Appendix for
"Habit Persistence and Keeping Up with the Joneses: Evidence from Micro Data"

November 2005

Appendix I

IA Sample Selection

The data set used in the analysis is a random sample of credit card accounts, active and not delinquent as of July 1999. Each account is followed for the period between the third quarter of 1999 and the third quarter of 2002. The unit of observation is the account in a given quarter. The card could be used by more than an individual and therefore I will refer to the decision maker behind it as household (HH).

Although the original sample covers the entire U.S. territory, I restrict the analysis to California, the only state for which retail sales are available at the city level and quarterly frequency. For the same reason, although the credit card data are available at monthly frequencies, I construct quarterly variables, a choice that has also the benefit of reducing the noise that plagues individual monthly consumption data. In constructing the sample, I exclude people whose accounts are inactive and those that don't use the card very often, in order to obtain a more meaningful measure of consumption. For an account to be in the sample the expenditure can never fall below $50 in any given quarter. The choice of this cutoff is meant to compromise between sample size and representativeness. According to Lim and Benjamin (2001), the average transaction amount on a credit card is $87, 112% higher than that made in cash. Following the literature, I also exclude retired account holders and people living in military areas, because their expenditures are influenced by special conditions and required specific modelling that is outside the scope of this paper. After this selection procedure, the sample contains 2,674 accounts.

Extreme outliers. Some of the series present a lot of variability. In order to address this issue, part of the previous literature excludes observations in which the growth rate of consumption is too large. Zeldes (1989) for example excludes observations in which the growth rate of expenditures is bigger than 1.1; Brav et al. (2002) exclude observations for which the growth rate goes from less than 1/2 to more than 2 or it is bigger than 5; Vissing-Jorgensen (2002) excludes observations for which the growth rate goes from less than 0.2 or more than 5. I keep all the observation because the distributions of expenditure growth rates is very symmetric and the cutoff points seem somewhat arbitrary. I have also tried to exclude observations whose growth rate is bigger than 5 and I get very similar results. Similarly, I have tried winsorizing the city retail sales data at the 5% and 95% cutoff points and obtained similar estimates.

Missing data. I exclude from the sample, for a given quarter, any observation with missing data on any of the variables included in the basic regression (including income). One of the reasons for missing observations is bankruptcy. In particular, 0.92% of the accounts ends up bankrupt at the end of the sample period. This makes the panel unbalanced, but doesn't biased the results, since in the estimation I control the financial health of the account through debt outstanding, amount charged off and credit constrained indicators.

IB Comparison with the U.S. Census 2000: Analysis of Demographic Characteristics and Borrowing Behavior

Table AI and Figure AI compare the distributions of income and age in the data set to those in the U.S. Census 2000. Both the distributions are very alike. The main discrepancy is due to the fact that HHs in the lower range of income and individuals in very young or old age are under-represented in the credit card data set. However, this phenomenon is to be expected given the nature of the variables analyzed: as documented in the Survey of Consumer Finances, HHs with lower income and whose head is younger than 35 or older than 65 are less likely to hold bank-type credit cards. For the same reason, the proportion of HHs that own the house in which they live is 74.88% in my data set, compared to the 66.2% in the Census. Finally, the percentage of married people is slightly less than the national average, 44.76% versus 52.4%, even though the sizeable amount of missing data, 35.27%, could be the cause of the difference. Table AII illustrates the breakdown of the data set by occupation: the "professional/technical" category is the most widely represented, accounting for 13.76% of the observations, followed by the "administrative/managerial" with 7.59%. An interesting feature related to this variable is that 2.51% of the HHs in the sample are headed by self-employed individuals.

