

Statistical Analysis of Market Penetration in a Mandatory Privatized Pension Market Using Generalized Logistic Curves

Victor M. Guerrero and Tapen Sinha
Instituto Tecnológico Autónomo de México

Abstract: In this paper we analyze market penetration of the Mexican Pension System. This market is unique in two respects: it is mandatory and it is private. Very few markets in the world have these two characteristics. In Mexico, the pension system became privatized in July 1997. By the end of 1999, more than 98% of workers (in the formal sector) had affiliated themselves with some private pension provider. We used two simple statistical methods of analysis of market share to draw some conclusions about how market share unfolds in a mandatory but privatized market. Our first descriptive analysis is based on a generalized logistic growth curve and the second on some simple linear regression fitting. Our results show that individual pension funds did not have similar growth patterns. Early market leaders (in terms of market share) did not necessarily stay leaders in the end. However, the first 12 months turned out to be critical for market share.

Key words: Growth curves, linearizing power transformation, logistic model, market penetration, saturation level.

1. Introduction

Mandatory privatized pension provides a unique opportunity to study market penetration. It is unique in two respects. (1) It is mandatory. Therefore, there is no choice about “buying”. (2) It is private. Therefore, a worker can affiliate himself or herself to any of the providers. Very few markets in the world have these two characteristics.

The objective of this paper is to provide a more general method (than fitting standard logistic growth curves) of empirically estimating market saturation. This method provides us with a way of classifying different companies in the market. One potential use of the model is to predict later growth in market share for every company based on early trends. Since the initial cost of staying in this market is very high (because of high initial fixed cost), a model that can be used for predicting the market share based on early behavior, can be very

useful. Thus, for example, companies can use this model to decide whether to “stick around” or cut their losses and exit from the market.

What is so special about mandatory participation in the market? Consider the market penetration for television sets (Young and Ord, 1989). It is customary to talk about market penetration as if every person would buy only one. That does not happen. On the other hand, mandatory pension schemes are unique: all workers have to have them and each person can have only one account. Law does not allow multiple accounts. Thus, market saturation is a better-defined concept in the context of a pension market than in the context of a market for television sets (or any other consumer good). There are several qualifications that need to be mentioned here. If a worker is in the informal sector of the labor market, he/she is not covered by social security at all. Informal sector refers to the part of the labor market that is not registered with the government. This sector pays no tax to the government. Most activities in the sector are paid for by cash only thus evading the entire banking network making it all but impossible to track. In Mexico, the informal sector accounts for more than 50% of the labor force. Therefore, a true market saturation will mean a coverage of all workers in both informal and in formal sectors. But since the informal market is not covered by the system of social security in Mexico, we leave it out of our discussion.

It might be argued that the informal market is getting formalized over time. There is no empirical evidence to that effect in Mexico over the past decade (see Sinha, 2000, p. 209). However, with panel data from a number of countries in Latin America, Packard (2002) finds evidence that private pension system is providing incentives for labor markets to be formalized. Strictly speaking, market saturation will exist only when the number of workers entering the pool of contributors (either from the informal labor market or as new entrants) is equal to the number of workers leaving the formal labor market (through retirement, death or moving to the informal sector).

Why does a private market make any difference? In most countries around the world (especially in other OECD countries), pension is by and large public (if it is mandatory) and pay as you go (this is the term commonly employed to indicate that current contribution is used to pay pension right holders of today). Private market allows us to study market share in an unfettered condition. In some countries with private pension (such as in Uruguay), government affiliated companies have been allowed to dominate the market (for a broad analysis of the Latin American privatization, see Sinha, 2000). This is not so in Mexico.

1.1 The Mexican privatized pension market

In December 1995, the Mexican Congress passed the New Social Security Law (Ley del Seguro Social) paving the way for the new system. A second set of laws

were passed in April 1996 to put in place the reforms in pension system (Ley de los Sistemas de Ahorro para el Retiro). This law allowed privatized-management of the country's pension system. It approved operation of investment management companies (Administradoras de Fondos para el Retiro or AFOREs). Each AFORE can now manage individual retirement funds (Sociedades de Inversion Especializadas en Fondos para el Retiro or SIEFORES). Each AFORE account holder contributes 6.5% of wages to the system (paid for by the employer). In addition, there is a "government contribution" of 5.5% of the minimum wages (for more details, see Sinha, 2002).

