0.94) [7.61-10.69]	152	7.66

Source: FNS-742 Verification Summary Reports and APEC study.

Note: Standard errors in parentheses; 90 percent confidence intervals in brackets.

APPENDIX B

SUMMARY OF EXTERNAL REVIEWER COMMENTS

A draft version of the econometric model and report were reviewed by staff at the Food and Nutrition Service's (FNS) Office of Analysis, Nutrition, and Evaluation and Child Nutrition Division and by two external reviewers. Both external reviewers acknowledged the difficulty of the modeling objective and were satisfied that the model and the methodology we developed achieved this objective. The reviewers provided a set of comments outlining potential revisions to the model. In this appendix, we summarize the major substantive comments of the external reviewers. We describe each issue and how we addressed it in the final version of the model and report.

1. Data Issues

Potential for Measurement Error in APEC Study Household Survey Data. One reviewer noted that income measured with survey data is typically measured with error, and indicated that the report should at a minimum discuss this potential problem and implications on our approach for updating estimates of erroneous payments. Two additional analyses were suggested. First, link the APEC Study household data to some administrative data on income (e.g., Social Security Records) for at least a subsample to determine the extent of measurement error. Second, move the "cutoff levels" for income-eligibility for free and reduced-price meal benefits around to see how sensitive the predictions are to potential errors.

Response: The APEC study implemented several approaches that were expected to minimize problems associated with measuring income using a household survey that should result in more accurate household income measures, but does not eliminate all measurement error. For example, the study used computerized-assisted personal interviewing, requested documentation of income sources and amounts and reconciled respondent reported sources and amounts against documented ones; the reference month for the interview coincided with the

month covered by the application and the interview was conducted within one or two months of this reference month. Unfortunately the study did not have resources nor was it feasible to link APEC household survey data to Social Security Records or other administrative data sources to directly examine the extent of measurement error in our sample. However, to get some sense of the effects that measurement error might have on results, for the main study we did examine the sensitivity of certification error estimates to changes in the meal benefit eligibility thresholds. For example, we found that increasing the thresholds for free and reduced-price eligibility both by 5 percentage points reduced certification error by 6.5 percent; increasing the thresholds by 10 percentage points, reduced the certification error rate by 15 percent. As we note in Chapter II of the final report, ultimately, there is not much we can do regarding measurement error in our household survey reports of income sources and amounts but acknowledge that the potential for error exists, since the best we can do with the modeling effort is to replicate the APEC study estimates (which are subject to measurement error themselves).

Missing Data on Districts in the FNS-742 Data File. One reviewer noted that the district-level FNS-742 data used for future predictions is missing some districts, and that some effort should be made to assess the characteristics of districts with missing FNS-742 data before simply assuming "the rates of erroneous payments in these districts are similar to rates in districts that are represented in the data," as we had done in the draft report.

Response: We were able to compare districts that did not have FNS-742 data to those that did on the basis of one important characteristic: numbers and percentages of students in each school meal certification category. Student counts for districts not represented in the FNS-742 were generated by subtracting totals of students by certification status in the FNS-742 from national totals available from the FNS national database, the Common Core Data, and the Private School Survey. We found similar percentages of students by certification status in the two

groups of districts. In FNS-742 districts, 34 percent of students were certified for free meals, 8 percent were certified for reduced price meals, and 59 percent were not certified for school meal benefits. In districts not represented in the FNS-742, these percentages were 35 percent, 7 percent, and 58 percent, respectively. Although it is possible that these two types of districts differ in other important ways, it is reassuring that they are similar in this one important dimension.

A more thorough assessment of this issue would require examining several district characteristics. For example, one approach would be to try to match FNS-742 districts to districts in the Common Core Data (CCD), and contrast characteristics of the districts included and missing in FNS-742. The difficulties inherent in matching districts was a major reason for focusing on FNS-742 data and not the other data sets (such as CCD or state agency data). The problem is that there is not a standard ID for matching between these data sets. While we were able to match our APEC study sample of approximately 90 districts to these other data sources in the model development, it was not straightforward and required considerable manual processing of the data. We did not have the resources to do this for thousands of districts.

