United States General Accounting Office

GAO

Report to Congressional Committees

March 2004

SMALL BUSINESS ADMINISTRATION

Model for 7(a) Program Subsidy Had Reasonable Equations, but Inadequate Documentation Hampered External Reviews

G A O

Accountability ★ Integrity ★ Reliability

GAO-04-9

March 2004

SMALL BUSINESS ADMINISTRATION

Model for 7(a) Program Subsidy Had Reasonable Equations, but Inadequate Documentation Hampered External Reviews

GAO Accountability • Integrity • Reliability

Highlights

Highlights of GAO-04-9, a report to Chairman and Ranking Minority Member, House Committee on Small Business; Ranking Minority Member, Senate Committee on Small Business and Entrepreneurship

Why GAO Did This Study

The Small Business Administration (SBA) approved about $8.6 billion in loan guarantees through its 7(a) loan program in fiscal year 2003. SBA must estimate the subsidy cost of this program. Since fiscal year 2003, SBA has been using econometric modeling to estimate the subsidy. This report reviews SBA's estimation methodology and equations, assesses the default and recovery rates the model produced, identifies ways to enhance the estimates' reliability, describes the process for developing the model, and analyzes SBA's data.

What GAO Recommends

SBA should (1) determine whether to include in the model other information from its new loan monitoring system, (2) periodically evaluate and update the model, and (3) document the model development process. OMB should require agencies to document the basis and process for developing their credit subsidy models.

SBA agreed with recommendations to improve the final model but SBA and OMB disagreed that the model development was inadequately documented and disagreed with our recommendations to improve such documentation and guidance.

However, given the difficulty experienced by reviewers due to inadequate documentation, we continue to recommend that SBA document the basis and process for developing its model and that OMB

www.gao.gov/cgi-bin/getrpt?GAO-04-9.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Davi D'Agostino at (202) 512-8678 or dagostinod@gao.gov.

What GAO Found

From an economics perspective, SBA's econometric equations were reasonable, and its model produced estimated default and recovery rates that were in line with historical experience. However, from an audit perspective, SBA's lack of documentation of the model development process precluded GAO, and others, from independently evaluating the model's development and determining if SBA used a sound and consistently applied method to select and reject model variables.

Taking into account economic reasoning and research, SBA's econometric equations for estimating defaults, prepayments, and recoveries were reasonable. SBA's equations used a limited set of variables; equations using other variables could also be reasonable but would produce different estimates. Since an estimate is an approximation, no one estimate can be considered accurate, and reasonable estimates can fall within a range of values. The model's estimated default and recovery rates were in line with recent historical experience. SBA could improve its estimation methodology by periodically checking for and correcting errors and should consider adding more borrower information, such as credit scores. Some errors in the model resulted in understating the estimated program costs.

SBA used the expertise of other agencies and a contractor to develop its model and worked closely with the Office of Management and Budget (OMB), which must approve the methodology agencies use to estimate subsidies. OMB officially approved the model in the fall of 2002.

SBA did not adequately document its model development process, including alternative variables considered and rejected, to enable external reviewers to assess the process that was used. Further, GAO and two other independent reviewers could not determine whether a bias existed in the model by systematically excluding variables to influence the subsidy rate in a particular direction. Adequate documentation, a key internal control, would enable SBA and other agencies to demonstrate the rationale and basis for key aspects of the model that provide important cost information for budgets, financial statements, and congressional decision makers and facilitate SBA's annual financial statement audit. Current OMB and other guidance is either silent or unclear about the level of documentation necessary for credit subsidy model development.

SBA had a process to help ensure data integrity and data consistency in the equations with the loan-level data in its databases. Although errors existed in SBA's data systems, the magnitude and nature of these errors were not likely to significantly affect the subsidy rate.

Contents

Letter

1

Results in Brief 5
Background 8
SBA's Equations Were Reasonable and Estimated Default, and Recovery Rates Were in Line with Historical Experience 11
SBA's Model Could Be Enhanced by Adding Information on Borrowers, Correcting Errors, and Updating Some Data 19
SBA Collaborated with OFHEO and OMB to Develop the Model 22
Lack of Adequate Model Documentation Hampered Independent Reviews of SBA's Model 23
SBA Had a Process to Help Ensure Data Quality and the Data Used in the Model and SBA's Loan Level Databases Were Consistent 29
Conclusions 32
Recommendations for Executive Action 33
Agency Comments and Our Evaluation 34

Appendixes

Appendix I: Objectives, Scope, and Methodology 40
Assessing the Reasonableness of the Model's Econometric Equations and Evaluating the Model's Estimated Default, Prepayment, and Recovery Rates 40
Identifying Additional Steps SBA Could take to Further Enhance the Reliability of the Model 40
Reviewing SBA's Process of Developing the Subsidy Model 41
Evaluating the Model's Supporting Documentation, Including Its Discussion of What Variables Were Tested and Rejected 41
Determining What Steps SBA Took to Ensure the Integrity of the Data Used in the Model and Whether These Data Were Consistent with Information in Its Databases 42

Appendix II: Analysis of Default, Prepayment, and Recoveries Econometric Equations 43
SBA's Default and Prepayment Equations 43
Effects of Including Additional Variables 48
SBA's Recovery Equation 54

Appendix III: Comments from the Small Business Administration 58
GAO Comments 63

Appendix IV: Comments from the Office of Management and Budget 65
GAO Comments 68

Appendix V: GAO Contacts and Staff Acknowledgments 70
GAO Contacts 70
Staff Acknowledgments 70

Tables

Table 1: Variable Names and Descriptions 45
Table 2: Multinomial Logistic Regression Coefficient Estimatesa 47
Table 3: Names and Descriptions of Additional Variables 49
Table 4: Multinomial Logistic Regression Coefficient Estimatesa 51
Table 5: Distribution of SIC Industry Codes in SBA's Loan Database Distribution of SIC Industry Codes in SBAs Loan Database 54
Table 6: Variable Names and Descriptions 55
Table 7: Recovery Model 56

Figures

Figure 1: Major Segments of the Model to Estimate 7(a) Subsidy Rate 3
Figure 2: Estimated Default Rates Compared with Average Default Experience from 1992 through 2001 17
Figure 3: Estimated Default Rates Compared with Fiscal Year 2001 Actual Default Experience 18

Abbreviations

CFO	Chief Financial Officer
FCRA	Federal Credit Reform Act of 1990
GDP	gross domestic product
NAIC	North American Industrial Classification
OFHEO	Office of Federal Housing Enterprise Oversight
OMB	Office of Management and Budget
SBA	Small Business Administration
SIC	Standard Industrial Classification

This is a work of the U.S. government and is not subject to copyright protection in the United States. It may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.

G A O
Accountability * Integrity * Reliability

United States General Accounting Office
Washington, D.C. 20548

March 31, 2004

The Honorable Donald A. Manzullo
Chairman, Committee on Small Business
House of Representatives

The Honorable Nydia M. Velazquez
Ranking Minority Member
Committee on Small Business
House of Representatives

The Honorable John F. Kerry
Ranking Minority Member
Committee on Small Business and Entrepreneurship
United States Senate

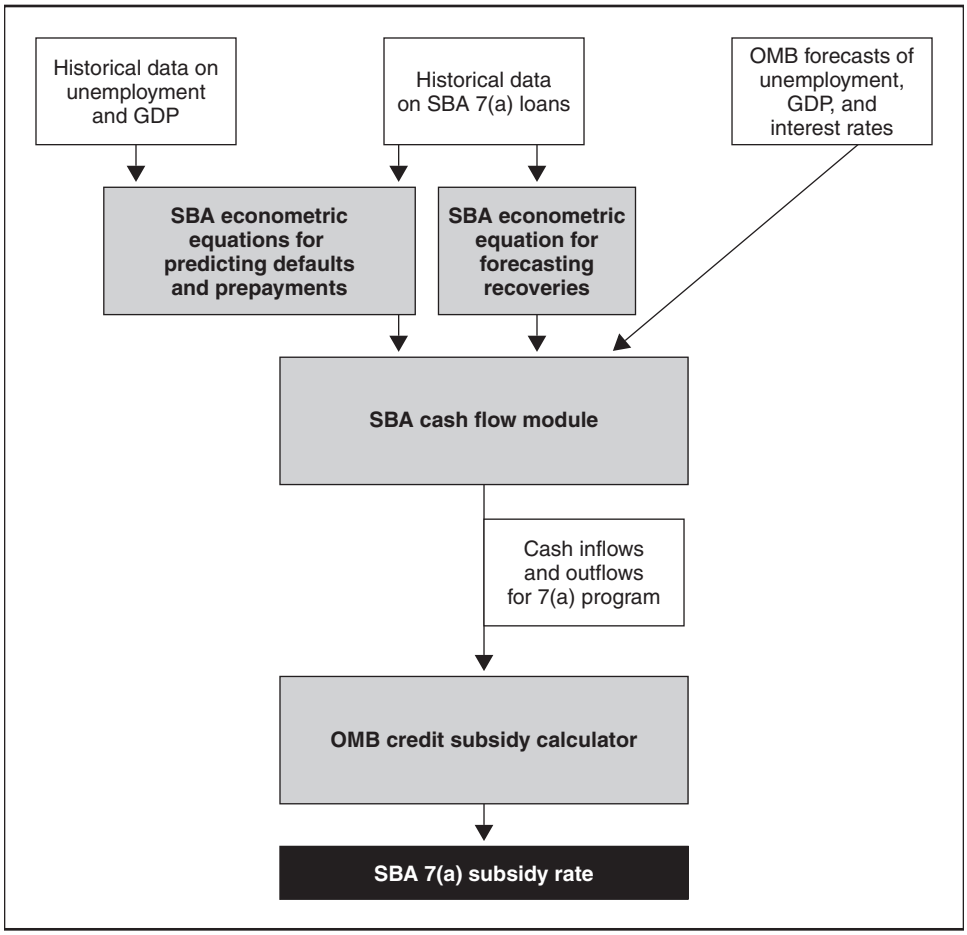
The 7(a) program is the Small Business Administration's (SBA) largest lending program for small businesses. SBA reported that it approved about $8.6 billion in loan guarantees in fiscal year 2003. The program provides loan guarantees of up to 85 percent for loans made to small businesses that are unable to obtain financing on reasonable terms in the private credit markets. Like most federal loan or loan guarantee programs, SBA's 7(a) program is subject to the Federal Credit Reform Act of 1990 (FCRA). FCRA requires most agencies with government lending programs to estimate annually the cost to the federal government of extending or guaranteeing credit over the life of the loans (the subsidy cost). Since an estimate is an approximation, no one estimate can be considered accurate with certainty, and reasonable estimates can fall within a range of values. Changes in estimation methodologies, variables, or data used to calculate an estimate, are likely to result in differences in the estimate. In fiscal year 2003, SBA implemented a new methodology to estimate the subsidy cost of the 7(a) program that is based on econometric modeling.[1] SBA officials told us that the new 7(a) model was the first step in a long-term effort to develop and implement new econometric models for their credit programs. Although this allowed SBA to build a model that responds to the need for greater sensitivity to a wider variety of factors than a model based on historical

[1]Econometric modeling is a series of techniques used to quantify relationships among a group of variables and is often used to forecast the value of economic variables such as loan defaults.

averages, SBA believes that this approach may not be appropriate for all its credit programs.

In order to calculate the subsidy cost of their programs, agencies must estimate the present value of future cash flows over the life of the program, which for the 7(a) program are principally affected by defaulted loans, prepayments of outstanding loans, recoveries on defaulted loans, and fees. The revised method SBA adopted for the subsidy calculation has four segments: (1) the econometric equations that are used to estimate the likelihood of defaults and prepayments, (2) the equations used to estimate the extent of recoveries, (3) the cash flow module, and (4) the Office of Management and Budget (OMB) Credit Subsidy Calculator, as shown in figure 1. The results of the first and second segments—the econometric equations—are a key input into the third. The third segment—the cash flow module—uses these results, along with OMB forecasts of interest rates, unemployment rates, and gross domestic product growth rates to estimate cash inflows from fees and recoveries on defaulted loans and outflows from claim payments on defaulted loans. The resulting cash flows are entered into the fourth segment, the OMB Credit Subsidy Calculator, which calculates the (1) present values of the cash flows and (2) the subsidy rate.

Figure 1: Major Segments of the Model to Estimate 7(a) Subsidy Rate



Source: GAO.

This report responds to your November 26, 2002, and December 11, 2002, requests that we review the methodology that SBA developed to estimate the subsidy costs of its 7(a) loan program for the fiscal year 2004 budget. As agreed with your staff, we (1) assessed the reasonableness of the model's econometric equations and evaluated the model's estimated default and recovery rates based on the 7(a) program's recent historical loan experience; (2) identified any additional steps SBA could take to further enhance the reliability of its subsidy estimate produced by the model; (3) described SBA's process for developing the subsidy model; (4) evaluated the model's supporting documentation including its discussion of what

variables were tested and rejected; and (5) determined what steps SBA takes to ensure the integrity of the data used in the model and determined whether these data are consistent with information in SBA's databases. We did not, however, validate SBA's model.

First, to analyze the model, we obtained from SBA copies of the model as approved by OMB in 2002, along with the loan-level data that were used to develop the subsidy estimates. We analyzed the econometric equations to determine whether they were reasonable based on the variables they included, the statistical techniques used, and the results obtained. For example, we determined whether the econometric equations included appropriate variables and whether the variables used in the equations were statistically significant. To evaluate the model's estimated default and recovery rates, we compared these rates with recent historical loan experience of the 7(a) program provided by SBA. Using SBA's data, we also calculated what SBA would have estimated for default and recovery rates based on the estimation methodology it used prior to its fiscal year 2003 budget submission. Second, to identify any additional steps SBA could take to enhance the reliability of its model, we considered additional types of data that SBA might collect and consider including in its econometric equations. As part of this analysis, we reviewed the academic literature on default modeling and interviewed officials with several banks engaged in similar efforts. Third, to describe SBA's process for developing the model we met with SBA and OMB officials. Fourth, to evaluate the model's supporting documentation, including its discussion of what variables were tested and rejected, we obtained and analyzed available relevant documents and met with SBA officials and their contractor who developed the model. We compared the information presented in SBA's model documentation with existing credit subsidy guidance. Finally, to determine what steps SBA took to ensure the integrity of the data used by the model and to determine whether these data were consistent with information in its databases, we assessed SBA's processes for ensuring data reliability. We examined the type and level of errors and evaluated the likelihood that they would significantly affect the credit subsidy estimates. We also compared the loan-level data used in the model with the data contained in SBA's databases. Appendix I discusses the details of our methodology.

We conducted our work in Washington, D.C. from December 2002 to March 2004 in accordance with generally accepted government auditing standards.

Results in Brief

Overall, we found that from an economics perspective, SBA's econometric equations were reasonable, and the SBA model produced estimated default and recovery rates that were in line with historical experience. However, from an audit perspective, SBA's lack of adequate documentation of the model development process precluded us from (1) independently evaluating the model's development; (2) determining whether SBA used a sound and consistently applied method to select and reject variables to be included in the model; and (3) determining whether a bias from selecting variables existed in the model.

We found that SBA's econometric equations for estimating defaults, prepayments, and recoveries were reasonable. SBA's equations used a limited set of variables; equations using other variables could also be reasonable but would produce different estimates. We also found that the model's estimated default and recovery rates were in line with recent historical experience. SBA's econometric equations related the likelihood of defaults and/or prepayments to several variables that economic reasoning and prior research suggested were appropriate to this type of model, and, at the time of our review, SBA used appropriate statistical techniques to identify the nature of these relationships. In addition, SBA's equations produced estimated relationships for defaults and prepayments that were consistent with expectations based on economic reasoning. For example, the likelihood of default was estimated to be higher when unemployment was higher. SBA's equations used a limited set of variables, and we found that equations using additional variables available to SBA that it did not include, such as measures of interest rates and the businesses' industry type, would also be reasonable. If SBA had used these alternative equations, it might have estimated a higher or lower subsidy rate. SBA did not include any economic variables in its equation for estimating recoveries, so that forecasted recovery amounts were not dependent on expected economic conditions. According to documentation provided by SBA of the work done to develop this equation, adding economic variables would not have increased the precision of the recovery rate estimates.