The sample compares well to the U.S. data with regard to HHs borrowing behavior as well. Panel A of Table AIII contains a comparison between my data set and a large multi-issuer credit card data set covering the period between 1995 and 1998 and used by Gross and Souleles (2002). Indebtedness is highly skew in both data sets; while the median debt outstanding is $0 and $70 respectively, the average debt outstanding is $1,486 for my data set and similar for Gross and Souleles. The figure more than doubles if we consider only those HHs that have debt outstanding, reaching $3,400. Both the mean and median credit limits are higher in my data set, probably reflecting the increase in credit availability over the period. Similarly, average and median interest rates are lower in the latter period, due to a trend in the reduction of interest rates that is reflected in the statistics on the rate changes as well.

Table AIII Panel B provides a comparison with the Survey of Consumer Finances. The samples are similar with regard to the percentage of people not paying the balance in full at the end of the month, which is estimated to be 44.4% in the SCF and 45.75% in my sample. Unfortunately, the different classification criteria cause some difficulties in comparing the statistics, as the SCF reports the total debt outstanding on all the credit cards available for the HH. Moreover, this survey has been proven to suffer from under-reporting of debt.

IC Credit Cards Expenditures as a Measure of Consumption

Data from the Consumer Expenditure Survey (CEX), show that in 2001 the average annual expenditure was $38,045. Following the selection criteria of Attanasio and Weber (1995), I construct an estimate of consumption of non-durable goods and services by excluding from the aggregate consumer durables, housing, health and education expenditures. This gives a quarterly expenditure between $4,765 and $5,053, depending on the classification criteria. According to the statistics presented above, 24% of this amount, between $1,143 and $1,212, is paid by credit card. The average expenditure on the accounts I analyze is $701.19. This means that on average the expenditures analyzed cover between 13.88 and 14.72 percent of non-durable goods and services consumption. This figure represents good news, because it indicates that on average the HHs in my data set use this credit card conspicuously and are therefore offering a good measure of their expenditures.

Appendix II

In this Section I model the consumption decisions of a household that borrows through a credit card and saves in a savings account. In each period t, household i chooses consumption $c_{i,t}$ and the amount $P_{i,t}$ of the credit card balance to pay back, with the objective of maximizing the expected value of a lifetime utility function:

$$\underset{\{c_{i,t};P_{i,t}\}_0^{T-1}}{Max} E_t \sum_{t=0}^{T-1} \beta^t U(c_{i,t}, h_{i,t}, H_{i,t}, \Theta_{i,t}) \tag{0.1}$$

subject to:

$$A_{i,t+1} = (A_{i,t} + Y_{i,t} - P_{i,t})R^f_{i,t} \tag{0.2}$$

$$B_{i,t+1} = (B_{i,t} - P_{i,t} + c_{i,t})\widetilde{R}^C_{i,t} \tag{0.3}$$

$$c_{i,t} \leq A_{i,t} + Y_{i,t} + \overline{B}_i - B_{i,t} \tag{0.4}$$

$$h_{i,t+1} = \zeta c_{i,t} \tag{0.5}$$

$$A_{i,t} \geq 0, \quad B_{it} \leq \overline{B_i}, \quad c_{i,t} > 0 \tag{0.6}$$

where $h_{i,t}$ is the level of the habit stock that HH i derives from its own past consumption, $H_{i,t}$ is the consumption of the reference group and $\Theta_{i,t}$ are demographic and socioeconomic characteristics; $A_{i,t}$ is the amount of money in the savings account at the beginning of period t, which earns the risk-free rate, $R^f_{i,t}$; $Y_{i,t}$ is the income realization; $B_{i,t}$ is the credit card balance at the beginning of period t, $\overline{B}_i$ the credit limit on the card, and $\widetilde{R}^C_{i,t}$ the gross interest rate the HH is charged on the balance outstanding. The value of $\widetilde{R}^C_{i,t}$ depends on whether or not the HH pays the balance in full and is given by the following expression:

$$\widetilde{R}^C_{i,t} = \begin{array}{ll} R^C_{i,t} > R^f_{i,t} > 1 & \text{if } P_{i,t} < B_{i,t} \\ 1 < R^f_t & \text{if } P_{i,t} \geq B_{i,t} \end{array}$$