In addition to the similarity the breakout of students by meal certification category in districts that did not have FNS-742 data compared to those that do, we believe that the missing data problem in FNS-742 is not a major issue for three additional reasons: (1) the predicted amounts and rates of erroneous payments for the study year using the districts in FNS-742 are quite close to those from the APEC study for SY 2005-06, which is based on nationally representative data; (2) The SFAs in the SY 2005 – 06 FNS-742 data file represent high percentages of the child nutrition state agencies, SFAs, and the total enrolled students in schools operating the NSLP and/or SBP nationwide that year; and (3) we expect overtime that district

response to FNS-742 will improve, as districts become more familiar with the reporting process and with improvements in the reporting process (e.g., online reporting).

2. Model Fitting Issues

Both external reviewers were satisfied with the overall model and methodology. They concluded that the approach for arriving at the final specification is justifiable and has good predictive power. They both felt reliance exclusively on FNS-742 data was reasonable.

The District-Level Model Should be Estimated Using Unweighted Data. Initially we estimated the district-level error rate specifications using weighted least squares to account for the multistage, clustered sample design. One reviewer correctly noted that the goal here is prediction not causal modeling. And while the school and student sample weights are needed to produce unbiased district-level estimates, it was recommended that we ignore the district selection probabilities in the actual regressions (except, perhaps, as an additional explanatory variable).

Response: We implemented the reviewer's suggestion for the final estimation approach. We estimated the six NSLP and six SBP equations using ordinary least squares (OLS) estimation techniques on unweighted data. We adjusted for the heteroskedasticity by estimating robust standard errors of the coefficients in the regression equations.

Conduct Model Testing on a Random Subsample of APEC Study Districts. Typically in prediction models, one would use a "set aside" subsample, randomly chosen from the full data set, to measure model performance. This prevents over-fitting within the study sample. Although recognizing the small number of districts in the APEC Study sample as a constraint, one reviewer suggested one possibility would be to split the sample (say 60 - 40 or 70 - 30) and use the larger sample to estimate the primary models and assess their performance. Then, using

those models predict the dependent variable in the set aside sample and use measures of goodness of fit to compare the models. The reviewer noted that in this situation, the authors should only use this as a double check—relying instead primarily on the approaches already used by the authors to develop and test the model specification.

Response: In our planning, we considered the split sample approach but decided against it. Given the small number of districts in the APEC Study sample, we concluded that reducing the sample size further with a split sample likely would lead to unreliable estimates.

Specification Sensitivity Tests: Approaches Other Than Tobit. We tested the sensitivity of the erroneous payments estimation results to various alternative specifications of the model. We assessed the assumption of a linear specification by estimating a Tobit model. The Tobit model accounts for the fact that the dependent variable (certification error rates) cannot take on values less than zero or greater than one. OLS estimation potentially yields estimates that are biased if a large proportion of the values of the dependent value are equal to zero or one. One reviewer suggested that rather than a Tobit model, which depends too much on distributional assumptions, it would be preferable to have compared modeling the error rates directly with modeling a logistic transform of the error rates (i.e., $g = \log(p) - \log(1-p)$, where p is a district-level error rate). Back-transforming the fitted logits creates biased estimates, so a second regression would be necessary (of p against the fitted logit).

Response: We used Tobit as a robustness check against the basic OLS specification. The results were not different. While the reviewer's point that a drawback of the Tobit model is that it makes strong assumptions about the distribution of the regression error term is correct, we did not reestimate the model using the technique that the reviewer recommended because we use this specification as a robustness check only and could not justify the additional resources required to implement it.

3. Derivation of Standard Errors

Estimating Standard Errors of the Predicted Amounts and Rates of Erroneous Payments. Our initial model included estimation of standard errors of our updated estimates of amounts and rates of erroneous payments in order to establish confidence intervals around these estimates. Specifically, for each district we estimated a variance for each of the six types of error in the NSLP and SBP. We then combined these estimates to get the variance of district-level overpayments, underpayments, and total erroneous payments. Finally, we combined the district-level estimates to generate the variance of national estimates of erroneous payments. One reviewer questioned some of the assumptions in our approach to estimating standard errors. Moreover, both reviewers noted that estimating standard errors of the regression model as well as the final predicted amounts and rates of erroneous payments are difficult and are probably not worth doing.

Response: We concurred with the recommendations of the reviewers to not estimate standard errors of the final predicted amounts and rates of erroneous payments.