SBA could enhance the model and the reliability of the subsidy estimate produced by the model by including additional information that SBA expects to have in the future and by correcting errors. SBA intends to collect new business and business-owner information to determine how it affects loan performance and such information may suggest variables that can be useful in the model. SBA's econometric equations used variables

from its current databases and economic indicators, such as gross domestic product (GDP) growth rates and unemployment rates, to forecast future defaults and prepayments. However, at the time of our review, SBA's current database did not include other information on businesses or business owners, such as information on borrowers' credit that is often used by private sector lenders to determine potential defaults and losses. Academic literature on default models suggests that such information is predictive of defaults. SBA has recently contracted to develop a loan monitoring system that is intended to track this information and allow the agency to determine how it affects loan performance. During our review of the model, we identified some errors that resulted in underestimates of the program costs of around $6.5 million or about 6.8 percent of the estimated cost of the program for fiscal year 2004.

To develop its subsidy model, SBA drew on the expertise of other government agencies and consulted with OMB officials. In February 2002, SBA entered into an arrangement with the Office of Federal Housing Enterprise Oversight (OFHEO), which has staff with expertise in econometric modeling, to assist in the development of the 7(a) subsidy model.[2] OMB also played a key role in the development of the model because FCRA requires OMB to approve the methodology that each federal agency uses to estimate the subsidy costs of its loan programs. Thus, SBA consulted with OMB officials during the model's development, and OMB officially approved the model in the fall of 2002. OMB officials said that their role in reviewing the model was primarily to provide oversight and ensure compliance with the law. Because at the time of our review, SBA routinely had its cash flow models reviewed by an independent third party, it hired an outside consultant to conduct limited reviews of the econometric equations and cash flow segment. The consultant identified some errors that SBA corrected.

SBA did not prepare adequate supporting documentation to enable us and other independent reviewers to understand and evaluate the process that SBA used to develop the model. While SBA provided some general documentation of its model development process, the documentation lacked adequate discussion of alternative variables or combinations of

[2]OFHEO was established as an independent entity within the Department of Housing and Urban Development. OFHEO's primary mission is ensuring the capital adequacy and financial safety and soundness of two government-sponsored enterprises—the Federal National Mortgage Association and the Federal Home Loan Mortgage Corporation.

variables that SBA considered, which variables were rejected for which reasons, and specific examples based on results of earlier regressions. As a result, we were unable to determine whether a bias in selecting variables existed in the model. SBA officials told us that they did not prepare this type of documentation because they believed that there was no specific requirement to do so. Current guidance is either silent or unclear about supporting documentation needed to explain the development of econometric models used to generate credit subsidy estimates for the budget and financial statements. However, maintaining adequate documentation on how such models were developed is a sound internal control practice that would provide SBA and other agencies the opportunity to more fully demonstrate and explain the rationale and basis for key aspects of their models that provide important cost information for budgets, financial statements, and congressional decision makers. This documentation would also help facilitate SBA's annual financial statement audit.

SBA hired a private contractor to reconcile the information submitted to it by 7(a) program lenders with the data stored in SBA's loan-level databases on a monthly basis and, at the time of our review, had an ongoing process to correct any errors that were found. Although errors existed in SBA's data systems at the time of our review, we determined that the magnitude and nature of these errors were not likely to significantly affect the subsidy rate. In addition, SBA officials told us that they performed various ad hoc reviews of the information in SBA's loan-level databases to assess its accuracy and were currently assessing various alternatives to further enhance its data integrity. On the basis of our analysis of a statistical sample of defaulted, prepaid, and active loans, as well as recoveries from defaulted loans, we found that the data SBA used to calculate the subsidy costs were consistent with the loan level data contained in SBA's actual databases at the time of our review.

This report contains three recommendations to SBA and one recommendation to OMB. We recommend that SBA (1) determine how best to include in the model borrower-specific information that it intends to collect in its new loan monitoring system; (2) establish a process for periodically revising the model to correct errors and to reflect any changes in the 7(a) program or other factors that could affect the subsidy estimate; and (3) prepare adequate documentation of the model development process including a detailed discussion of alternative variables or combinations of variables that were considered, tested, and rejected and criteria for doing so. We also recommend that OMB require that agencies

document the basis for credit subsidy estimates and reestimates, including the process followed for selecting model methodologies over alternatives and variables tested and rejected with the basis for excluding them.

We received comments on a preliminary draft of this report from SBA and OMB. SBA agreed with the findings and the first two recommendations related to the final model. OMB had no comments. While a draft of this report was at the agencies for comment, we continued to pursue additional documentation that SBA had that might further explain its 7(a) model development process, including what variables were selected and rejected and why. This final report discusses the lack of adequate documentation and recommends improvements in SBA's documentation of the development process for its credit subsidy models and in OMB's Circular A-11 guidance. SBA generally disagreed with our findings and recommendations related to the lack of adequate documentation supporting the model's development process. OMB disagreed with our recommendation that it revise Circular A-11. However, in light of the consistent difficulty experienced by three independent reviewers of SBA's 7(a) credit subsidy model, including SBA's financial statement auditors, we continue to recommend that SBA enhance its credit subsidy model documentation and that OMB require agencies to document the basis and process used to develop credit subsidy models, including understanding the model's basis and the variables that were selected and rejected.

Background

FCRA was enacted, among other reasons, to provide more accurate measures of the costs of federal loan programs and to more accurately compare costs among credit programs and between credit and noncredit programs. FCRA requires agencies with loan guarantee programs to estimate the subsidy cost, or the cost to the government, of their loan guarantees over the life of the loan. To calculate the subsidy costs, agencies must calculate, on a cohort[3] basis, the net present value of the forecasted cash flows for the program, which for SBA included estimated defaults, recoveries, and fees related to the 7(a) program. In addition, as part of this process, SBA must determine the effects of loan prepayments on the cash flows. Under FCRA, SBA provides information that generates a single subsidy rate and does not provide information about any uncertainty in its

[3]A cohort includes those direct loans or loan guarantees of a program for which a subsidy appropriation is provided in a given fiscal year even if the loans are not disbursed until subsequent years.

estimate of the rate or other factors affecting the rate, such as prepayments or defaults.

Prior to its 2003 budget submission, SBA's methodology for estimating the subsidy on its 7(a) loans used historical averages for defaults and recoveries based on loan data going back to 1986 as the basis for estimates of future defaults and recoveries. This approach resulted in fairly stable subsidy estimates on a yearly basis as it included a sufficient volume of historical information that smoothed out fluctuations in economic conditions from year to year. However, this approach resulted in SBA consistently overestimating defaults and recoveries. In previous work, we found that SBA overestimated defaults by about $2 billion from fiscal years 1992 to 2000.[4]

In an effort to improve the accuracy of its subsidy estimate, SBA implemented a new methodology based on econometric modeling to estimate the subsidy cost for the fiscal year 2003 and 2004 budget submissions. Econometric modeling has advantages over historical averaging. For example, to the extent that data are available, it can take into account the effects of changes in such factors as economic conditions, program rules, and loan types on defaults and prepayments.

All forecasts are uncertain, and this uncertainty has multiple causes. When relationships among economic variables are estimated, uncertainty may arise from the choice of variables used in the model, from the degree of precision with which the strength of the relationships is estimated, and from uncertainty about the future values of the independent variables used in the forecasting equation. Excluding a variable that should be in a forecasting model can reduce the quality of the model. For example, if some industries have high default rates, then excluding industry variables will tend to underestimate default costs in years when many loans go to high risk industries and overstate default costs in years when many loans go to low risk industries. The choice of variables to be used in a model results from a process of professional judgment and balancing the risks of including too many or too few variables. Economic theory and statistical tests play an important role in these decisions. The remaining sources of uncertainty, the precision of the estimated relationships and uncertainty about future values of independent variables, are often beyond the control

[4]U.S. General Accounting Office, *SBA's 7(a) Credit Subsidy Estimates*, GAO-01-1095R (Washington, D.C.: Aug. 21, 2001).

of those building the model. The precision of the effects of the independent variables is determined largely by the amount of data available to the analyst, and uncertainty about future values of independent variables is inherent in any forecast.

Internal control is a major part of managing an organization and this includes controls over data gathering and processing, such as SBA's data on 7(a) loans. As mandated by the Federal Managers' Financial Integrity Act of 1982, the Comptroller General issues standards for internal control in the federal government.[5] These standards provide the overall framework for establishing and maintaining internal control and for identifying and addressing major performance and management challenges and areas at greatest risk of fraud, waste, abuse, and mismanagement. According to these standards, internal control comprises the plans, methods, and procedures used to meet missions, goals, and objectives. Control activities are the policies, procedures, techniques, and mechanisms that enforce management's directives and help ensure that actions are taken to address risks. Control activities are an integral part of an entity's planning, implementing, reviewing, and accounting for government resources and achieving effective results. They include a wide range of diverse activities including controls over information processing. These controls are established to ensure that all data inputs are received, are valid, and outputs are correct. Agency management should design and implement internal control based on the related costs and benefits. No matter how well designed and operated, internal control cannot provide absolute assurance that all agency objectives will be met and, thus, once in place, internal control provides reasonable, not absolute, assurance of meeting an agency's objectives.

[5]U.S. General Accounting Office, *Internal Control: Standards for Internal Control in the Federal Government*, GAO/AIMD-00-21.3.1 (Washington, D.C.: November 1999).

SBA's Equations Were Reasonable and Estimated Default, and Recovery Rates Were in Line with Historical Experience

We found that the econometric equations that SBA used to estimate defaults, prepayments, and recoveries were reasonable, although other equations could also be reasonable. SBA uses an appropriate statistical technique for identifying the nature of these relationships. In addition, SBA's equations produced estimated relationships for defaults and prepayments that were consistent with expectations based on economic reasoning. We found that there were additional variables available to SBA that it did not include in its equations, such as measures of interest rates and the borrower's industry type that would also be reasonable and would produce different subsidy rates. In addition, SBA did not include any economic variables in its equation for estimating recoveries. According to documentation provided by SBA to estimate recoveries on defaulted loans, adding economic variables would not have increased the precision of the recovery rate estimates. Finally, we found that the new model's estimated default and recovery rates were in line with recent historical experience.

Variables in SBA's Default and Prepayment Equations Were Appropriate

The econometric equations that SBA used at the time of our review related the likelihood that a borrower would either default on or prepay a loan to several variables that economic reasoning and prior research suggested were appropriate to include in these types of equations. These variables included: (1) characteristics of the borrower's business, such as whether it was a sole proprietorship, partnership, or corporation; (2) characteristics of the loan, such as the amount borrowed; and (3) two measures of economic conditions, the unemployment rate in the state where the loan was made and the GDP growth rate. Economic reasoning and prior research suggested that differences in borrower and loan characteristics and economic conditions were likely to influence defaults and prepayments. For example, prior research suggested that new businesses were less likely to survive than were established businesses and thus were more likely to default.[6] Prior research also suggested that the likelihood of default on loans made to partnerships or corporations should be less than it was for loans made to sole proprietors, while the likelihood of prepayment should be greater. Details about SBA's econometric equations are found in appendix II.

[6]See, for example, Brian Headd, "Business Success: Factors Leading to Surviving and Closing Successfully," Office of Advocacy, U.S. Small Business Administration. (This paper is part of a series of papers distributed by the U.S. Bureau of the Census and is based on research conducted by the author when he worked there, but does not represent the official views of either SBA or the Census.)

SBA's Statistical Technique and Estimated Relationships for Prepayments and Defaults Were Appropriate

At the time of our review, SBA used an appropriate technique known as multinomial logistic regression[7] to identify whether the variables included in its model were important influences on the likelihood that a borrower would either default on or prepay a loan and to estimate the magnitude of these relationships. This technique, which has been used in other models of this type, was appropriate because it corresponded to the decision-making process that borrowers faced. When deciding whether to default on the loan, prepay the loan, or keep it active, using this technique, SBA produced estimates of both the probability of default and the probability of prepayment.[8]

The relationships that SBA's equations estimated between different variables and the likelihood of defaults and prepayments were consistent with economic reasoning. For example, SBA's default equation suggested that defaults were more likely when unemployment was higher, and the rate of increase in gross domestic product was lower. Both of these estimated relationships were consistent with economic reasoning because it was less likely borrowers would continue paying their debts when more people are out of work, and the economy was growing less rapidly or in decline.

SBA's prepayment equation also suggested that prepayments were more likely when loans were made under the SBA Express Program, for which SBA guaranteed a smaller percentage of the loan amount than it did under the regular 7(a) business loan program. This result was consistent with our expectations because the smaller guarantee was likely to make lenders more cautious in making lending decisions, such that firms borrowing through this program may have been more creditworthy than firms borrowing through the regular program. In turn, the businesses' enhanced creditworthiness may have led to more prepayments because these

[7]Multinomial logistic regression is a technique used to estimate the probability of an event occurring when the variable of interest, such as the status of a loan, is best presented in categories rather than as continuous numbers. In this case, the categories might be default, prepay, or still active. Economists generally prefer this method to simpler techniques that provide less realistic estimates.

[8]SBA's model is based on quarters of the year, and the unit of analysis is the individual loan for as long as it remains active. So, if a loan was active for 16 quarters before being prepaid, there will be 16 observations as to whether a borrower had defaulted, prepaid, or made regular payments on the loan in that quarter. All these observations are used in estimating the likelihood of default and prepayment.

businesses may have been relatively more financially stable and may have been more likely to pay off their loans early. The details of SBA's default and prepayment equations, which show these relationships, are in appendix II.

Other Default and Prepayment Equations Would Also Be Reasonable and Lead To Different Subsidy Rate Estimates

We identified additional variables available to SBA, but not included in the model, that also influenced the likelihood of defaults and prepayments. The choice of variables included in a model reflects the modelers' professional judgment and different equations using different sets of variables can all be considered reasonable. To analyze the effect of adding additional variables, we tested SBA's model to estimate the 2003 subsidy cost using additional variables that (1) measured the current interest rate on 1-year U.S. Treasury bills and (2) considered the industry in which the borrowing firm operates. The interest rate could be important as either another measure of general business conditions or as a specific measure of the cost of capital. The industry in which the borrowing firm operates could be important if default and/or prepayment rates vary among industries, and the distribution of loans among industries varies over time. In addition, banks have traditionally recognized that the financial performance of a borrower depends on the nature of the business supporting the loan, the structure of the loan, and the financial condition of the firm. At the time of our review, SBA's econometric equations contain information on the loan and the firm but did not include information on the firm's business.

The estimates produced by our testing suggest that these variables also influenced the likelihood of defaults and prepayments occurring and, therefore, that equations using these variables could also be reasonable.[9] However, there are additional considerations that could be important in deciding whether to include a measure of interest rates in the default and prepayment equations. Specifically, including an interest rate variable would mean that forecasted interest rates would be used with the results of the econometric equations (and forecast values of other economic variables) to forecast future defaults and prepayments. The fact that forecasting interest rates is difficult may be a reason for not including an interest rate variable, even if the variable appears to be significantly related to the historical likelihood of default or prepayment. Furthermore, at

[9]We also wanted to use a variable measuring firm size since larger firms may have more resources they can use to avoid default in the event of adverse business conditions. However, SBA's database did not include data on firm size.

present, forecasted interest rates are low relative to the interest rates that prevailed over most of the period from which the data were drawn to develop SBA's equations, potentially limiting the usefulness of including an interest rate variable.