Equation (0.2) describes the evolution of the savings account balance: at the end of period t, the account contains the initial funds, plus that period income, minus the credit card payment. Analogously, (0.3) shows that the credit card balance at the beginning of period t+1 consists of the unpaid balance from the previous period, $B_{i,t} - P_{i,t}$, and any new expenditure charged on the card. Equation (0.4) states that consumption cannot exceed the sum of the resources on hand and the unused part of the credit line. Equation (0.5) describes the evolution of the habit stock of the HH, which is assumed to depend on last period consumption only in order to make the empirical analysis more tractable. The specification of the evolution of local aggregate consumption and the demographic characteristics is not necessary for the derivation of the optimal consumption rule and is left for later.[1] Finally, equations (0.6) represents the no short sales constraint, the borrowing limit, and the condition that consumption must be strictly positive, respectively.

I solve the maximization problem delineated above by expressing it in recursive form. To make the model more tractable and help focusing on the key elements of the problem, I make the following simplifying assumptions.

Assumption 1 *Household income lies between $\underline{Y_i}$ and $\overline{Y}_i$ and evolves according to $Y_{i,t} = Y_i + \epsilon_{i,t}$, where Y_i is HH's i mean income and $\epsilon_{i,t}$ is i.i.d. and has mean 0 and variance σ^2.*

This assumption implies that in any given period the income realization doesn't depend on previous history. Despite being restrictive, it is aimed at capturing the presence of uninsurable income shocks that causing fluctuations around the income level expected by the household or, alternatively, random unexpected expenses. Exploring the effect of more general income processes constitutes material for future research.

Assumption 2 *The credit limit is set to be $\overline{B_i} = \sum_0^T \frac{\overline{Y_i}}{R^f_{i,t}}$, the present value of an income stream in which the highest possible income $\overline{Y_i}$ is realized in every period.*

Assumption 3 *There is no default.*

[1] $H_{i,t}$ and $\Theta_{i,t}$ are therefore suppressed in the rest of the section for notational simplicity.

Notice that the optimal payment rule for the HH is to employ all the resources on hand to pay back as much as possible of the outstanding credit card balance. The intuition is that since the household will be charged for $c_{i,t}$ only in period t+1, paying the balance doesn't subtract any resource from current consumption. On the contrary, since there is a wedge between borrowing and lending rates, a bigger payment increases the wealth of the HH because it reduces the amount on which it is charged the high interest rate. Therefore, the only choice variable left for the household is $c_{i,t}$.

The solution of the problem delineated in (0.1)-(0.6) is complicated by the presence of a discontinuity in the value function at the point in which the HH switches from lending at the risk-free rate to borrowing at a higher interest rate. To solve this problem I use as state variable to be the amount of resources on hand, $Z_{i,t} = A_{i,t} + Y_{i,t} - B_{i,t}$. The problem can be then defined over two different regions: the first, region B, in which the HH has negative resources on hand and thus borrows at a high interest rate; and the second, region $\overline{B}$, in which it has enough resources to pay the balance in full, put some money into the savings account and earn the risk-free rate. Since the point of discontinuity coincides with the point in which the HH switches regions, within each region continuity and differentiability are satisfied and standard solution techniques can be applied.

More precisely, the problem of a HH that in period t is not borrowing can be expressed as follows:[2]

$$V_t^{\overline{B}}(Z_t,h_t) = \max_{\{c_t\}} u(c_t,h_t) + \beta\left[\int_{-Z_tR_t^f+c_t}^{\overline{Y}} V_{t+1}^{\overline{B}}(Z_{t+1},h_{t+1})f(Y)dY + \int_{\underline{Y}}^{-Z_tR_t^f+c_t} V_{t+1}^{B}(Z_{t+1},h_{t+1})f(Y)dY\right] + \lambda_t(Z_t+\overline{B}-c_t) \quad (0.7)$$