Workers can choose any AFORE for contribution. Once an AFORE is chosen, no change can be made for one year. It is possible to choose a different AFORE every year without any financial penalty. AFOREs are allowed to charge "management fees" either as a percentage of contribution, or as a percentage of value accumulated or any combination thereof. Most AFOREs charge as a percentage of contribution. AFOREs have to inform the affiliate (an affiliate is a worker who has an account with a particular AFORE) about their account at least once a year. The statement would include information about the accumulated value, contribution during the year and any charges the account has incurred. The AFOREs started to collect compulsory and voluntary contributions from February of 1997. Contribution to the new system became compulsory for all private sector workers in September 1997. Note that all formal sector workers are not necessarily private sector workers because approximately 23% of formal sector workers work for the federal, state or municipal governments. Among the workers in the formal sector, very few agricultural workers or self-employed workers are included.

CONSAR, the regulatory body of the AFOREs in Mexico, had issued 17 licenses by the end of 1997. Three of the AFOREs have merged with others. Consequently, as of February 28, 1999 there are 14 AFOREs left in the market. All of these mergers had to be approved by CONSAR as per regulation. After the rearrangement, the fourteen AFOREs that are left in the market are Banamex, Bancomer, Bancrecer, Banorte, Bital, Garante, Genesis, Inbursa, Principal, Profuturo GNP, Santander, Tepeyac, Siglo XXI, and Zurich. The merits and demerits of the system are analyzed in Espinosa and Sinha (2000).

2. Statistical Methodology

2.1 Growth curves

Let Z_t be a variable that indicates the trend of market penetration at time $t = 1, \dots, N$. Thus Z_t is measured as the number of elements of the population owning a particular item (in this case, membership to an AFORE). We assume

that the saturation level of the market is the constant $a < \infty$, so that $0 < Z_t < a$ and $Z_t \rightarrow a$ as $t \rightarrow \infty$. There are three commonly used models of market penetration (see Meade, 1984): (i) Modified exponential, (ii) Logistic and (iii) Gompertz. Here we shall employ an extension of the logistic curve, obtained as follows. First, note that the equation

$$Z_t = \frac{a}{1 + c \exp(-\beta t)} \quad (2.1)$$

with $c, \beta > 0$, can also be expressed as

$$\log[Z_t/(a - Z_t)] = \alpha + \beta t \quad (2.2)$$

with $\alpha = \log(1/c)$. Taking log is possible here by the assumption that Z_t is positive and bounded by above by a . Thus, the (natural) logarithm is employed basically as a linearizing transformation of the variable

$$Z_t^* = \frac{Z_t}{a - Z_t} = \frac{Z_t/a}{1 - Z_t/a} \quad (2.3)$$

which can be interpreted as the odds of the event of owning a particular item of interest and whose empirical probability of success is Z_t/a . Therefore, when $Z_t^* > 1$ we can say that the probability of the event is greater than that of its complement. This change of variable is used to extend the range of the original variable from $0 < Z_t < a$ to $0 < Z_t^* < \infty$.

Rather than imposing the use of logarithms, we shall let the data lead us to an appropriate linearizing transformation chosen within the following power family that includes the logarithm as a special case. That is, we suggest selecting the data transformation within the family indexed by the value of λ in

$$\begin{aligned} T_\lambda &= Z_t^{*\lambda} && \text{if } \lambda > 0 \\ &= \log(Z_t^*) && \text{if } \lambda = 0 \\ &= -Z_t^{*\lambda} && \text{if } \lambda < 0 \end{aligned} \quad (2.4)$$

This power transformation family is similar to the Box-Cox family of transformations, defined for $Z_t^* > 0$ by $Z_t^{*(\lambda)} = (Z_t^{*\lambda} - 1)/\lambda$ if $\lambda \neq 0$ and $\log(Z_t^*)$ if $\lambda = 0$. In fact, both (2.4) and the Box-Cox transformation keep the same trend direction as the variable Z_t^* . However (2.4) is discontinuous at $\lambda = 0$, but continuity of $T_\lambda(Z_t^*)$ in λ is not required by the method that will be employed here.

Guerrero (2000) proposed a method to select the index of the transformation (2.4) that most adequately validates the assumption that

$$T_\lambda(Z_t^*) = \alpha + \beta t + \epsilon_t \quad (2.5)$$

for $t = 1, 2, \dots, N$, where $\{\epsilon_t\}$ is a sequence of zero-mean random error terms with constant variance. The method was designed as a data description device applicable to (short) time series. It allows $\{\epsilon_t\}$ to be integrated of order 0 (stationary) or integrated of order 1 (nonstationary), so that 0 or 1 differences are needed to make the sequence $\{\epsilon_t\}$ stationary, and it employs the corresponding optimal linear estimator of β in each case. A useful reference for model (2.5) with different orders of integration (including fractional integration) for the random sequence $\{\epsilon_t\}$ is Deo and Hurvich (1998). When $\{\epsilon_t\}$ is integrated of order 1 we need the additional assumption that the sequence started at some finite point in the past with a fixed initial condition given by the value 0.