We found that including either the interest rate on 1-year Treasury bills or the industry in which the borrowing firm operates as a variable in the default and prepayment equations changed the estimated cost of the program. (See app. II.) According to SBA's model, the estimated subsidy rate for loans disbursed in 2003 was 1.04 percent. This estimate increased to 1.13 percent with the industry identifiers included and decreased to 0.76 percent with the inclusion of the interest rate on 1-year Treasury bills. In addition, when we included both the interest rate variable and the industry identifiers, we estimated a subsidy rate of 0.83 percent. Because interest rates are difficult to predict and have recently been quite low, we conducted tests to determine how sensitive the estimate was to small changes in forecasted interest rates. We found that it is not very sensitive to such changes. For example, when we increased the forecasted values above those included in the official OMB forecast by 10 percent, we estimated a subsidy rate of 0.80 percent while when we decreased the forecasted values by 10 percent we estimated a subsidy rate of 0.73 percent.

The range of estimated subsidy rates that result from including additional variables was roughly comparable to the range that resulted from using different economic assumptions. We tested the sensitivity of SBA's estimated subsidy rate to small changes in the forecast values of the GDP growth rate and the unemployment rate by reestimating the subsidy rate with SBA's model but used both more optimistic and more pessimistic assumptions about future economic conditions.[10] With the more optimistic assumptions, we estimated the subsidy rate decreased to 0.81 percent while with the more pessimistic assumptions we estimated that it increased to 1.28 percent.

[10]The optimistic assumptions were that the GDP growth rate was 10 percent higher than the OMB forecast, and the unemployment rate was 10 percent lower. The pessimistic assumptions were a 10 percent lower GDP growth rate and a 10 percent higher unemployment rate.

Estimates of Recoveries Depended Only on Age of Loan, Not Economic Conditions

SBA's model also included a separate econometric equation for estimating recoveries, which are the amounts of defaulted loans that were eventually recouped by collection efforts, such as the liquidation of assets. In this equation, the cumulative net recovery rate[11] for a cohort of loans was estimated as a function only of the age of the loans in that cohort. In particular, this equation did not include any economic variables, so forecasted recovery rates were estimated to resemble historical recovery rates even though economic conditions in the future might be quite different from the past. According to documentation provided by SBA of the work done to develop this equation, adding economic variables would not have increased the precision of the recovery rate estimates.[12]

The Model's Estimated Default and Recovery Rates Were in Line with Historical Experience

Our evaluation of the model's estimated default and recovery rates found that these rates were in line with historical experience of the 7(a) program. There are some limitations to evaluating expected future loan performance compared with historical data because over time the economy changes and underwriting criteria and other factors that affect loan performance may also change. Therefore, one would not expect the estimated loan performance to exactly mirror historical experience. However, these types of comparisons are useful to evaluate the model's estimated default and recovery cash flows. Because recently issued loans do not have significant experience and historical data can be summarized in several ways, we evaluated the new model's estimated default and recovery rates compared with historical data in two ways to determine whether the estimates were in line with historical experience.

In August 2001, we reported that from fiscal year 1992 through fiscal year 2000, SBA overestimated the cost of the 7(a) program by about $1 billion, primarily because it overestimated defaults by approximately $2 billion. Over this same period, SBA's estimated recoveries closely matched actual loan performance. SBA's prior method to estimate costs was based on averages of historical loan performance. As previously discussed, SBA's current model estimated defaults significantly differently than the prior

[11]The cumulative net recovery rate for a cohort of loans is defined as cumulative net recoveries to date divided by cumulative defaulted dollars to date.

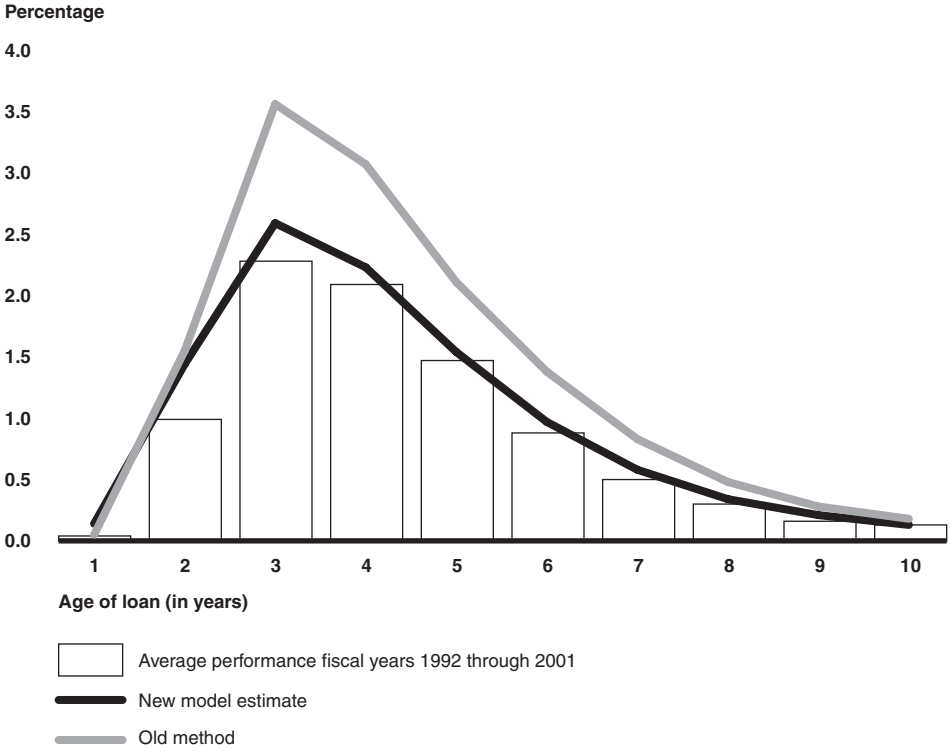
[12]Economic reasoning might suggest that recovery rates would be lower when economic conditions are unfavorable, but other attempts to incorporate economic variables into recovery rate equations have not been successful.

method in that it considered economic variables and loan specific information. Meanwhile, at the time of our review, the model continued to estimate recoveries based on historical patterns.

While it was currently not possible to determine the accuracy of the model's estimated default rate, as shown in the following two figures, the rate appeared to more closely match recent historical experience than SBA's previous method. Figure 2 shows how the model's estimated default rate compared with the estimated default rates calculated with SBA's previous method and with the average default experience of loans issued between 1992 and 2001.[13] We could have included more or fewer years of loans in our analysis, but we believe data since 1992 are sufficient to evaluate the model's estimated default rate compared with historical experience because it included several years of loans that have been through their peak default period, which for 7(a) loans is generally between years 2 and 5.

[13]The average default experience was calculated based on the actual default experience for loans issued between 1992 and 2001 (all years referred to are fiscal years) depending on the year of the loan when the default occurred. For example, year 1 average defaults are based on the average of actual first-year defaults that occurred for loans between 1992 and 2001. Year 2 average defaults are based on the average of actual second year-defaults that occurred for loans issued between 1992 and 2000. Year 10 average defaults are based on the average of actual tenth year defaults that occurred for loans issued in 1992.

Figure 2: Estimated Default Rates Compared with Average Default Experience from 1992 through 2001



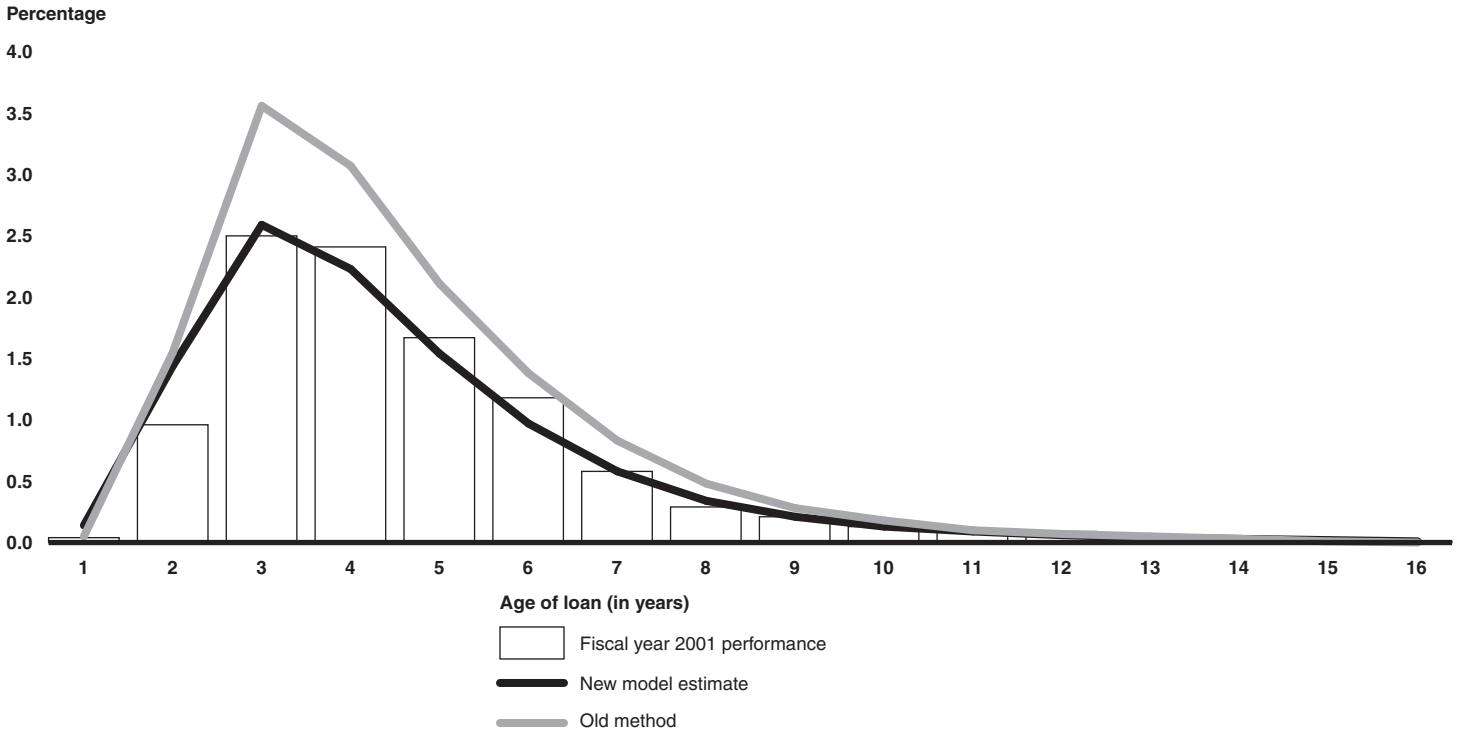
Source: GAO.

As previously mentioned, since historical data may be summarized differently, figure 3 shows how the new model's estimated default rate compared with the estimated default rate calculated with SBA's previous method and to actual default experience during fiscal year 2001 for the loans issued since 1986.[14] This comparison allowed us to evaluate the

[14]SBA officials told us they did not have the resources available to provide these data through fiscal year 2002.

estimated default rate over a longer period of time since data from older loans that have been outstanding for a longer period of time was included.[15]

Figure 3: Estimated Default Rates Compared with Fiscal Year 2001 Actual Default Experience



Source: GAO.

[15]We calculated the actual default experience during fiscal year 2001 for the loans issued since 1986 based on the default experience of those loans during fiscal year 2001 and the age of those loans. For example, the default experience of loans issued in 1991, which were in their eleventh year during 2001, was compared with the estimated default rates projected for the eleventh year.

SBA's Model Could Be Enhanced by Adding Information on Borrowers, Correcting Errors, and Updating Some Data

SBA could enhance the reliability of its model's estimates by adding information on both the businesses and the owners to the econometric equations and reestimating the equations and by correcting errors in the model. The econometric equations SBA used at the time of our review to predict default and prepayments included some variables describing the businesses and loans and two economic indicators, GDP and unemployment rates. But they did not include some variables other analysts and financial institutions often use that are associated with businesses and business owners, such as credit scores. In addition, during our review, we found some errors that resulted in underestimating the cost of the 7(a) program that was included in the fiscal year 2004 President's Budget. Correcting these errors would have increased the estimated cost of the program by about $6.5 million.

Including Additional Information on Businesses and Business Owners Could Enhance the Model's Reliability

The quantitative relationships between the default and prepayment rates and the current independent variables would probably change if new information were included. In our review of the literature and discussions with large banks, additional information was mentioned as having an influence on defaults and prepayments. The information cited was more detail on the loans, the business, and on business owners, including credit scores.[16]

Our review of the academic literature and discussions with some commercial lenders indicated that private lenders often include variables SBA did not consider in forecasting the financial performance of small businesses.[17] At the time of our review, the current SBA model included loan variables (age and term) and some business variables (new business indicators, form of ownership, and loan amount, among others) but was missing detailed information on businesses that can help predict financial viability. These variables include earnings, capital, payment records, and

[16]A credit score is a numerical measure of a borrower's creditworthiness based on a statistical analysis of past financial behavior and current financial obligations.

[17]Berger, Allen N., Frame, W. Scott, Miller, Nathan H. "Credit Scoring and the Availability, Price, and Risk of Small Business Credit," Federal Reserve Board, Mimeographed, April 2002; Caouette, John B., Altman, Edward I., Narayanan, Paul. *Managing Credit Risk: The Next Great Financial Challenge*. New York: John Wiley & Sons Inc., 1998; and W. Scott, Padhi, Machael, Woosley, Lynn, *The Effect of Credit Scoring on Small Business Lending in Low- and Moderate-Income Areas*, *Frame*. Federal Reserve Bank, Atlanta, Working Paper 2001-6, Unpublished, April 2001.

available collateral, all of which have been shown to affect creditworthiness and likelihood of default. Profit levels, for example, help predict a business's ability to generate cash internally to cover loan payments. Records of debt payments help determine whether a business can cover its obligations, while available collateral tells a lender whether a business has the resources to cover outstanding debts during a financial crisis. Adding and periodically updating this information could enhance the predictive ability of SBA's econometric model by providing more accurate estimates of potential defaults and prepayments.

In addition, analysts and banks have found that variables describing business owners can aid in evaluating credit risk, and many large banks have started to underwrite and monitor small businesses using credit scores. Information from business owners' credit records, such as income, personal debt, employment tenure, homeownership status, and previous personal defaults or delinquencies, can help predict delinquencies and defaults in the businesses themselves. Although at the time of our review SBA's current model did not include variables that measure these characteristics, the agency was developing a new loan monitoring system that SBA officials told us was intended to track this type of information. This is an important issue since, if banks use credit scores and the SBA does not, the SBA may be left with riskier loans. SBA could then determine whether such variables also reflect risks in SBA loans and could be used to help evaluate the costs of SBA loan guarantees.

SBA's 2004 Subsidy Rate Estimate Included Errors

During our review of the model used to generate the cost estimate of the 7(a) subsidy that was included in the fiscal year 2004 budget, we found errors that resulted in underestimates of program costs of about $6.5 million. Based on the estimated subsidy rate and the projected loan volume included in the fiscal year 2004 President's Budget, the estimated cost of the program was about $94.9 million. If the errors we found had been detected and corrected by SBA before the budget was submitted, the estimated cost of the program with the same projected loan volume would have increased to about $101.4 million.

These errors related to SBA's method of estimating recoveries, annual guarantee fee cash flows, and projections of borrower interest rates. First, the recovery estimates were based on the assumption that loans would be issued during fiscal year 2003 instead of during fiscal year 2004, although default and prepayment estimates were based on the later year. As a result, the model estimated that recovery cash flows would occur 1 year early,

affecting the net present value[18] of the cash flows and the subsidy rate. Second, formulas SBA used to summarize the output of the cash flow segment of the model indicated that the same annual guarantee fees collected during the first quarter of fiscal year 2004 would be collected from about years 5–27, even though the fees would decline as loan balances were paid off. SBA officials indicated that these two errors would be corrected before the submission of the 2005 budget. Third, in estimating the cost of loans issued in the future, SBA assumed the loans would have characteristics similar to those of loans issued during fiscal year 2001. However, SBA did not adjust the borrower interest rates to levels that would be more appropriate for loans to be issued during fiscal year 2004. SBA officials indicated that this adjustment was not necessary because it would not significantly affect the cost of the program. However, SBA had made this adjustment when it calculated the subsidy cost for loans to be issued during fiscal year 2003. When we corrected the previously described errors, the estimated cost of the program for fiscal year 2004 increased by $6.5 million. We also found an error related to estimating prepayment penalties. SBA officials stated that they were aware of this error but believed that fixing it would be complicated and that these cash flows would be immaterial to the cost of the program. In the officials' view, fixing the error would not be cost beneficial.