$$\text{subject to:} \qquad Z_{t+1} = Z_tR_t^f + Y_{t+1} - c_t \quad (0.8)$$

$$h_{t+1} = \zeta c_t \quad (0.9)$$

where $V_t^{\overline{B}}(Z_t,h_t)$ is the value function; the expression in brackets is the expectation of the future value function, taken with respect to next period income realization; and the last expression is the product of the Lagrange multiplier and the resource constraint. Equations (0.8) and (0.9) describes the evolution of the state variables. The cutoffs of the two integrals show that whether next period the HH falls in one region rather than the other is determined by the resources accumulated from the previous period, $A_{i,t}$, and the realization of the income shock: if $Y_{i,t+1}$ is high enough that the HH is able to repay the credit card balance in full, then it keeps staying in the non-borrowing region; otherwise, it will start borrowing. When the HH makes its consumption decision in period t, it knows that this choice will determine the credit card balance and therefore the likelihood of staying in the non-borrowing region. As time passes and the HH accumulates or decumulates resources, the probability of falling into a certain region gets smaller and smaller, independently from the income realization and the consumption choice.

The formulation of the problem of a HH that enters period t with an unpaid balance is very similar to (0.7).The only difference is the law of motion of the state variable:

$$V_t^{B}(Z_t,h_t) = \max_{\{c_t\}} u(c_t,h_t) + \beta\left[\int_{(-Z_t+c_t)R_t^C}^{\overline{Y}} V_{t+1}^{\overline{B}}(Z_{t+1},h_{t+1})f(Y)dY + \int_{\underline{Y}}^{(-Z_t+c_t)R_t^C} V_{t+1}^{B}(Z_{t+1},h_{t+1})f(Y)dY\right] + \lambda_t(Z_t+\overline{B}-c_t) \quad (0.10)$$

$$\text{subject to:} \qquad Z_{t+1} = Y_{t+1} - (-Z_t+c_t)R_t^C \quad (0.11)$$

$$h_{t+1} = \zeta c_t \quad (0.12)$$

The maximization problem delineated above generates the following Euler equations for a HH that at period t is in the non-borrowing region:

[2] From now on I suppress the subscript i, for notational simplicity.

$$u^c(c_t,h_t) = \beta \int_{-Z_t R_t^f + c_t}^{\overline{Y}} [u_{t+1}^c R_{t+1}^f + \beta R_{t+1}^f \zeta E_Y(u_{t+2}^h) - \zeta u_{t+1}^h] f(Y) dY \\ + \beta \int_{\underline{Y}}^{-Z_t R_t^f + c_t} [u_{t+1}^c + \beta \zeta E_Y(u_{t+2}^h) - \zeta u_{t+1}^h] f(Y) dY \quad (0.13)$$

and for one that is in the borrowing region:

$$u^c(c_t,h_t) = \beta R_t^C \int_{(-Z_t+c_t)R_t^C}^{\overline{Y}} \left[[u_{t+1}^c R_{t+1}^f + \beta R_{t+1}^f \zeta E_Y(u_{t+2}^h)] R_t^C - \zeta u_{t+1}^h \right] f(Y) dY \\ + \int_{\underline{Y}}^{(-Z_t+c_t)R_t^C} \left[[u_{t+1}^c + \beta \zeta E_Y(u_{t+2}^h)] R_t^C - \zeta u_{t+1}^h \right] f(Y) dY \quad (0.14)$$

These results are very intuitive. The household decides how much to consume today versus tomorrow by weighting future utility and different interest rates by the probability that it will actually face them. If the HH consumes \$1 less today it looses $u^c(c_t)$ and gains the following: next period the credit card balance will be \$1 lower and so one more dollar will be available for consumption, yielding a utility of $u^c(c_{t+1})$; if, in addition to this gain, the income realization is high enough that the HH is able to repay the balance in full, it will earn the gross risk-free rate on the dollar moved through time and the utility will be $u^c(c_{t+1})R_{t+1}^f$. These events realize with probabilities $\int_{\underline{Y}}^{(-Z_t+c_t)R_t^C} f(Y)dY$ and $\int_{(-Z_t+c_t)R_t^C}^{\overline{Y}} f(Y)dY$, respectively.