By substituting (2.3) in (2.4) and using (2.5) without the random error, we get the extended logistic curve given by

$$\begin{aligned} Z_t &= a/[1 + (\alpha + \beta t)^{-1/\lambda}] && \text{if } \lambda > 0 \\ &= a/[1 + \exp(-\alpha - \beta t)] && \text{if } \lambda = 0 \\ &= a/[1 + (-\alpha - \beta t)^{-1/\lambda}] && \text{if } \lambda < 0 \end{aligned} \quad (2.6)$$

This is a more flexible family of growth curves than (2.1), which is obtained as a special case of (2.6) when $\lambda = 0$. The index λ could be estimated by Nonlinear Least Squares if we were willing to assume a particular behavior for $\{\epsilon_t\}$ (e.g. that the errors are independent and identically distributed or that they follow an autoregressive process of order 1), but we will not pursue this line of investigation here. Instead we shall apply a method that leaves the structure of $\{\epsilon_t\}$ to be determined by the data themselves in so far as being integrated of order 0 or 1.

2.2 Selection of the linearizing transformation

Guerrero's (2000) method consists of calculating a sequence of λ values that trace the concavity or convexity in $\{Z_1^*, Z_2^*, \dots, Z_N^*\}$ by means of

$$\lambda_{t+2} = Z_{t+1}^*/(Z_{t+2}^* - Z_{t+1}^*) - Z_t^*/(Z_{t+1}^* - Z_t^*) \quad (2.7)$$

for $t = 1, 2, \dots, N - 2$, where $\lambda_{t+2} > 0$ implies (local) concavity and $\lambda_{t+2} < 0$ (local) convexity in the data around time $t+2$. To summarize the global behavior of the data the following measures of central tendency can be used and they become natural estimates of λ ,

$$\bar{\lambda} = [Z_{N-1}^*/(Z_N^* - Z_{N-1}^*) - Z_1^*/(Z_2^* - Z_1^*)]/(N - 2) \quad (2.8)$$

and

$$\lambda_{Med} = \text{Median}\{\lambda_{t+2} : t = 1, 2, \dots, N - 2\} \quad (2.9)$$

Selection of (2.8) or (2.9) can be made on the basis of the absence or presence of extreme values, respectively.

For each estimated value, two estimates of the slope parameter are calculated, the Ordinary Least Squares (OLS) estimator

$$\hat{\beta}_{OLS} = \sum_{t=1}^N \frac{[t - (N + 1)/2][T_{\lambda}(Z_t^*) - \bar{T}_{\lambda}(Z^*)]}{\sum_{t=1}^N [t - (N + 1)/2]^2} \quad (2.10)$$

and the average difference estimator

$$\hat{\beta}_{DIF} = [T_{\lambda}(Z_N^*) - T_{\lambda}(Z_1^*)]/(N - 1) \quad (2.11)$$

with $\bar{T}_{\lambda}(Z^*)$ the average of $\{T_{\lambda}(Z_1^*), T_{\lambda}(Z_2^*), \dots, T_{\lambda}(Z_N^*)\}$ and $\hat{\alpha} = T_{\lambda}(Z^*) - \hat{\beta}_{OLS}(N + 1)/2$ the accompanying estimate of $\hat{\beta}_{OLS}$. These linear estimates are the most efficient in a statistical sense whenever the order of integration is 0 or 1, respectively (see Deo and Hurvich, 1998). Discrimination between the order 0 or 1 is performed by comparing the coefficient of determination, R^2 , they produce. To this end, we should recall that only the following expression works well when dealing with transformed data (see Kvålseth, 1985)

$$R^2 = 1 - \frac{\sum_{t=1}^N \left\{ Z_t^* - T_{\lambda}^{-1} \left[\hat{T}_{\lambda}(Z_t^*) \right] \right\}^2}{\sum_{t=1}^N (Z_t^* - \bar{Z}^*)^2} \quad (2.12)$$

where $\hat{T}_{\lambda}(Z_t^*)$ denotes an estimated value in the transformed scale and $T^{-1}[T_{\lambda}(Z_t^*)]$ is the corresponding estimated value brought back to the original scale of Z_t^* . Thus, with respect to the choices (2.10) and (2.11) we have that