Cohort Data Could Be Updated

In addition, the model could also be further enhanced if SBA were to update the model to include new information as it becomes available. For example, SBA used the 2001 cohort of loans to generate estimates of the 2003 and 2004 subsidy. But, they were not sure if they were going to use the 2002 cohort of loans for the 2005 estimate because they said that updating the cohort is complicated as a result of changes in program policies or in the composition of the 7(a) loan portfolio. However, the model would likely produce more reliable estimates if the most recent loan data were being used to generate the forecast rather than continuing to use an older cohort of loans.

[18]Present value is the worth of the future stream of returns or costs in terms of money paid immediately. In calculating present value, prevailing interest rates provide the basis for converting future amounts into their "money now" equivalents.

SBA Collaborated with OFHEO and OMB to Develop the Model

SBA contracted with OFHEO economists, with expertise in econometric modeling of mortgage defaults and prepayments, to develop its subsidy model, which included determining the variables to be included in the econometric equations. SBA consulted with OMB officials, who are required by FCRA to approve agency subsidy estimates. SBA also hired a private consulting firm to conduct a limited review of the model as part of its ongoing review process to minimize errors in estimating the subsidy.

SBA Entered into an Agreement with OFHEO to Develop the Subsidy Model

In February 2002, SBA entered into an agreement with OFHEO to assist in developing the subsidy model. According to SBA staff, they selected OFHEO because it had staff with expertise and experience in econometric modeling and was less expensive than a private contractor.[19] According to SBA staff, the OFHEO economists followed a four-step process to develop the model. The first step was refining and building the data set that would be used to generate the estimates. The data set OFHEO used was constructed from the SBA databases that were used to track loan payment history and personal financial information on borrowers. The second step was the design and estimation of the default, prepayment, and recovery equations, including the selection of variables for these equations. The third step of the process was the construction of the cash flow module, and, the fourth step was the construction and testing of the model that OFHEO would deliver for use by SBA.

OMB Officials Approved SBA's Model

OMB officials also played a key role in the development of the model because, under FCRA, OMB has final responsibility for approving estimation methodologies and determining subsidy estimates. SBA officials said they consulted with OMB during the model's development until OMB approved it in the fall of 2002. OMB officials told us that they considered the model to be an improvement over the previous method that SBA used to calculate the program subsidy rate because it used better data and the econometric equations allowed for more accurate estimates of future cash flows. In addition, SBA could now use the model to consider both programmatic and economic variables in estimating the subsidy rate. For

[19]The Economy Act, 31 U.S.C. 1535, permits federal agencies to enter into agreements with other federal agencies for goods or services if the agency contracting the service cannot obtain the goods or services as conveniently or economically by contracting with a private source.

example, they said SBA could model how such variables as lender type affected the subsidy rate.[20] In reviewing the model, OMB officials told us that they focused on the methodology of the model, the cash flow projections, appropriate use of variables in the econometric equations, and the validity of the data used to make the calculations. They approved the model in November 2002.

SBA Hired a Private Consulting Firm to Review the Model

SBA hired a private consulting firm to conduct an independent limited review of the model for September 2002 to October 2002, as part of its ongoing process to identify errors before OMB approved the model. The consulting firm assessed the model conceptually and evaluated its underlying computer programming—specifically, the key data inputs that were the primary source of the model's cash flows and the model's programming specifications (to ensure they were correctly coded and that the code functioned properly). The firm also assessed the model's compliance with the relevant statutes and regulations and conducted scenario testing to evaluate how it performed under different economic assumptions. The consulting firm concluded that although the model performed reasonably well in estimating the subsidy cost, SBA had made errors in estimating loan guaranty and servicing fees, the calculation of recoveries, and prepayment penalties. SBA made changes to the model to address the identified discrepancies for fees and recoveries, the net effect of which was, to increase the subsidy rate estimate by about 36 percentage points. The consulting firm also determined that the model lacked adequate documentation and they were, therefore, unable to review the econometric component of the model. However, OFHEO subsequently provided SBA with a report documenting the model's development to a limited extent.

Lack of Adequate Model Documentation Hampered Independent Reviews of SBA's Model

In developing its new econometric model, SBA did not prepare adequate supporting documentation to enable independent reviewers to understand and evaluate the process that was used. For example, the independent contractor SBA hired to review the 7(a) credit subsidy model was hampered by the lack of adequate documentation and, as a result, this team's review of the model's theoretical basis and its working features was severely limited. While SBA later developed some general documentation

[20]The 7(a) program has three classifications of lenders: regular, certified, and preferred lenders.

of its model development process, this documentation did not contain, among other things, an adequate discussion of alternative variables, or combinations of variables, that it considered, tested, and rejected, and the reasons for rejecting them. SBA officials told us that they did not prepare this type of documentation because they believed that there was no specific requirement to do so. Current guidance is either silent or unclear about supporting documentation needed to explain the development of econometric models used to generate credit subsidy estimates for the budget and financial statements. Nevertheless, we believe that maintaining adequate documentation on how such models were developed is a sound internal control practice that would provide SBA and other agencies the opportunity to demonstrate and explain the rationale and basis for key aspects of their models that provide important cost information for budgets, financial statements, and congressional decision makers. Moreover, as a practical matter, this documentation would help facilitate SBA's and other agencies' annual financial statement audits.

SBA's 7(a) Credit Subsidy Model Documentation Was Inadequate for Outside Reviewers

BearingPoint, the independent contractor hired to perform an initial review of the SBA 7(a) credit subsidy model prior to its finalization, was hampered by the lack of adequate documentation. In response to our inquiry, the contractor stated that the team did not validate the model which, from an audit perspective, would have encompassed a more robust effort. In its final report to SBA, the contractor reported that SBA lacked sufficient supporting documentation for a "thorough review of its [the model's] theoretical basis (including alternative modeling methodologies explored), its working features, or the update and maintenance procedures necessary to use the model on an ongoing basis. This lack of adequate documentation severely limited our ability to assess certain critical parts of the model in detail, including its econometric components." Further, the contractor recommended that "SBA develop a robust set of documentation to support this model" including "the modeling methodology, alternate methodologies considered, data inputs and outputs, and model maintenance and update requirements."

In its January 30, 2004, audit report, Cotton and Company, the independent public accounting firm, identified in its internal control report 9 specific

deficiencies in the model's documentation.[21] These deficiencies included, for example, a lack of technical references for the statistical method used for the performance of the model, the absence of mathematical specifications, the fact that important variables were not clearly identified, and that units of measure for key variables were not specified. In addition, the audit report stated that the documentation that was provided was "self-contradictory" about the quality of the default and prepayment model and lacked a discussion of the assumptions and limitations of SBA's modeling approach. In responding to the independent public accountant's internal control report, SBA's Chief Financial Officer generally agreed with the report's findings, including the deficiencies in SBA's model documentation, and stated that the internal control report presented "fundamentals of good financial management and SBA is committed to accomplishing as many of these items as possible in the coming year."

In response to BearingPoint's recommendation, SBA's OFHEO contractor prepared some documentation for the model, but this documentation was not sufficient to allow us and SBA's financial statement auditor to gain an adequate understanding of certain key parts of the model development process. For example, the documentation that SBA provided included a broad overview of how the model works, a list of the variables that the final econometric equations included, the estimated coefficients of the equations, and figures showing how well the equations fit the data during the historical period. For some variables, SBA's documentation indicated how the variables were expected to influence default or prepayment probabilities, but did not provide any reasons, conceptual justification, or supporting empirical analysis. Some of these statements seemed intuitive, such as when the output of the economy increases, as measured by the percent change in real GDP, it is expected that default rates will drop. However, other statements were not intuitive. For example, SBA's documentation indicated that larger loans were expected to default at elevated levels and did not include any support for this assertion.

Additionally, the model documentation did not explain in sufficient detail why SBA excluded some variables. Rather, the model documentation included a table of 29 variables that were tested and rejected and stated that the information presented was "a list of most variables tested." The

[21]U.S. Small Business Administration, Office of the Inspector General Auditing Division, *Audit of SBA's FY 2003 Financial Statements*, audit report 4-10, (Washington, D.C.: Jan. 30, 2004).

documentation also provided a general overview about why these 29 variables were excluded. SBA's documentation stated that "variables were removed for a variety of reasons. Some of the reasons include—insignificant, highly correlated with other variables, low economic importance (significant but impact on probabilities was negligible), inconsistent results (variable was not robust to different specifications), and incoherent results (results could not be reconciled with any economic logic)." While the documentation that SBA provided to us contained acceptable reasons that economists could cite in rejecting variables, the documentation's lack of specificity did not allow us to determine which variables were rejected for which reasons. Further, we were unable to determine whether these were the only criteria or whether they were consistently applied throughout the model development process.

SBA and the OFHEO contractor told us that, during the model development process, approximately 800 pages of raw testing information were generated and retained in an electronic file. They further stated that these 800 pages were not organized in any fashion and that there was no summary document or road map with greater detail than the model documentation provided us that would describe the variable-testing process or the results of that process in an understandable fashion. In addition, SBA and the contractor told us that the variables reflected in the 800 pages were not recorded in English words, but rather in mnemonics, and that there was no crosswalk or key still in existence to decode the mnemonics. Based on these representations by SBA and its contractor, we initially concluded that this information would be of questionable or no usefulness in assessing SBA's development of the assumptions and selection of variables used in the modeling process.

SBA eventually provided us access to the 800 pages of material that contained some information on variables that were considered and rejected. This document was a partial compilation of analyses conducted during the model development process with no explanation or discussion of what was learned from each analysis conducted. Thus, on its own, this document provided little additional information regarding the process that SBA's contractor followed in developing the econometric equations used in the subsidy model. Further, the document was written in mnemonics and was not organized in any logical manner. In addition, SBA officials could not identify any specific parts of this documentation that related to alternative variables that were considered and rejected during the model development process.

Documenting the basis for selecting and rejecting variables from an econometric model used to develop credit subsidy estimates is an important internal control that would also help to provide financial statement auditors reasonable assurance that a bias was not introduced into the credit subsidy estimates by systematically excluding variables to influence the subsidy rate in a particular direction. Statement on Auditing Standards Number 57, *Auditing Accounting Estimates* (SAS No. 57), states that "even when management's estimation process involves competent personnel using relevant and reliable data, there is potential for bias in the subjective factors." When evaluating the reasonableness of an estimate, the auditor should concentrate on, among other things, "key factors and assumptions that are subjective and susceptible to misstatement and bias." Because of the nature of econometric models and the effect that variables used have on future loan default and prepayment projections, auditors need to understand both what was included and excluded from the model to assess the reasonableness of the credit subsidy estimate from a financial accounting perspective.

As our work demonstrated, changing the variables that were included in the model changed the subsidy rate. Because of the lack of adequate documentation on SBA's 7(a) model development process, we were unable to determine whether a bias in selecting variables existed in the model. Further, SBA's lack of adequate documentation on the 7(a) model development process could have impeded our ability to reach a conclusion on SBA's loan accounts in connection with the audit of the consolidated financial statements of the federal government.

Specific Guidance on Credit Subsidy Model Development Documentation Is Limited

Currently, there is limited specific guidance on the nature and extent of documentation that agencies must prepare related to the development of models to generate credit subsidy estimates. OMB Circular A-11, Preparation, Submission, and Execution of the Budget, provides guidance on how agencies should prepare credit subsidy estimates. Circular A-11 does not include any guidance to the agencies for documenting their model development process including selection and rejection of variables for use in the models that generate federal credit subsidy estimates. However, Federal Financial Accounting and Auditing Technical Release 6, Preparing Estimates for Direct Loan and Loan Guarantee Subsidies under the Federal Credit Reform Act Amendments to Technical Release 3: Preparing and Auditing Direct Loan and Loan Guarantee Subsidies under the Federal

Credit Reform Act,[22] provides some implementation guidance about the nature and extent of documentation agencies should have for their models. Technical Release 6 states that agencies should document the cash flow model(s) used and the rationale for selecting the specific methodologies. Agencies should also document the sources of information, the logic flow, and the mechanics of the model(s) including the formulas and other mathematical functions. In addition, because the model is the basis for budget and financial statement credit subsidy estimates, this documentation also facilitates an OMB budget analyst's review, if the analyst is not involved in the development process, the external financial statement audit, and other independent reviews. Technical Release 6 also states that agency documentation for subsidy estimates and reestimates should be complete and stand on its own, enabling an independent person to perform the same steps and replicate the same results with little or no outside explanation or assistance. In addition, if the documentation were from a source that would normally be destroyed, then copies should be maintained in the file for the purposes of reconstructing the estimate.

Technical Release 6 does not specifically address expected documentation of an agency's model development process, including a detailed discussion of alternative variables that are considered, the reasons for their rejection, and specific examples based on results of earlier regressions. Nevertheless, in our view, the documentation principles in this Technical Release represent sound internal control practice that could also be applied to an agency's development of a model used to generate budget and financial statement credit subsidy estimates. Such documentation would introduce transparency into an agency's budget process and enable agencies' models and the resulting estimates to withstand scrutiny and inquiry from independent reviewers. For example, such documentation would allow validation of an agency's model by independent reviewers, and provide reasonable assurance that the agency selected and rejected assumptions and variables for the model on a sound basis. Further, this documentation would help demonstrate to congressional stakeholders sound decision making and stewardship over millions of dollars in appropriated funds.

[22]Technical Release 6 was issued by the Accounting and Auditing Policy Committee (AAPC) a permanent committee established by the Federal Accounting Standards Advisory Board whose mission is to promulgate accounting standards for federal government reporting entities. The AAPC's role is to assist the federal government in improving financial reporting by providing solutions to accounting and auditing related issues. Technical Release 6 provides implementation guidance for agencies to prepare and report credit subsidy estimates.

SBA Had a Process to Help Ensure Data Quality and the Data Used in the Model and SBA's Loan Level Databases Were Consistent

Calculating a reliable credit subsidy estimate requires that the key cash flow data, such as defaults or recoveries and the timing of these events be reliable, or the credit subsidy estimate could be affected. Internal control standards call for agencies to have a process to help ensure the completeness, accuracy, and validity of all transactions processed. SBA's monthly reconciliation process, combined with lender incentives and loan sales, helped ensure the quality of the underlying data used in its credit subsidy estimation process. Although at the time of our review, some errors in its data existed in SBA's databases, the nature and magnitude of these errors was unlikely to significantly alter the subsidy rate. Further, we tested the data used by SBA's new econometric model and found them to be consistent with the data in SBA's loan systems at the time of our review.

SBA Had a Process to Identify and Correct Data Errors

The primary method that SBA used to help ensure the integrity of its loan data is its Form 1502 reconciliation process. Reconciliations are an important internal control established to ensure that all data inputs are received and are valid and all outputs from a particular system are correct. This process, which has been in effect since October 1997, utilized an SBA contractor to conduct monthly matches of borrower data submitted by 7(a) program lenders on SBA's Form 1502 to the information in the agency's Portfolio Management Query Display System to help ensure the completeness and accuracy of the agency's data. The information on the Form 1502 included a wide variety of data for an individual loan, some of which was used in the credit subsidy estimation process, and included, among other things, loan identification number; loan status such as current, past due, or in liquidation; loan interest rate; the portion of the loan guaranteed by SBA; and the ending balance of the loan's guaranteed portion. Errors identified by this match were loaded each month into SBA's Portfolio Management Guaranty Information System, and it was accessed by the various district office staff to work with lenders to correct the erroneous data.