Analogously, in the case of (0.14), consuming one dollar less today means that the credit card balance next period will be R_t^C dollars less.[3] The utility deriving from this intertemporal transfer will be $u^c(c_{t+1})R_t^C$ if the HH doesn't have enough resources to pay the balance in full in period t+1, and $u^c(c_{t+1})R_t^C R_{t+1}^f$ in case it does and can invest the dollar charged on the credit card at the risk-free rate.

The presence of the habit stock generates an additional effect due to the fact that when the HH consumes one dollar less today it increases tomorrow's utility not only directly, but also by decreasing the habit level. This effect is given by $-\zeta u_{t+1}^h > 0$. [4] The extra dollar consumed in period t+1 will increase the habit stock of period t+2 at a cost in term of utility given by $\beta \zeta E_Y(u_{t+2}^h)$ or $\beta \zeta R_{t+1}^f E_Y(u_{t+2}^h)$, depending on the HH asset position.

It is also possible to draw a parallel with the Euler equation traditionally obtained in the literature, where the HHs are assumed to be able to borrow and lend at the risk-free rate. In particular, if in case $\overline{B}$ we assume that the resources on hand are high enough that the HH will always pay the balance in full, then (0.13) collapses to the usual Euler equation:

$$u^c(c_t) = \beta \left\{ \int_{\underline{Y}}^{\overline{Y}} u^c(c_{t+1}) R_{t+1}^f f(Y) dY \right\} \quad (0.15)$$

References

[1] Attanasio, Orazio, and Weber, Guglielmo. (1995), "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *The Journal of Political Economy*, Vol. 103, No. 6, pp. 1121-1157.

[2] Brav, Alon, Constantinides, George M., and Geczy, Christopher C. (2002), "Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence." *The Journal of Political Economy*, Vol. 110, No. 4, pp. 793-824.

[3] Gross, David B., and Souleles, Nicholas S. (2002), "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data" *The Quarterly Journal of Economics*, Vol. 117, No. 1, pp. 149-186.

[4] Lim, Paul J., and Benjamin, Matthew. (2001) "Digging Your Way Out of Debt," *U.S. News and World Report.*

[3] Since in period t the HH was not able to pay the balance completely it was charged interest on both the unpaid portion of the balance and any new purchases.

[4] Recall that an increase in the habit stock decreases utility and therefore $u^h < 0$.

[5] State Board of Equalization. *Taxable Sales in California,* various issues.

[6] State Board of Equalization. (1999) *Sales and Use Taxes: Exemptions and Exclusions*, No. 61.

[7] Vissing-Jorgensen, Annette. (2002), "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution." *The Journal of Political Economy*, Vol. 110, No. 4, pp. 825-853.

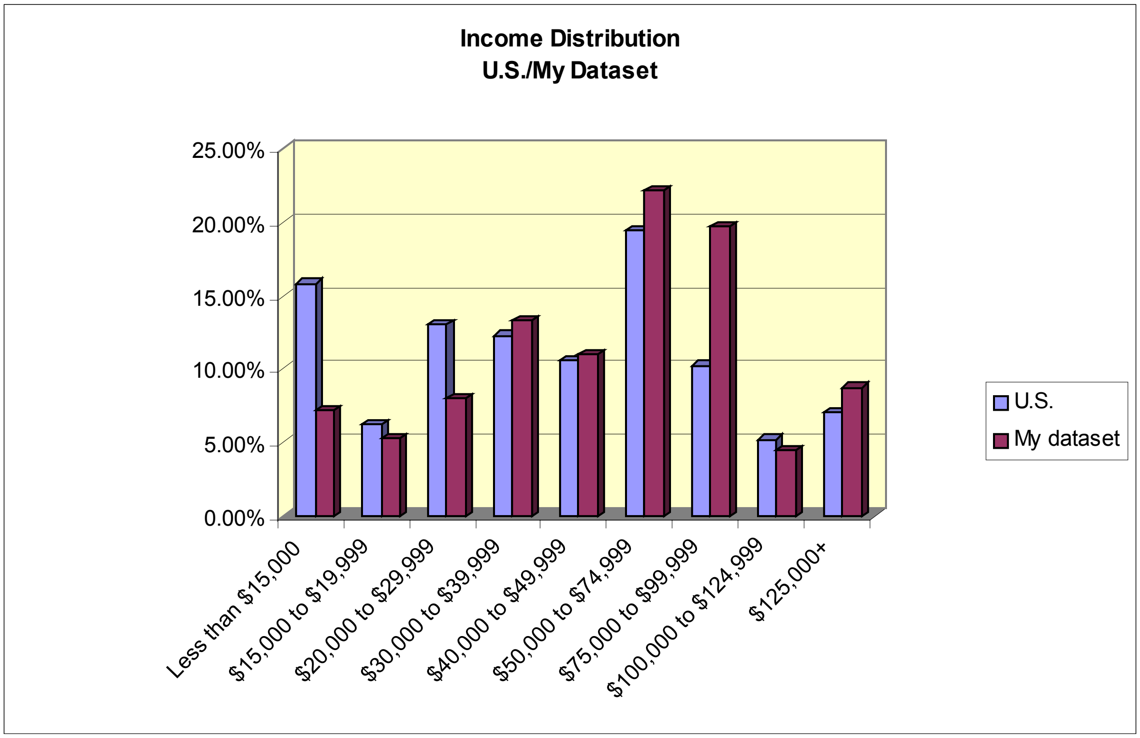
[8] Zeldes, Stephen P. (1989), "Consumption and Liquidity Constraints: An Empirical Investigation" *The Journal of Political Economy*, Vol. 97, No. 2, pp. 305-346.

Table AI
Demographic Characteristics
Comparison with U.S. Census data

This Table contains a comparison between the breakdown by age and income of my data set versus the U.S. Census.

	Age	
	My dataset CA	*U.S. Census*
20 to 24 years	7.11%	9.44%
25 to 34 years	17.61%	19.85%
35 to 44 years	24.91%	22.47%
45 to 54 years	24.01%	18.75%
55 to 59 years	8.12%	6.70%
60 to 64 years	5.12%	5.38%
65 to 74 years	9.16%	9.15%
75 to 84 years	3.48%	6.15%
85 years and over	0.49%	2.11%
Total	**100%**	**100%**
	Income	
	My dataset CA	*U.S. Census*
Less than $15,000	7.21%	15.85%
$15,000 to $19,999	5.33%	6.25%
$20,000 to $29,999	8.04%	13.02%
$30,000 to $39,999	13.32%	12.27%
$40,000 to $49,999	11.04%	10.62%
$50,000 to $74,999	22.12%	19.46%
$75,000 to $99,999	19.74%	10.23%
$100,000 to $124,999	4.44%	5.20%
$125,000+	8.75%	7.09%
Total	100%	100%
	Other Demographic Characteristics	
	My dataset CA	*U.S. Census*
Home owner	74.68%	66.20%
Renter	9.69%	33.80%
Missing	15.63%	0.00%
Married	44.76%	52.40%
Single	19.97%	47.60%
Missing	35.27%	0.00%

Figure AI



Source: U.S. Census and credit card data set.