Although we did not independently test the data match conducted by SBA's contractor or the field office staff's correction of identified errors, we reviewed summary reports of the errors in the Guaranty Loan Reporting System for each district office over a 4 month period during fiscal year 2003 and found that most of these reported errors were resolved during the month the errors were identified. During the months we reviewed, the percentage of errors resolved ranged from a low of about 65 percent to a

high of nearly 89 percent.[23] Although one month we reviewed had only a 65 percent resolution rate, leaving 4,860 errors uncorrected at the end of the month, as explained in the following paragraph, not all of these errors would affect the subsidy estimate and this number is relatively small compared to the large volume of loan transaction level data used in the credit subsidy estimation process. Our review of the underlying data used in the model showed that about 5.7 million data records were used to record the quarterly loan performance of 392,315 loans from 1988–2001.

In order to assess whether the remaining errors in SBA's data base would likely have a significant affect on the credit subsidy estimation process, we reviewed the 38 different error codes that are reported monthly by the Guaranty Loan Reporting System and found that less than half of these error codes were related to data used by the econometric model and, as a result, could have affected the credit subsidy estimate. For example, the Guaranty Loan Reporting System identified errors for lender contact name and phone number—data that were not used by the new econometric model and would not affect the subsidy estimate. Other error codes relating to the guaranteed portion principal balance or whether a loan was in liquidation status could affect the credit subsidy estimate if the number of errors and their dollar volume were significant.

We reviewed a 6-month summary error report from the Guaranty Loan Reporting System for activity between February and July 2003 and found that, for those error codes that could affect the credit subsidy estimate, only two of these codes had error rates that exceeded 1 percent of the transactions. One of these codes indicated that the loan status was not correct because the loan was in liquidation and had an average error rate of about 1.4 percent for the 6-month period we reviewed. The other error code indicated that the bank did not report any information for a particular loan and had an average error rate of about 2.4 percent for the same time period. The remaining 11 error codes that could have affected the credit subsidy estimate had rates of less than 1 percent. We assessed the error rates on this report in aggregate to determine if these could affect the credit subsidy estimate and found that the average aggregate error rate was about 6.5 percent during this period. However, given that most of these errors were corrected in the month the error was identified, it was unlikely that the

[23]The number of errors does not equal the number of loans that had errors since a single loan can have multiple errors.

remaining uncorrected errors would affect the credit subsidy estimate at the time of our review.

Lender Incentives and Loan Sales Help Ensure Data Integrity

In addition to the monthly loan data reconciliation process, lender incentives also helped ensure the integrity of the underlying data used in the credit subsidy estimates. In accordance with current SBA policy, the agency can reduce or completely deny a lender's claim payment if the defaulted loan data are not correct. According to SBA officials, this policy gives the 7(a) program lenders an incentive to correct data errors because it helps ensure they will be paid the full guarantee amount if the borrower subsequently defaults on the loan. SBA provided us with repair and denial data for fiscal years 1999 through the first three quarters of fiscal year 2003 showing that the agency exercised these options 2,177 times during this time, totaling at least $69.9 million.[24]

Further, an ancillary benefit of SBA's loan sales program was to help ensure data integrity. Prior to a sale, SBA district office staff, as well as contractors, reviewed loan files as part of the "due diligence" reviews to provide accurate information about the loans available for sale to potential investors so that they may make informed bids. SBA officials told us that prior to selling a loan, discrepancies between the lenders' data and SBA had to be resolved.

Data Used by the Econometric Model Were Consistent with SBA Databases

In order to assess the consistency between the data used in SBA's econometric approach and the data in SBA's loan system, we selected and tested a stratified random sample of 400 items to test key data that could affect the credit subsidy estimate and found no errors.[25] Specifically, we randomly selected 100 default and recovery transactions and compared the amounts and transaction dates between the loan system data and loan-level data used for the credit subsidy estimate. In addition, we randomly selected

[24]According to SBA officials, the actual repair and denial amounts are higher than this because many lenders release SBA of its guaranty obligations rather than having repairs or denials on their lender record.

[25]Because we followed a probability procedure based on random selections, our sample is only one of a large number of samples that we might have drawn. Since each sample could have provided different estimates, we express our confidence in the precision of our particular sample's results as a 95 percent confidence interval. This is the interval that would contain the actual population value for 95 percent of the samples we could have drawn. As a result, we are 95 percent confident that the number of data errors in key aspects of the model, such as default and recovery dates and amounts and loan status, do not exceed 1 percentage point of the key data used in the model.

100 loans identified by the model to be prepaid and reviewed the loan histories in SBA's database and determined that all of these loans were paid off prior to their scheduled termination date. Further, we tested 100 additional loans and compared their status such as current, paid off, or default to ensure their status in the model was correct and found no errors.

We also assessed the magnitude of 7(a) loans that were excluded from the model in order to determine whether excluding these potentially valid loans would likely affect the credit subsidy estimate. Our earlier work on SBA's previous 7(a) credit subsidy model that primarily used historical averages of defaults and recoveries found that excluding loans from certain years that had higher default rates would lower the overall average default rate. Excluding large numbers of loans from this model would likely have a similar effect on the estimated subsidy rate. To assess the magnitude of excluded loans, we reviewed the computer coding for the econometric model and found that SBA excluded loans when critical data for the model were missing such as the initial disbursement date, the loan amount, or demographic information on the borrowers. For most of the years between 1988 and 2001, the number of loans excluded because they lacked these essential data ranged from 1 percent to 2 percent and overall, we concluded that the degree of excluded loans was acceptable and would not significantly affect the credit subsidy estimation calculation, at the time of our review.

Conclusions

Overall, we found that from an economics perspective, SBA's econometric equations for its 7(a) credit subsidy model were reasonable. However, from an audit perspective, SBA's lack of adequate documentation of the model development process precluded us from (1) independently evaluating the model's development; (2) determining whether SBA used a sound and consistently applied method to select and reject variables to be included in the model; and (3) determining whether a bias in selecting variables existed in the model.

Based on our review, SBA's econometric equations for estimating defaults, prepayments, and recoveries, which were used to derive the estimate of its fiscal year 2004 subsidy costs, were reasonable. This model's methodology has the potential to produce more reliable estimates than the previous method of using historical averaging to project the estimated program cash flows because this model relies on economic reasoning in addition to historical program data. However, the precision of any econometric model is limited because any estimate produced by such a model should be

considered one point in a range within which the actual subsidy cost will likely fall. Because the budget process requires agencies to select a specific estimate rather than project a range, there will likely be some variance between the forecasted and actual subsidy amounts. Using additional data that SBA anticipates gathering in its new loan monitoring system, such as borrower-specific data, could further enhance the reliability of SBA's estimates of the subsidy cost. Therefore, further enhancements could produce more reliable results.

Although the errors we identified in the model did not materially affect the subsidy cost estimate, they did indicate that the process SBA used to validate the model could be improved. Therefore, it is important to invest the resources needed to periodically reevaluate the underlying assumptions of any model to ensure that they are correct and comprehensive, and that any errors or erroneous assumptions are corrected so that the model continues to yield reasonable results.

While we found SBA's equations to be reasonable from an economics perspective, the lack of adequate documentation of the model's development process hampered three independent reviews of the 7(a) model. Notwithstanding the current lack of clear OMB Circular A-11 guidance, SBA could benefit from applying the documentation principles embodied in Technical Release 6 to the development of the 7(a) econometric model and other credit subsidy estimation models it has recently developed or is currently developing. Without adequate documentation, SBA will be unable to transparently demonstrate the rationale and basis for key aspects of models that provide important cost information for budgets, financial statements, and congressional decision makers. Although OMB provides guidance on how agencies should prepare credit subsidy estimates in Circular A-11, it does not include any guidance to the agencies for documenting their model development process including the selection and rejection of variables for use in the models that generate federal credit subsidy estimates. A lack of improved OMB guidance for model documentation will continue to hamper adequate external oversight and validation of models used to generate credit subsidy estimates.

Recommendations for Executive Action

We are making three recommendations to SBA and one to OMB. To further enhance the reliability of SBA's subsidy estimates, we recommend that the SBA Administrator take the following two actions:

- determine how best to include in future subsidy models borrower-specific information, such as credit scores and loan-to-value ratios, to be collected in the new loan monitoring system; and

- ensure that the model remains reasonable by establishing a process for periodically evaluating the model to correct any errors and revising it to reflect changes in the 7(a) business loan program or other factors that could affect the subsidy estimate.

To demonstrate and explain the rationale and basis for the 7(a) econometric model and all other models developed, we recommend that the SBA Administrator take the following action:

- prepare and retain adequate documentation of the model development process including a detailed discussion of the alternative variables or combinations of variables that were considered, tested, and rejected, as well as the reasons for rejecting them.

To facilitate (1) validation of models used to generate credit subsidy estimates, (2) external oversight, and (3) financial statement audits, we recommend that the Director, OMB, take the following action:

- revise OMB Circular A-11 to require that agencies document the development of their credit subsidy models, including the process followed for selecting modeling methodologies over alternatives, and variables tested and rejected, along with the basis for excluding them.

Agency Comments and Our Evaluation

We provided an initial draft and a revised draft, based on our review of additional model documentation, to both SBA and OMB for review and comment. While our initial draft was at the agencies for comment, we continued to pursue additional documentation that SBA had to further explain its 7(a) model development process, including what variables were selected, rejected, and why. When we eventually obtained access to the 800 pages of SBA material, we determined that it was not organized and included no road map to describe the variable testing process or its results. We concluded that this information was of questionable or no usefulness to our assessment of SBA's modeling process. We addressed the weaknesses in SBA's documentation in the revised draft report and provided it to SBA and OMB for comment. In commenting on the initial draft, SBA's Chief Financial Officer (CFO) generally agreed with our findings and the first two recommendations related to actions to further enhance the reliability of the

model's subsidy estimates. OMB did not provide any comments on the initial draft report. We received comments on the revised draft from SBA's CFO who generally disagreed with our findings and recommendations related to the lack of adequate documentation supporting the model's development process. We also received comments on the revised draft from the OMB Assistant Director for Budget and the Controller who disagreed with our recommendation that OMB revise Circular A-11. Their written comments are reprinted in appendixes III and IV, respectively, and are summarized below. Both agencies provided technical comments that we have incorporated into the report as appropriate.

In commenting on our final draft report, SBA stated that it had provided us with extensive documentation, briefings, and explanations about how the model was developed. We met with SBA officials and their contractor who constructed the model and discussed their methodology, but we were unable to corroborate this information with the documentation they subsequently provided. SBA's comment letter stated that it provided us with 800 pages of material that contained some information on variables that were considered and rejected. During our subsequent review of this material, we found that this documentation was a partial compilation of analyses conducted during the model development process with no explanation or discussion of what was learned from each analysis conducted. After reviewing all of this documentation, as discussed in the report, we concluded that it provided little additional information to enable us to understand and corroborate the process and criteria that SBA used to select and reject variables for its 7(a) model.

Our conclusions regarding the lack of adequate documentation for the model's development process were consistent with those of both the independent contractor SBA hired to review the model in 2002 prior to its implementation and the independent public accounting firm that audited SBA's fiscal year 2003 financial statements. As part of its January 30, 2004, audit report, the independent public accounting firm identified in its internal control report 9 specific deficiencies in the model's documentation. These deficiencies included, for example, a lack of technical references for the statistical method used for the performance of the model, the absence of mathematical specifications, that important variables were not clearly identified, and that units of measure for key variables were not specified. In addition, the audit report stated that the documentation that was provided was "self-contradictory" about the quality of the default and prepayment model and lacked a discussion of the assumptions and limitations of SBA's modeling approach. While SBA's CFO

agreed with the independent accounting firm's findings regarding the lack of adequate documentation for the credit subsidy model, he disagreed with similar weaknesses identified in our report.

SBA disagreed that its lack of adequate documentation on the 7(a) model development process could impede our ability to reach a conclusion about SBA's loan accounts in connection with the audit of the consolidated financial statements of the federal government. Instead, SBA believed mandating additional documentation would establish a new and unnecessary requirement. Our comment was in regard to our responsibility as the auditor of the consolidated financial statements of the federal government and does not establish a new or unnecessary requirement for SBA. For the consolidated financial statement audit, we evaluate the reasonableness of credit program estimates based on audit guidance in SAS No. 57.[26] In auditing estimates, SAS No. 57 states that an auditor should consider, among other things, the process used by management to develop the estimate, including determining whether or not (1) relevant factors were used, (2) reasonable assumptions were developed, and (3) biases influenced the factors or assumptions. SBA's lack of adequate documentation of the 7(a) model development process impaired our ability to make such an assessment.

OMB disagreed with the recommendation that Circular A-11 should be revised and believed that the report did not demonstrate that revisions were needed. OMB officials commented that they worked closely with SBA during the model development process and believed that the documentation SBA provided to OMB was adequate for them to determine that the subsidy estimates and reestimates were reasonable. OMB also did not concur with our statement that a lack of improved OMB guidance hampered adequate external oversight. Unlike OMB, in this case, we and other external reviewers did not have the opportunity to work with SBA during the model development process and, as a result, relied on oral explanations and documentation provided by SBA staff and its contractor who developed the model. Further, we attempted to corroborate SBA's statements with the documentation that SBA provided. However, as we reported, three independent external reviews of SBA's 7(a) model were hampered by a lack of adequate documentation of SBA's model development process. We reaffirm our conclusion that adequate

[26]SAS No. 57 became effective for audits of financial statements for periods beginning on or after January 1, 1989.

documentation is needed for the SBA 7(a) model's development and that independent external review and oversight will continue to be hampered without a requirement to provide adequate documentation about how econometric models are developed.

OMB stated that Ernst and Young was able to independently validate SBA's 7(a) model with the available documentation. According to OMB, this firm stated that the 7(a) model assumptions and methodology appeared to be reasonable and accurate. We obtained and reviewed the reports OMB cited and found that the firm was not hired to validate or review the same segments of the model that we reviewed. This series of reports was related to the cash flow module of the 7(a) model, as well as the model used to calculate reestimates, but did not review the econometric equations or the model's development process. In its report, the firm explicitly stated that it was not reviewing the same parts of the model that we reviewed. We confirmed this information in conversations with the accounting firm's engagement partner and concluded that this firm's work was not relevant to the findings and conclusions presented in our report.

OMB also commented that SAS No. 57 states that internal controls over accounting estimates may or may not be documented. While SAS No. 57 does state that the process for preparing accounting estimates may not be documented, it also states that auditors should assess whether there are additional key factors or alternative assumptions that need to be included in the estimate and assess the factors that management used in developing the assumptions. Further, SAS No. 57 states that auditors should concentrate on key factors and assumptions that are subjective and susceptible to misstatement and bias. We believe this includes the selection and rejection of variables that can be included in the model. Without adequate documentation on the credit subsidy model development process, it is difficult for auditors to fulfill their responsibilities to assess these areas.

OMB also commented that SBA fulfilled the management responsibilities described in SAS No. 57 regarding internal controls for accounting estimates. We disagree with this statement and point out that SAS No. 57 provides guidance for auditing accounting estimates as part of conducting financial statement audits rather than directing agency management's actions. Management's responsibility for internal controls are contained in our "Standards for Internal Control in the Federal Government," which states, among other things, that "internal control and all transactions and other significant events need to be clearly documented, and the

documentation should be readily available for examination."[27] Further, as previously stated, Cotton and Company also identified the lack of adequate model documentation as an internal control weakness. Moreover, SBA's CFO generally agreed with the independent public accountant's report's findings, including the deficiencies in SBA's model documentation, and stated that the internal control report presented "fundamentals of good financial management and SBA is committed to accomplishing as many of these items as possible in the coming year."

OMB also stated that requiring agencies to prepare additional documentation of the variables tested and rejected would be unduly burdensome. We disagree with this statement and note that this documentation would only need to be prepared when a model is developed or when significant updates are implemented. Further, this requirement would be consistent with other segments of OMB Circular A-11 that require agencies to provide supporting documentation for their budget submissions. However, as we mentioned in the report, there is currently no explicit guidance for agencies to document the development of the models that are used to generate credit subsidy estimates.

OMB also commented that we received sufficient information to test alternative variables to measure the reasonableness of the final SBA credit subsidy model. We note that our work demonstrated that using additional variables that were also reasonable changed the subsidy estimate. We believe that this work highlights the need for agencies to document their basis for rejecting variables or combinations of variables from their final credit subsidy models. By documenting this work, agencies will be able to demonstrate to independent reviewers that a bias from variable selection does not exist in the final model.