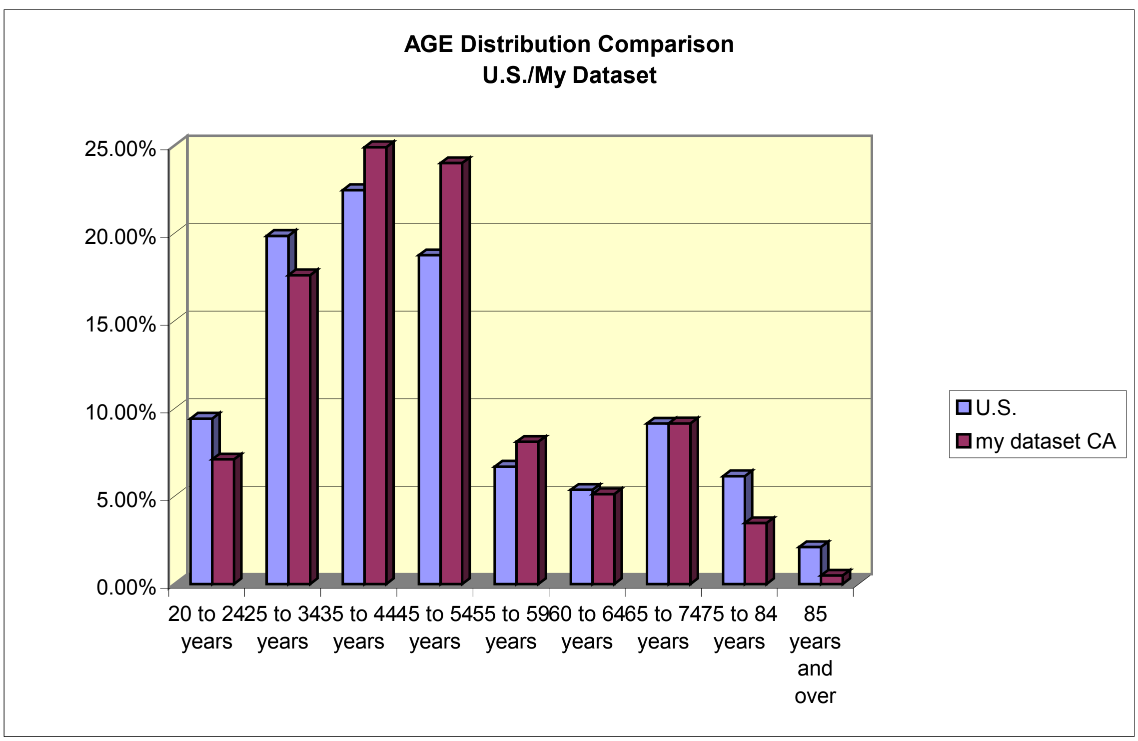


Table AII
Occupation:
Comparison with U.S. Census 2000

My Dataset

Occupation	*Percent*
Administrative/Managerial	7.59%
Clerical/White_Collar	3.25%
Craftsman/Blue_Collar	3.93%
Farmer	0.11%
Housewife	0.67%
Military	0.19%
Professional/Technical	13.76%
Sales/Service	1.91%
Self_Employed	2.51%
Student	1.65%
Other	64.44%
Total	100%