Both agencies provided technical comments that we incorporated into the report as appropriate. The written comments of both agencies are reprinted in appendixes III and IV.

We are sending copies of this report to the Chair of the Senate Committee on Small Business and Entrepreneurship, other appropriate congressional committees, the Administrator of the Small Business Administration, and

[27]U.S. General Accounting Office, *Internal Control: Standards for Internal Control in the Federal Government*, GAO/AIMD-00-21.3.1 (Washington, D.C.: November 1999).

the Director of the Office of Management and Budget. We also will make copies available to others upon request. In addition, the report will be available at no charge on the GAO Web site at http://www.gao.gov.

If you have any questions about this report, please contact me at (202) 512-8678 or dagostinod@gao.gov or Katie Harris, Assistant Director, at (202) 512-8415 or harrism@gao.gov. Key contributors to this report are listed in appendix V.

Davi M. D'Agostino
Director, Financial Markets
and Community Investment

Objectives, Scope, and Methodology

As agreed with your staff, we (1) assessed the reasonableness of the model's econometric equations and evaluated the model's estimated default, prepayment, and recovery rates based on the 7(a) program's recent historical loan experience; (2) identified additional steps the SBA could take to further enhance the reliability of its subsidy estimate produced by the model; (3) reviewed SBA's process for developing the subsidy model; (4) evaluated the model's supporting documentation, including its discussion of what variables were tested and rejected; and (5) determined what steps SBA has taken to ensure the integrity of the data used in the model and determined whether these data are consistent with information in its databases. We did not validate SBA's model.

Assessing the Reasonableness of the Model's Econometric Equations and Evaluating the Model's Estimated Default, Prepayment, and Recovery Rates

To analyze the model, we obtained from SBA copies of the model as approved by the Office of Management and Budget (OMB), along with the loan-level data that were used to develop the subsidy estimates. We analyzed the econometric equations to determine whether they were reasonable based on the variables they included, the statistical techniques used, and the results obtained. For example, we determined whether the econometric equations included appropriate variables and whether the variables used in the equations were statistically significant. To evaluate the model's estimated default and recovery rates, we compared these rates with recent historical loan experience of the 7(a) program provided by SBA. Using SBA's data, we also calculated what SBA would have estimated for default and recovery rates based on the estimation methodology it used prior to its fiscal year 2003 budget submission. (See app. II for a detailed discussion of our analysis of the reasonableness of the model's econometric equations.)

Identifying Additional Steps SBA Could take to Further Enhance the Reliability of the Model

To identify additional steps SBA could take to enhance the reliability of its model, we considered additional types of data that SBA might collect and consider including in its econometric equations. As part of this analysis, we reviewed the academic literature on default modeling and interviewed officials with several banks engaged in similar efforts.

Reviewing SBA's Process of Developing the Subsidy Model

To determine SBA's process for developing the model, we met with SBA officials in the Chief Financial Office who were responsible for estimating the 7(a) program subsidy costs. We also met with OMB officials who were responsible for approving the model. Finally, we also reviewed available documentation on the model's development provided by SBA and the report by the private consultant who reviewed the model.

Evaluating the Model's Supporting Documentation, Including Its Discussion of What Variables Were Tested and Rejected

To evaluate the model's supporting documentation, including its discussion of what variables were tested and rejected, we obtained and analyzed available relevant documents and met with SBA officials and their contractor who developed the model. We compared the information presented in SBA's model documentation with existing credit subsidy guidance including OMB Circular A-11 and Federal Financial Accounting and Auditing Technical Release 6: *Preparing Estimates for Direct Loan and Loan Guarantee Subsidies under the Federal Credit Reform Act Amendments to Technical Release 3: Preparing and Auditing Direct Loan and Loan Guarantee Subsidies under the Federal Credit Reform Act*. We also assessed the impact the lack of documentation would have on SBA's financial statement audit by comparing the documentation with Statement on Auditing Standards Number 57, *Auditing Accounting Estimates*. SBA and its contractor told us that 800 pages of raw testing information contained in an electronic file was not organized in any fashion, and that there was no summary document or road map that had greater detail than the model documentation provided us that described the variable-testing process or the results of that process in an understandable fashion. In addition, SBA and the contractor told us that the variables reflected in the 800 pages were not recorded in English words, but rather in mnemonics, and that there was no crosswalk or key still in existence to decode the mnemonics. Thus, no documentation existed that would link the variable names used in the programming to a table of variable descriptions. We obtained and reviewed a copy of this documentation and confirmed the representations of SBA and its contractor.

Determining What Steps SBA Took to Ensure the Integrity of the Data Used in the Model and Whether These Data Were Consistent with Information in Its Databases

To determine what steps SBA took to ensure the integrity of the data used by the model, we met with SBA officials to gain a general understanding of the agency's data integrity efforts. We also assessed the number of errors that were resolved by the district offices each month by analyzing 4 months of fiscal year 2003 field office activity from the Form 1502 Guaranty Loan Reporting System. We further assessed whether the remaining errors at the end of the month would likely affect the credit subsidy estimate by analyzing the types of errors tracked by the system and determining which errors affected data used by the new model. We also assessed the magnitude of these errors by analyzing 6 months of fiscal year 2003 activity in the Guaranty Loan Reporting System. To determine whether the data in the new model was consistent with data in SBA's loan-level databases, we selected and tested a stratified random sample of 400 key data elements that could affect the credit subsidy estimate.[1] Specifically, we randomly selected 100 default and 100 recovery transactions and compared the amounts and transaction dates between the loan system data and loan-level data used for the credit subsidy estimate; 100 loans identified by the model to be prepaid and reviewed the loan histories in SBA's database to determine whether all of these loans were paid off prior to their scheduled termination date; 100 additional loans and compared their status such as current, paid off, or default to determine if their status in the model agreed with SBA's loan-level databases.

[1]We are 95 percent confident that the number of data errors in key aspects of the model, such as default and recovery dates and amounts and loan status, do not exceed 1 percent of the data population.

Analysis of Default, Prepayment, and Recoveries Econometric Equations

This appendix provides more detail on the three econometric equations that the Small Business Administration (SBA) used to estimate the subsidy rate for its 7(a) loan guarantee program and the expanded equations that we developed. These equations are used to forecast defaults, prepayments, and recoveries. The first section of this appendix describes the variables that SBA used in the default and prepayment equations and presents SBA's estimated coefficients. The second section explains how we created the variable that we used to represent the borrower's industry and presents the estimated coefficients from our expanded default and prepayment equations. The third section describes the equation that SBA used to forecast recoveries and presents the estimated coefficients from that equation.

SBA's Default and Prepayment Equations

In its new model for estimating the subsidy rate for the 7(a) loan program, SBA uses multinomial logistic regression to estimate the likelihood of defaults and prepayments as functions of a variety of explanatory variables. Because multinomial regression is a simultaneous estimation process, the default and prepayment equations are identically specified (that is, the same explanatory variables are used in each equation). SBA conducts its analysis at the level of the individual loan, using loans that were disbursed from 1988 through 2001. For each loan, SBA's data set contains an observation for each quarter that the loan is active. For example, if a loan prepays at the end of the third year (counting the disbursement year as the first year), then it is active during 12 quarters and, therefore, there are 12 observations for that loan in the data set.

For each observation, the dependent variable measures whether in that quarter the borrower defaults on the loan, prepays the loan, or keeps it active. As a result, the coefficients in the default or prepayment equation are estimates of the association of each explanatory variable with the likelihood of the loan defaulting or prepaying in that quarter.

There are several categories of explanatory variables included in the default and prepayment equations. The first group consists of a set of dummy variables that indicate the age of the loan. These variables thus serve to reflect the fact that prepayment and default behavior change as a loan seasons. Specifically, there is a dummy variable for each of the first ten quarters of the life of a loan. From the eleventh quarter to the thirty-fourth quarter, there is a dummy variable for each two consecutive quarters. Finally, if a loan remains active past an age of thirty-four quarters, there is one more dummy variable.

The second set of explanatory variables concern loan characteristics. A set of dummy variables indicates the contractual term of the loan at origination. The categories are less than 5 years, 5 to up to 10 years, 10 years to up to 15 years, and 15 years or greater. Less than five years serves as the omitted category in the regression. Loan amount is another characteristic and is measured in millions of dollars. SBA also includes a dummy variable that shows whether a loan was delivered through the SBA Express Program. Also known as Subprogram 1027, this program allows lenders to originate a loan using their own loan documents instead of SBA documents and processing, but the loan guarantee is only up to 50 percent. By comparison, the typical SBA guarantee is almost 80 percent. Finally, there is a set of dummy variables for type of lender: Regular, Preferred, and Certified. In the regression, the regular type serves as the omitted category.

The next set of explanatory variables provides information on the borrower. A set of dummy variables identifies ownership structure. The categories are sole proprietorship, corporation, or partnership. Sole proprietorship is the omitted category in the regression. An additional dummy variable indicates whether the borrower is a new business. Finally, there is a set of dummy variables that indicate the U.S. Census Bureau region where the borrower is located.

The final set of explanatory variables contains two measures of economic conditions. The first is the state unemployment rate where the borrower is based. The source for these data is the U.S. Bureau of Labor Statistics. The second is the quarterly percentage change in gross domestic product. SBA obtained these data from the U.S. Bureau of Economic Analysis.[1] Table 1 summarizes the explanatory variables.

[1]Using data obtained from SBA, we were able to successfully replicate their equations.

Table 1: Variable Names and Descriptions

Variable name	Variable description
Age dummy variables	**If loan is:**
i1	1 if loan is 1 quarter old, else 0
i2	1 if loan is 2 quarters old, else 0
i3	1 if loan is 3 quarters old, else 0
i4	1 if loan is 4 quarters old, else 0
i5	1 if loan is 5 quarters old, else 0
i6	1 if loan is 6 quarters old, else 0
i7	1 if loan is 7 quarters old, else 0
i8	1 if loan is 8 quarters old, else 0
i9	1 if loan is 9 quarters old, else 0
i10	1 if loan is 10 quarters old, else 0
i1112	1 if loan is 11 or 12 quarters old, else 0
i1314	1 if loan is 13 or 14 quarters old, else 0
i1516	1 if loan is 15 or 16 quarters old, else 0
i1718	1 if loan is 17 or 18 quarters old, else 0
i1920	1 if loan is 19 or 20 quarters old, else 0
i2122	1 if loan is 21 or 22 quarters old, else 0
i2324	1 if loan is 23 or 24 quarters old, else 0
i2526	1 if loan is 25 or 26 quarters old, else 0
i2728	1 if loan is 27 or 28 quarters old, else 0
i2930	1 if loan is 29 or 30 quarters old, else 0
i3132	1 if loan is 31 or 32 quarters old, else 0
i3334	1 if loan is 33 or 34 quarters old, else 0
i35p	1 if loan is older than 34 quarters, else 0
Loan characteristics	**If term of loan is:**
t5_10	1 if term of loan is at least 5 years but less than 10, else 0
t10_15	1 if term of loan is at least 10 years but less than 15, else 0
t15p	1 if term of loan is 15 years or more, else 0
sub1027	1 if loan delivered through SBA Express Program, else 0
loan_amt	Gross guaranteed disbursed amount in millions
Lender_PLP	1 if lender is part of the Preferred Lender Program, else 0

(Continued From Previous Page)

Variable name	Variable description
Lender_CLP	1 if lender is part of the Certified Lender Program, else 0
Borrower characteristics	**If borrower is:**
Corporation	1 if borrower is incorporated, else 0
Partnership	1 if borrower is a partnership, else 0
NewBusiness	1 if borrower is a new business, else 0
Northeast	1 if located in U.S. Census Bureau's Northeast Region, else 0
Midwest	1 if located in U.S. Census Bureau's Midwest Region, else 0
South	1 if located in U.S. Census Bureau's South Region, else 0
Economic conditions	
Urate	Unemployment rate in the state where firm is located
pc_gdp96	Quarterly percent change in constant dollar GDP

Source: GAO.

The coefficients in the SBA equations indicate that the probability of both defaults and prepayments generally increase and then decline as a loan seasons. Defaults peak during the eighth quarter while prepayments peak around quarters 27 and 28. Longer-term loans are less likely to default or prepay. By comparison, larger loans are more likely to default or prepay. Good economic conditions, as reflected by the coefficients on unemployment and the percentage change in gross domestic product, reduce the chances of default and increase the likelihood of prepayment. The positive coefficients on the variable for new business indicate that such firms are more likely to default and prepay. Corporations and partnerships are less likely to default and more likely to prepay than sole proprietors. Finally, loans granted under Subprogram 1027 are less likely to default and more likely to prepay. Table 2 presents the coefficients in SBA's default and prepayment equations as well as some summary statistics.

Table 2: Multinomial Logistic Regression Coefficient Estimates[a]

	Predicting to defaults	Predicting to prepayments
Variables	**Base model**	**Base model**
Constant	-9.7650	-5.2762
i1	2.1151	1.1203
i2	3.1174	1.6016
i3	3.8158	1.9374
i4	4.2247	2.1063
i5	4.5187	2.2865
i6	4.6659	2.4113
i7	4.7487	2.5805
i8	4.8211	2.7080
i9	4.8068	2.8163
i10	4.8121	2.9133
i1112	4.8033	3.0540
i1314	4.7772	3.1439
i1516	4.7101	3.3111
i1718	4.6214	3.4554
i1920	4.6136	3.6945
i2122	4.5156	3.5201
i2324	4.4297	3.6685
i2526	4.2945	3.8222
i2728	4.3414	4.0106
i2930	4.2515	3.6142
i3132	4.2036	3.7143
i3334	4.1378	3.7914
i35p	4.1027	3.9950
t5_10	-0.0462[a]	-0.6568
t10_15	-0.7596	-1.1013
t15p	-0.7395	-1.1014
sub1027	-0.5800	0.0812
Loan_amt	0.2578	0.1189
Corporation	-0.0434	0.0989
Partnership	-0.1982	0.0211[a]
Northeast	0.3612	-0.2054
Midwest	0.2184	-0.1869
South	0.4142	-0.0928

(Continued From Previous Page)

	Predicting to defaults	Predicting to prepayments
Variables	**Base model**	**Base model**
Lender_PLP	-0.1761	0.0824
Lender_CLP	-0.1688	-0.0014[b]
NewBusiness	0.2773	0.0678
urate	0.1043	-0.0957
Pc_gdp96	-0.1261	0.0661
Summary statistics for multinomial logistic regression models		
		Base model
N of Observations		5,736,628
Likelihood Ratio Chi Sq		120,478
Degrees of Freedom		76
Significance levels		<.0001

Source: GAO.

[a]Except as noted, significance of coefficients is less than or equal to .0001. Significance of coefficients marked (a): < .05; those marked (b) had significance greater than .05.

Effects of Including Additional Variables

Although we found that SBA's default and prepayment equations are reasonable, we evaluated the impact of including additional variables in those equations and found that equations containing some additional variables are also reasonable. In particular, we found that when measures of interest rates and the industry of the borrower are included, these factors appear to be significantly related to the likelihood of defaults and prepayments. Table 3 presents the descriptions of the additional variables.

Table 3: Names and Descriptions of Additional Variables

Variable name	Variable description
tbill	Interest rate on 1 year U.S. Treasury Bills
Agri_etc	1 if firm is in agriculture, else 0
Mine_Const	1 if firm is in mining or construction, else 0
Manuf	1 if firm is in manufacturing, else 0
Wholesale	1 if firm is in wholesale trade, else 0
Trans_etc	1 if firm is in transportation, communication, or utilities, else 0
Retail	1 if firm is in retail trade, else 0
Finan_etc	1 if firm is in finance, insurance, or real estate, else 0

Source: GAO.

Table 4 presents the coefficients from three alternative specifications of the default and prepayment equations, respectively, as well as, for comparison purposes, the coefficients from SBA's equations. The first pair of alternative equations include an interest rate variable, the second pair include a set of dummy variables that identify the borrower's industry, and the third pair include both the interest rate variable and the industry-specific dummy variables.

The interest rate variable that we use is the interest rate on 1-year Treasury bills. We selected that rate, in part, because of the availability of forecasted values for it that would be consistent with the forecasted values SBA uses for other economic indicators in forecasting future defaults and prepayments.

To create the industry-specific dummy variables, we used data from SBA that identified the borrower's industry category, using either the Standard Industrial Classification (SIC) codes or the North American Industrial Classification (NAIC) codes. The NAIC is the Department of Commerce's current system for classifying businesses into industries and in 1997 the NAIC codes replaced the SIC codes that Commerce previously used. When possible, for loans that had NAIC codes, but not SIC codes, we converted the NAIC code into the corresponding SIC code. We aggregated the SIC codes into broader categories defined by the first digit of the code. To reduce the number of dummy variables, we aggregated some small categories. In particular, we aggregated mining and construction and combined the small number of firms classified in the public administration industry with firms in the service industry and used that category as the

omitted category in our regressions. As a result, the coefficients on the industry-specific dummy variables should be interpreted as the difference in the likelihood of default and prepayment from the likelihood for the service category. Table 5 shows how loans in SBA's database are distributed among categories defined by single-digit SIC codes.

Table 4: Multinomial Logistic Regression Coefficient Estimates[a]

Variables	Base model	Base+ T-bill	Base + SIC codes	Base + SIC + T-bill
Predicting to defaults				
Constant	-9.765	-9.903	-9.958	-10.078
i1	2.115	2.116	2.110	2.111
i2	3.117	3.119	3.109	3.110
i3	3.816	3.819	3.806	3.809
i4	4.225	4.228	4.213	4.216
i5	4.519	4.523	4.506	4.510
i6	4.666	4.671	4.655	4.660
i7	4.749	4.755	4.737	4.742
i8	4.821	4.828	4.811	4.817
i9	4.807	4.815	4.798	4.805
i10	4.812	4.821	4.803	4.811
i1112	4.803	4.813	4.795	4.804
i1314	4.777	4.789	4.769	4.780
i1516	4.710	4.723	4.703	4.715
i1718	4.621	4.634	4.616	4.628
i1920	4.614	4.625	4.611	4.620
i2122	4.516	4.526	4.509	4.518
i2324	4.430	4.441	4.421	4.431
i2526	4.295	4.308	4.292	4.304
i2728	4.341	4.354	4.332	4.343
i2930	4.252	4.263	4.231	4.242
i3132	4.204	4.216	4.198	4.209
i3334	4.138	4.151	4.131	4.143
i35p	4.103	4.121	4.084	4.100
t5_10	-0.046[a]	-0.046[a]	-0.064	-0.063
t10_15	-0.760	-0.761	-0.738	-0.739
t15p	-0.740	-0.739	-0.709	-0.708
sub1027	-0.580	-0.565	-0.553	-0.541
Loan_amt	0.258	0.259	0.278	0.279
Corporation	-0.043	-0.043	-0.084	-0.083
Partnership	-0.198	-0.199	-0.199	-0.199
Northeast	0.361	0.365	0.355	0.358
Midwest	0.218	0.224	0.210	0.215
South	0.414	0.418	0.433	0.436

(Continued From Previous Page)

Variables	Base model	Base+ T-bill	Base + SIC codes	Base + SIC + T-bill
Lender_PLP	-0.176	-0.171	-0.175	-0.170
Lender_CLP	-0.169	-0.171	-0.176	-0.177
New business	0.277	0.279	0.278	0.279
Urate	0.104	0.107	0.102	0.104
pc_gdp96	-0.126	-0.129	-0.124	-0.126
T-bill		0.022		0.020
Agri_etc			-0.537	-0.537
Mine_Const			0.306	0.306
Manuf			0.319	0.318
Wholesale			0.202	0.201
Trans_etc			0.208	0.208
Retail			0.443	0.443
Finan_etc			-0.146[a]	-0.145[a]
Predicting to prepayments				
Constant	-5.276	-4.917	-5.293	-4.932
i1	1.120	1.119	1.121	1.119
i2	1.602	1.597	1.603	1.598
i3	1.937	1.931	1.937	1.930
i4	2.106	2.097	2.108	2.098
i5	2.287	2.275	2.288	2.275
i6	2.411	2.398	2.413	2.398
i7	2.581	2.565	2.580	2.564
i8	2.708	2.691	2.709	2.691
i9	2.816	2.797	2.817	2.797
i10	2.913	2.893	2.913	2.892
i1112	3.054	3.032	3.055	3.032
i1314	3.144	3.117	3.146	3.118
i1516	3.311	3.281	3.312	3.282
i1718	3.455	3.427	3.456	3.427
i1920	3.695	3.670	3.694	3.669
i2122	3.520	3.497	3.521	3.497
i2324	3.669	3.644	3.668	3.642
i2526	3.822	3.793	3.823	3.792
i2728	4.011	3.982	4.010	3.980
i2930	3.614	3.587	3.613	3.584
i3132	3.714	3.685	3.714	3.684

(Continued From Previous Page)

Variables	Base model	Base+ T-bill	Base + SIC codes	Base + SIC + T-bill
i3334	3.791	3.760	3.789	3.756
i35p	3.995	3.952	3.992	3.948
t5_10	-0.657	-0.659	-0.652	-0.654
t10_15	-1.101	-1.100	-1.092	-1.091
t15p	-1.101	-1.101	-1.091	-1.091
Sub1027	0.081	0.048[a]	0.078	0.045[a]
Loan_amt	0.119	0.119	0.110	0.110
Corporation	0.099	0.097	0.091	0.089
Partnership	0.021[a]	0.024[a]	0.020	0.022
Northeast	-0.205	-0.214	-0.206	-0.215
Midwest	-0.187	-0.200	-0.186	-0.200
South	-0.093	-0.101	-0.091	-0.099
Lender_PLP	0.082	0.072	0.085	0.075
Lender_CLP	-0.001[b]	0.003[b]	-0.001[b]	0.003[b]
NewBusiness	0.068	0.064	0.077	0.073
Urate	-0.096	-0.103	-0.096	-0.104
Pc_gdp96	0.066	0.076	0.065	0.076
Tbill		-0.059		-0.059
Agri_etc			0.012[b]	0.010[b]
Mine_Const			0.058	0.059
Manuf			0.029[b]	0.032[a]
Wholesale			0.077	0.080
Trans_etc			0.138	0.139
Retail			-0.004[b]	-0.004[b]
Finan_etc			0.057[a]	0.056[a]
Summary statistics for multinomial logistic regression models				
N of Observations	5,736,628	5,736,628	5,710,096	5,710,096
Likelihood Ratio Chi Sq	120,478	121,081	121,718	122,318
Degrees of Freedom	76	78	90	92
Significance levels	<.0001	<.0001	<.0001	<.0001

Source: GAO.

Note: Models with SIC codes are based on a smaller number of cases due to missing SIC values.

[a]Except as noted, significance of coefficients is less than or equal to .0001. Significance of coefficients marked (a): < .05; those marked (b) had significance greater than .05.

[b]Models including SIC codes are based on a smaller number of cases due to missing SIC values.

Table 5: Distribution of SIC Industry Codes in SBA's Loan Database Distribution of SIC Industry Codes in SBAs Loan Database

SIC industry codes	N of loans	Percent of loans
Agriculture, forestry, fishing	12,280	3.0
Mining and construction	22,349	5.5
Manufacturing	49,807	12.3
Wholesale trade	29,464	7.3
Transport, communication, utilities	13,276	3.3
Retail trade	130,278	32.3
Finance, insurance, real estate	6,054	1.5
Service industries	134,623	33.4
Public administration	160	0.0
Missing	5,252	1.3
Total	**403,543**	**100.0**

Source: GAO.

The coefficients for the interest rate on 1-year Treasury bills are positive and highly significant for the default equations, as expected, and negative and highly significant for the prepayment equation. Most of the coefficients for the industry-specific dummy variables are also statistically significant. As can be seen in Table 4, the coefficients for most of the other variables in the equations are not much different in the alternative specifications from their values in SBA's equations.

SBA's Recovery Equation

SBA uses an ordinary least squares regression equation to estimate the relationship between the cumulative net recovery rate for a cohort of loans and the age of the loans in that cohort. This equation differs from the default and prepayment equations in that there are no economic or programmatic variables. As a result, forecasted recoveries on new loans will follow the historical pattern of recoveries on previously disbursed loans and will not depend on forecasted economic conditions. In addition, the unit of analysis is the cohort of loans rather than individual loans. The recovery equation uses ordinary least squares to regress the cumulative net recovery rate on a set of dummy variables for the age of the cohort. The cumulative net recovery rate is defined as cumulative net recoveries to date divided by cumulative defaults to date. Each dummy variable covers two

quarters ranging from quarters 1 and 2 to quarters 55 and 56. As expected for a cumulative dependent variable, the coefficients are generally increasing. In addition, except for the variable indicating the first two quarters, they are highly statistically significant. The adjusted R^2 is .9776, showing a good fit. Table 6 gives the names and descriptions for variables in the recovery equation while table 7 shows the coefficients for that equation.

Table 6: Variable Names and Descriptions

Variable name	Variable description
Cohort age 1-2	1 if in quarters 1 or 2, else 0
Cohort age 3-4	1 if in quarters 3 or 4, else 0
Cohort age 5-6	1 if in quarters 5 or 6, else 0
Cohort age 7-8	1 if in quarters 7 or 8, else 0
Cohort age 9-10	1 if in quarters 9 or 10, else 0
Cohort age 11-12	1 if in quarters 11 or 12, else 0
Cohort age 13-14	1 if in quarters 13 or 14, else 0
Cohort age 15-16	1 if in quarters 15 or 16, else 0
Cohort age 17-18	1 if in quarters 17 or 18, else 0
Cohort age 19-20	1 if in quarters 19 or 20, else 0
Cohort age 21-22	1 if in quarters 21 or 22, else 0
Cohort age 23-24	1 if in quarters 23 or 24, else 0
Cohort age 25-26	1 if in quarters 25 or 26, else 0
Cohort age 27-28	1 if in quarters 27 or 28, else 0
Cohort age 29-30	1 if in quarters 29 or 30, else 0
Cohort age 31-32	1 if in quarters 31 or 32, else 0
Cohort age 33-34	1 if in quarters 33 or 34, else 0
Cohort age 35-36	1 if in quarters 35 or 36, else 0
Cohort age 37-38	1 if in quarters 37 or 38, else 0
Cohort age 39-40	1 if in quarters 39 or 40, else 0
Cohort age 41-42	1 if in quarters 41 or 42, else 0
Cohort age 43-44	1 if in quarters 43 or 44, else 0
Cohort age 45-46	1 if in quarters 45 or 46, else 0
Cohort age 47-48	1 if in quarters 47 or 48, else 0

(Continued From Previous Page)

Variable name	Variable description
Cohort age 49-50	1 if in quarters 49 or 50, else 0
Cohort age 51-52	1 if in quarters 51 or 52, else 0
Cohort age 53-54	1 if in quarters 53 or 54, else 0
Cohort age 55-56	1 if in quarters 55 or 56, else 0

Sources: SBA and OFHEO.

Table 7: Recovery Model

Variable	Coefficient	Standard error	T - statistic
Cohort age 1-2	.0134	.0298	0.45
Cohort age 3-4	.0495	.0081	6.11
Cohort age 5-6	.0478	.0083	5.79
Cohort age 7-8	.0656	.0083	7.95
Cohort age 9-10	.0821	.0086	9.56
Cohort age 11-12	.1096	.0086	12.76
Cohort age 13-14	.1356	.0090	15.12
Cohort age 15-16	.1706	.0090	19.01
Cohort age 17-18	.1994	.0094	21.20
Cohort age 19-20	.2263	.0094	24.06
Cohort age 21-22	.2535	.0099	25.56
Cohort age 23-24	.2806	.0099	28.30
Cohort age 25-26	.3077	.0105	29.26
Cohort age 27-28	.3359	.0105	31.94
Cohort age 29-30	.3661	.0112	32.56
Cohort age 31-32	.3897	.0112	34.66
Cohort age 33-34	.4066	.0121	33.48
Cohort age 35-36	.4271	.0121	35.17
Cohort age 37-38	.4327	.0133	32.53
Cohort age 39-40	.4499	.0133	33.82
Cohort age 41-42	.4480	.0149	30.12
Cohort age 43-44	.4622	.0149	31.07
Cohort age 45-46	.4624	.0172	26.92
Cohort age 47-48	.4746	.0172	27.63
Cohort age 49-50	.4860	.0210	23.11

(Continued From Previous Page)

Variable	Coefficient	Standard error	T - statistic
Cohort age 51-52	.4982	.0210	23.68
Cohort age 53-54	.5099	.0298	17.14
Cohort age 55-56	.5192	.0298	17.45
Summary statistics			
Adjusted R^2	.9776		
Observations	393		

Source: GAO.

Comments from the Small Business Administration

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

U.S. SMALL BUSINESS ADMINISTRATION
WASHINGTON, D.C. 20416

OCT 17 2003

Ms. Davi D'Agostino, Director
Financial Markets and Community Investments Division
General Accounting Office
Washington, DC

Dear Ms. D'Agostino,

Thank you for the opportunity to review and comment on GAO's report entitled "Model For 7(A) Program Is Reasonable But Could Be Enhanced." We feel the report presents a thorough description of the model and the work that GAO was asked to complete.

As you know, SBA calculated significant re-estimates that resulted in large transfers of funds to the US Treasury from 1992-2001 for the 7(a) General Business Program. These re-estimates occurred primarily because SBA was using a historical average to model defaults. This method inherently reduces the ability to predict accurately in times where economic and program changes are taking place that do not repeat historical trends.

As a result, SBA began this project with the mission of building a model that would result in more accurate estimates and lower levels of re-estimates in the future. SBA felt strongly that the model should include assumptions about the economy since these appeared to be a cause for overestimating defaults. SBA also wanted the model to be useful in making decisions on programmatic changes.

In order to implement this mission, SBA management decided to hire contract support to provide additional expertise and ensure an objective process and result. SBA's choice of OFHEO allowed us access to experienced research economists with a past history in modeling loan programs at a very reasonable cost to the government.

By building a model that produces reasonable results based on well documented economic theory, SBA feels that the mission has been accomplished, and the results of the improved model will continue to be evident over the coming years. However, SBA also recognizes that work in this area is an ongoing effort. Periodic review is a constant necessity in order to ensure the models reflect the 7(a) program, and incorporating new data can lead to better knowledge of borrower and lender behavior.

GAO has made 2 recommendations and SBA agrees with them both. On the first recommendation, SBA intends to review any additional data that becomes available to assess if it would be useful in enhancing the accuracy of the model. Our Deputy CFO and subsidy team have been involved in reviewing the credit scoring product acquired by

SBA IS AN EQUAL OPPORTUNITY EMPLOYER AND PROVIDER

the Agency for loan monitoring. As the data becomes available, we will analyze it and assess its appropriateness for inclusion in the model.

GAO has also recommended that SBA establish a process for revising the model to correct errors and reflect changes in the 7(a) program. In FY 2003, SBA established an annual schedule for updating the coefficients used in the formulae based on additional years of data. We will continue to have these updates validated by an independent validation party, as is part of our internal controls process. We also intend to improve the automation of these models which will serve to decrease the potential for human error. SBA and OMB agreed to correct the error identified in GAO's review in the 2004 Budget Request, and did so as of October 1, 2003.

We appreciate the opportunity to comment on this report.

Sincerely,

Tom Dumaresq
Chief Financial Officer

U.S. SMALL BUSINESS ADMINISTRATION
WASHINGTON, D.C. 20416

FEB 19 2004

Davi M. D'Agostino
Director
Financial Markets and Community Investment
United States General Accounting Office
Washington, DC

Dear Ms. D'Agostino,

Thank you for this opportunity to comment on the draft GAO report on the Small Business Administration's (SBA) 7(a) program subsidy estimate ("Model for 7(a) Program Subsidy Estimate Had Reasonable Equations but Lack of Key Documentation Hampered Review"). SBA has the following comments.