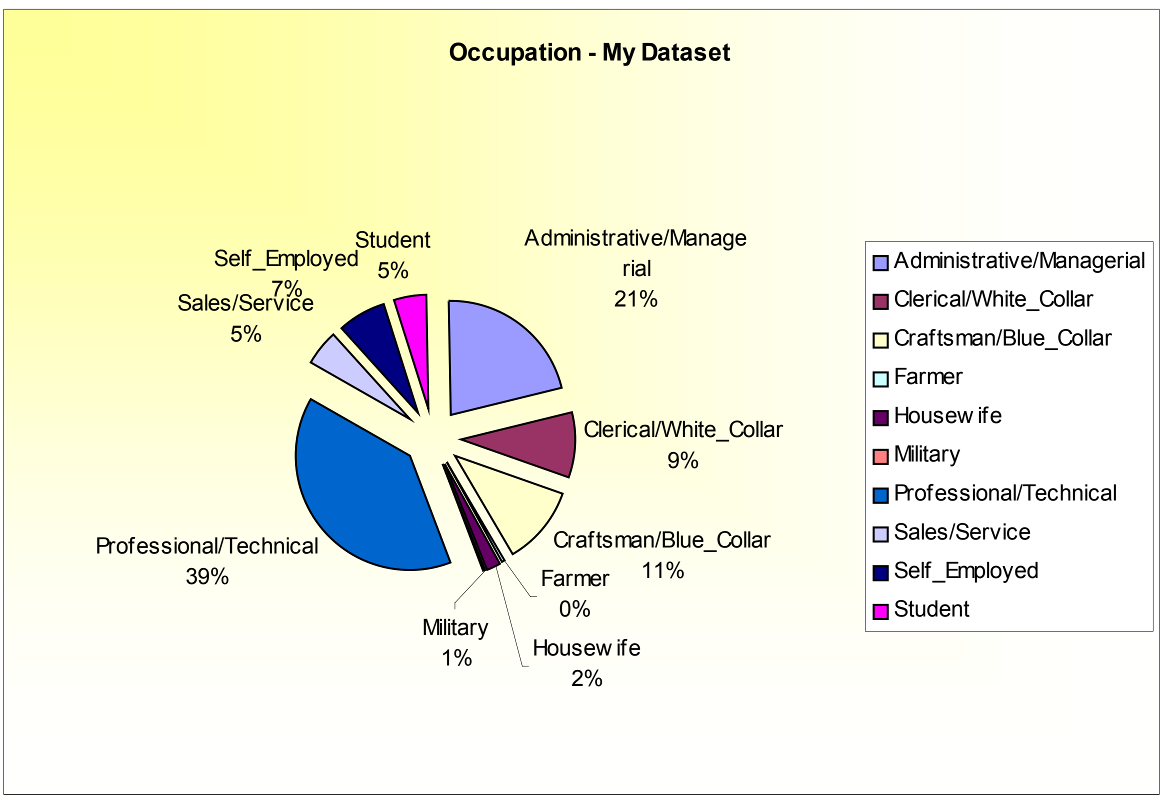


Table A III
Panel A
Financial Characteristics
Comparison with multi-issuer credit card dataset

	My dataset*		Gross and Souleles (2002)**	
	Mean	Median	Mean	Median
debt	$1,486.10	$0	$1,349.00	$70
debt|debt>0	$3,408.36	$2,821	$2,809.00	$2,120
credit limit	$8,823.51	$10,000	$6,207.00	$5,000
D credit limit	$154.15	$0	$76.80	$0
if D credit limit~=0	$1,181.65	$1,000	$1,985.00	$1,000
interest rate	16.13	14.99	16.60	17.20
D interest rate	-0.109	0	0.036	0
if D interest rate~=0	-1.012	-2.25	0.914	0.25

* The period analyzed is Aug.1998-Jul.2002

** Source: Gross and Souleles (2002), Table I. The period analyzed is Jan. 1995-Jan.1998

Panel B
Financial Characteristics
Comparison with the Survey of Consumer Finances

	My dataset		SCF	
	Mean	Median	Mean	Median
All	$3,408.36	$2,821	$4,100	$1,900
by Age				
Less than 35	$2,952	$2,312	$4,000	$2,000
35-44	$3,578	$2,967	$4,300	$2,000
45-54	$3,880	$3,202	$4,200	$2,300
55-64	$3,092	$2,463	$4,100	$1,900
65-74	$3,276	$2,810	$5,200	$1,000
older than 75	$3,720	$3,233	$1,900	$700
by Income Percentiles				
Less than 20	$3,039	$2,516	$2,100	$1,000
20-39.9	$3,285	$2,814	$2,800	$1,200
40-59.9	$3,820	$3,207	$3,700	$2,000
60-79.9	$3,551	$2,850	$4,700	$2,300
80-89.9	$3,215	$2,643	$7,200	$3,800
90-100	$3,836	$3,401	$6,600	$2,800
by Housing Status				
Home Owner	$3,559	$3,033	$4,500	$2,100
Renter	$3,405	$2,826	$3,400	$1,200