- **Page 1**

See comment 1.

GAO states on the first page under "What GAO found," that SBA "did not adequately document its model development Process". This statement should be qualified, as it is later in the document, by making it clear that there is no requirement for SBA to document the model development process as GAO recommends.

- **Page 1**

"SBA officials told us that the new 7a model was the first step in a long term effort to develop and implement new econometric models for their credit programs."

See comment 2.

Comment: SBA developed the 7a model using econometrics for many reasons, including the fact that there was a strong feeling in SBA and outside that the economy affects loan performance. However, SBA also knew that other factors affected loan performance and needed to be identified along with the economic factors. As such, SBA would like the following sentence added: **Although this allowed SBA to build a model that responds to the need for greater sensitivity to a wider variety of factors than a model based on historical averages, this approach may not be appropriate for all the credit programs.**

- **Page 7**

See comment 3.

SBA disagrees with the new section about the lack of documentation. Please add: **GAO was given 800 pages of data with information on variables that were considered and rejected. SBA supplemented this extensive document with many hours of briefings and explanations.**

Federal Recycling Program Printed on Recycled Paper

- **Page 7**

"However, maintaining documentation on how such models were developed is a sound internal control practice that would provide SBA and other agencies the opportunity to demonstrate and explain the rationale and basis for key aspects…"

See comment 4.

Comment: SBA would like GAO to add the words "**more fully**" after "opportunity to demonstrate". SBA feels that it has demonstrated and explained the rationale adequately according to the current formal and informal guidance. SBA followed the guidance that is available and a full explanation of the variables that SBA chose is supported by the documentation.

- **Page 25**

See comment 5.

SBA briefed GAO thoroughly on the issue of variables considered and rejected. GAO was provided with a list of the variables, and the reasons they were rejected, as well as the 800 page testing results document that was reviewed as a part of the government-wide financial statement audit.

- **Page 25**

See comment 6.

Add: **SBA hired an independent contractor in 2002 to review the model prior to its finalization as part of its review and validation process. This occurred prior to the completion of the documentation.** "The independent contractor hired to perform an initial review of the SBA 7(a) credit subsidy model prior to its finalization was hampered by the lack of detailed model documentation. In response to our inquiry, the contractor stated that it did not validate the model which, from an audit perspective, would have encompassed a more robust effort. In its final report to SBA, the contractor reported that SBA lacked sufficient supporting documentation for a "thorough review of its [the model's] theoretical basis (including alternative modeling methodologies explored), its working features, or the update and maintenance procedures necessary to use the model on an ongoing basis. This lack of documentation severely limited our ability to assess certain critical parts of the model in detail, including its econometric components." Further, the contractor recommended that "SBA develop a robust set of documentation to support this model" including "the modeling methodology, alternate methodologies considered, data inputs and outputs, and model maintenance and update requirements."

See comment 7.

"Nevertheless, **GAO believes** that maintaining sufficient documentation on how such models were developed is a sound internal control practice that would provide SBA and other agencies the opportunity to demonstrate and explain the rationale and basis for key aspects of their models that provide important cost information for budgets, financial statements, and congressional decision makers. Moreover, as a practical matter, this documentation would help facilitate SBA's and other agencies' annual financial statement audits".

Comment: Please add the words in bold.

- **Page 28**

See comment 8.

See comment 9.

See comment 10.

See comment 11.

GAO makes a statement that "SBA's lack of documentation on the 7(a) model process could impede our ability to conclude on SBA's loan accounts in connection with audit of the consolidated financial statements of the federal government."

Comment: SBA strongly disagrees with this particular comment because it establishes a new and unnecessary requirement. SBA has followed the available guidance, including Tech release 6 (formerly this was covered by Tech release 3) and SFFAS statements 2 and 18. GAO agrees that the models produce a reasonable result. GAO has not proven bias in either direction.

SBA believes that GAO's definition of "independent person" needs to be clarified to include and "be an informed reader who is familiar with statistical and econometric analysis, and the tools used to perform this activity."

Please add in the conclusion section: "SBA chose an independent party to ensure that bias from selection did not exist, and further that SBA's tests show that there is no identifiable bias in this model. It should be noted however that some degree of bias can exist in all forecasting models, including those using a simple historical average."

Thank you again for this opportunity to comment on the draft report.

Sincerely,

Thomas A. Dumaresq
Chief Financial Officer

The following are GAO's comments on the Small Business Administration's letter dated February 19, 2004.

GAO Comments

1. Highlights page was adjusted to reflect SBA's position.

2. We adjusted the report text to recognize SBA's position. See page 1.

3. We adjusted the text to reflect that SBA subsequently provided us access to this documentation and provided a description of the documentation as well as an assessment of its usefulness in assessing the model development process. See page 26.

4. We adjusted the text of the report. See page 7.

5. We acknowledge that SBA briefed us on the variables that were selected and rejected and that we could not corroborate this with the supporting documentation that SBA provided. See the Agency Comments and Our Evaluation section of the report pages 35-36.

6. We do not concur with SBA that this change should be made to the report because it is redundant with the information provided on pages 24 and 25.

7. We adjusted the report text to clarify our position.

8. We do not concur with SBA. See the Agency Comments and Our Evaluation section of the report on page 36.

9. We concur with SBA's assertion that we have not proven that the model had a bias. Our report states that we were unable to determine whether such a bias existed because of SBA's insufficient documentation.

10. We concur with SBA's definition of an independent person in the context of this report and point out our team that reviewed the 7(a) model met SBA's definition of an independent person. However, any revisions of the definition of an independent person would need to be made by the Federal Accounting Standards Advisory Board.

11. We do not concur with SBA's statement that an independent party ensured that the 7(a) model was free from bias from variable selection. As we discussed, neither Bearing Point nor Ernst and Young, both of

which SBA asserted were independent reviewers who ensured the model was free of bias, assessed the variable selection process. Bearing Point reported that its review was severely limited by the lack of documentation and did not assess the econometric segment of the model. Ernst and Young reported that, at the request of SBA, it did not assess the econometric component of the model. Thus, neither of these firms could assess whether a bias existed from the variable selection process.

We also do not concur with SBA's statement that its tests show that there was no identifiable bias in the model. While SBA may have tested its final model for bias, the agency has not provided us with any supporting documentation of these analyses. Further, testing the model would not identify this type of bias. Rather, an analysis of the variable selection process and whether it was consistently applied to all variables tested would more likely reveal whether such a bias existed in the final model.

We also do not concur with SBA's suggested change to the conclusions of our report regarding whether a possible bias existed in the final model. The bias that is described in our report would result from variable selection or rejection. SBA discusses a statistical bias that suggests that over the historical period the chosen model systematically either under predicts or over predicts the likelihood of defaults or prepayments. To provide reasonable assurance that a bias was not introduced into the subsidy rate estimate through the choice of particular equations from among the set of reasonable equations, adequate documentation of the basis for selecting and rejecting variables is an important internal control. We were unable to determine whether this type of bias existed because of the lack of documentation on the model development process.

Comments from the Office of Management and Budget

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

FEB 18 2004

Ms. Davi M. D'Agostino
Director
Financial Markets and Community Investment
United States General Accounting Office
Washington, DC 20548

Dear Ms. D'Agostino,

Thank you for this opportunity to comment on the draft General Accounting Office (GAO) report on the Small Business Administration's (SBA) 7(a) program subsidy estimate ("Model for 7(a) Program Subsidy Estimate Had Reasonable Equations but Lack of Key Documentation Hampered Review").

The Office of Management and Budget (OMB) was pleased that GAO's draft report concluded that SBA's new "econometric equations were reasonable, and its model produced estimated default and recovery rates that were in line with historical experience."[1] As we explained in our interviews with GAO during this review, OMB has worked closely and diligently with SBA staff on the development of the new model, which we believe is a vast improvement to the prior cost estimating techniques used by SBA. However, as we explain below, OMB disagrees with the recommendation that we revise OMB Circular A-11, as we believe that the draft report does not demonstrate that a revision is needed.

See comment 1.

The draft report recommends that OMB revise Circular A-11 to require agencies to document the development of their credit subsidy models, including "processes for selecting modeling methodologies over alternatives and variables tested and rejected along with the basis for excluding them."[2] According to the draft report, this documentation would facilitate OMB and external party model review and financial statement audits. OMB agrees that model documentation is important for both budget and financial statement purposes. We do not believe, however, that the draft report findings demonstrate that more documentation is needed.

See comment 2.

Consistent with our practice for all Federal credit agencies and to fulfill the responsibilities of the OMB Director under the Federal Credit Reform Act of 1990, as amended,[3] we worked closely with SBA during the model development process. Accordingly, we believe that the documentation SBA provided to OMB was adequate to determine that the subsidy estimates and reestimates published in recent President's Budgets are reasonable. In addition, as SBA explains in their letter on the draft report to

[1] GAO draft report, Highlights page and p. 5.
[2] GAO draft report, p. 37.
[3] 2 U.S.C. § 661.

you, Ernst & Young independently validated the model with available documentation from SBA. Ernst & Young found that "the 7(a)'s Model's assumptions and methodology appear to be reasonable and accurate" and that there was "reasonable accuracy of model calculations and output."[4] GAO has confirmed the results of these reviews in its report, finding that the "econometric equations were reasonable."[5] Consequently, we do not concur with the draft report's statement that "lack of improved OMB guidance for model documentation hampers adequate external oversight and validation of models used to generate credit subsidy estimates."[6]

See comment 3.

Within the draft report, GAO also referenced the requirements in the Statement on Auditing Standards (SAS) No. 57 as the basis for reporting that "SBA did not prepare adequate supporting documentation to enable independent reviewers to understand and evaluate the process that SBA used." The standard, however, explains that "[m]anagement is responsible for establishing a process for preparing accounting estimates. Although the process may not be documented or formally applied, it normally consists of …[i]dentifying the relevant factors that may affect the accounting estimate… [d]eveloping assumptions that represent management's judgment of the most likely circumstances and events with respect to the relevant factors…[d]etermining the estimated amount based on the assumptions and other relevant factors." We believe that SBA has fulfilled the management responsibilities as stated in SAS No. 57, as validated by Ernst & Young, and as referenced within the GAO draft report, thus providing independent reviewers with adequate documentation to understand and evaluate its estimates.

See comment 4.

Further, to require agencies to prepare additional documentation of the "variables tested and rejected along with the basis for excluding them"[7] would be unduly burdensome. Since the draft report indicates that your staff received and tested those final decisions, namely, the model itself and its estimated default and recovery rates and determined that they were reasonable, we do not think the draft report adequately makes the case that a change is needed. Finally, GAO received the underlying loan data used as a basis for model development. With these data, GAO has the information necessary to test alternative variables to measure the reasonableness of the model.

See comment 5.

OMB agrees that agencies should be encouraged to keep adequate documentation of complex statistical models to assist in their review and improvement over time. However, OMB recognizes that agencies should have discretion in the level of record-keeping that is required, based on the importance of the models, ease of replication of results, and other factors.

[4] Ernst & Young report "Task 6: Independent Review of SBA 7(a) Subsidy Model, Part 2;" November 18, 2003, p. 2.
[5] GAO draft report, Highlights page and p. 5.
[6] GAO draft report, p. 36.
[7] GAO draft report, p. 37.

See comment 6.

The draft report states that a revision of OMB's Circular A-11 would facilitate "external financial statement audit[s]."[8] We believe that this rationale misconstrues the fundamental nature and purpose of the Circular. Circular A-11 provides guidance on issues affecting the Budget, including both formulation and execution of credit subsidy estimates. The Circular does not provide guidance on internal control as it relates to financial statement audits. Accordingly, even if we agreed with the draft report that additional guidance is needed, which we do not, we believe that Circular A-11 would not be the appropriate location to include instruction on matters affecting agency financial statement audits.

See comment 7.

Moreover, such instruction is already available in standards and guidance elsewhere. For example, the GAO report references the Federal Accounting Standards Advisory Board's ("FASAB") Technical Release 6, "Preparing Estimates for Direct Loan and Loan Guarantee Subsidies Under the Federal Credit Reform Act" ("TR6"). This guidance addresses internal control and the proper documentation an agency should maintain to support the assumptions used in subsidy calculations (paragraphs 20-22). TR6 was developed by an interagency task force, whose members included GAO, OMB, and Federal credit agencies, under the auspices of the Accounting and Auditing Policy Committee of FASAB. This guidance is authoritative, falling within Level C on the hierarchy of the Federal generally accepted accounting principles (footnote 21 on page 29 of GAO's draft report should be changed to reflect that TR6 is, in fact, authoritative). This guidance specifically outlines requirements to document "key assumptions" of models, that is, those which have the greatest effect on the subsidy estimate. We believe that the documentation provided by SBA, which includes documentation of the key assumptions underlying the 7(a) subsidy estimate, satisfies the TR6 documentation requirements.

In closing, we want to reaffirm that OMB takes seriously its responsibilities in overseeing the Federal Credit Reform Act. However, the draft report does not demonstrate that a revision to the procedures set forth in OMB's Circular A-11, or TR6, is necessary.

Thank you again for this opportunity to comment on the draft report.

Sincerely,

Linda M. Springer
Controller

Richard P. Emery
Assistant Director for Budget

[8] GAO draft report, p. 29.

The following are GAO's comments on the Office of Management and Budget's letter dated February 18, 2004.

GAO Comments

1. We do not concur with OMB and believe that in light of the consistent difficulty experienced by three independent reviews of SBA's 7(a) model, our report makes a case for the need to enhance the guidance in Circular A-11 to require agencies to document the process they used to develop the model. See the Agency Comments and Our Evaluation section of the report pages 36-37.

2. We do not concur with OMB. See Agency Comments and Our Evaluation section of the report page 37.

3. We do not concur with OMB. See Agency Comments and Our Evaluation section of the report pages 37– 38.

4. We do not concur with OMB. See Agency Comments and Our Evaluation section of the report page 38.

5. While we concur that agencies need to have discretion in the level of documentation that they maintain when dealing with inconsequential matters, we do not agree that such discretion should be allowed in clearly consequential activities such as the development of the 7(a) model. We reaffirm our position that OMB needs to enhance its guidance regarding the need for adequate documentation for the credit subsidy model development process.

6. We concur with OMB that the fundamental nature and purpose of Circular A-11 is not to provide guidance on internal controls as it relates to financial statement audits. However, the primary focus of this report is on credit subsidy estimates which are prepared in accordance with Circular A-11. Also, the financial statement audit is an important validation of the credit subsidy estimates included in the budget. We reaffirm our conclusion and recommendation that enhanced guidance on credit subsidy model development would facilitate external review, including those performed by OMB, of the credit subsidy estimate. Because of the relationship between the credit subsidy estimates prepared for the budget and those used in the financial statements, the enhanced guidance would benefit both the financial statement audit and budgetary review.

7. Report language was revised to address technical points about Technical Release 6. However, as we discussed, this guidance does not specifically require documentation of credit subsidy model development.

GAO Contacts and Staff Acknowledgments

GAO Contacts

Davi M. D'Agostino (202) 512-8678
M. Katie Harris (202) 512-8415

Staff Acknowledgments

In addition to those individuals named above, Jay Cherlow, Dan Blair, Edda Emmanueli-Perez, Mitch Rachlis, Marcia Carlsen, Beverly Ross, Susan Sawtelle, and Mark Stover made key contributions to this report.

GAO's Mission

The General Accounting Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to e-mail alerts" under the "Order GAO Products" heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Public Affairs

Jeff Nelligan, Managing Director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G Street NW, Room 7149
Washington, D.C. 20548

PRINTED ON RECYCLED PAPER

United States
General Accounting Office
Washington, D.C. 20548-0001

Official Business
Penalty for Private Use $300

Address Service Requested

Presorted Standard
Postage & Fees Paid
GAO
Permit No. GI00