$$\begin{aligned} \hat{T}^{-1}[T_{\lambda}(Z_t^*)] &= (\hat{\alpha} + \hat{\beta}_{OLS}t)^{1/\lambda} && \text{if } \lambda \neq 0 \\ &= \exp(\hat{\alpha} + \hat{\beta}_{OLS}t) && \text{if } \lambda = 0 \end{aligned} \quad (2.13)$$

is the adjusted value when OLS is used. Whereas in the other case we get

$$\begin{aligned} \hat{T}^{-1}[T_{\lambda}(Z_t^*)] &= (T_{\lambda}(Z_{t-1}^*) + \hat{\beta}_{DIF})^{1/\lambda} && \text{if } \lambda \neq 0 \\ &= \exp[T_{\lambda}(Z_{t-1}^*) + \hat{\beta}_{DIF}] && \text{if } \lambda = 0 \end{aligned} \quad (2.14)$$

for $t = 2, \dots, N$ and $\hat{T}^{-1}[T_{\lambda}(Z_1^*)]$.

If the true value of the saturation level were known (as it happens when Z_t^* represents a proportion, in which case $a = 1$) Guerrero's method can be applied as originally proposed. Nevertheless, in many applications the saturation level should be considered as an unknown parameter that has to be estimated from the

available data. In that case, we suggest to search for the value of the saturation level that maximizes R^2 , over a grid of “ a ” values.

3. Data Analysis

The data set consists of the number of accounts recorded in each of the 14 existing AFOREs. At the beginning of 1999, seven of the 14 managed to capture about twelve million accounts (out of a market size of 14 million accounts). The rest together has only two million accounts. For the purpose of this analysis, we shall group the 7 smallest AFOREs in one called *Others*. Thus we shall analyze here the data on the following AFOREs: (1) Banamex, (2) Bancomer, (3) Banorte, (4) Bital, (5) Garante, (6) Profuturo, (7) Santander, (8) Others and (9) the Total number of accounts. The corresponding data are shown in the Appendix (expressed in thousands of accounts from February 1997 to February 2000). We decided to work first with the variable $Z = \text{Total Accounts}$ by applying the methodology described in the previous section. By so doing we found that the optimum saturation level (providing the largest R^2 over a grid of “ a ” values) was $\tilde{a} = 21900$. Correspondingly, the concavity or convexity in Total Accounts was numerically evaluated as indicated in Table 1 (see also Figure 1 to appreciate it visually), where the values shown were calculated using equation (2.7).

Table 1: Summary of results for total accounts ($\tilde{a} = 21900$)

Month	Feb 97	Mar	Apr	May	Jun	Jul	Aug	Sep 97
λ	-	-	0.60	0.32	0.69	0.22	1.18	-0.27
Month	Oct 97	Nov	Dec	Jan 98	Feb	Mar	Apr	May 98
λ	3.51	1.72	3.28	9.49	0.39	2.05	4.14	-12.18
Month	Jun 98	Jul	Aug	Sep	Oct	Nov	Dec	Jan 99
λ	9.02	2.94	-3.42	-1.79	6.64	-3.01	8.25	11.00
Month	Feb 99	Mar	Apr	May	Jun	Jul	Aug	Sep 99
λ	-4.28	-7.38	6.62	-2.74	-7.04	6.94	1.02	2.98
Month	Oct 99	Nov	Dec	Jan 00	Feb			
λ	-22.18	0.47	13.68	4.45	-10.74			

Global values: $\bar{\lambda} = 0.76, \lambda_{Med} = 1.02$.

Table 1 allows us to say that the behavior of the odds is convex from April to September, 1997 (except perhaps in August) indicating a greater than linear growth during that period, then it becomes concave from October, 1997 to

April, 1998 (excepting February). From May 1998 onwards, the pattern in the odds is unclear since it changes several times from convex to concave and *vice versa*, therefore a linear growth may be deemed reasonable. Since there are many extreme values in Table 1, the best summary is provided by the median of the values, that is $\lambda_{Med} = 1.02$, whose proximity to unity indicates that the global behavior of the odds may be well described as linear.

Thus, a reasonable linearizing power transformation for Total accounts has index $\lambda = 1$. With this value already fixed we searched again for the best estimate of the saturation level, obtaining as a result $\hat{a} = 20400$. This number is very close to 20139.34, which is the estimate provided by IMSS for the “potential formal labor market size” (quoted in CONSAR, 1999). Using these values of λ and \hat{a} we obtained the fit shown in Figure 1 for the original variable Z_t and in Figure 2 for the odds Z_t^* , calculated using equation (2.4).

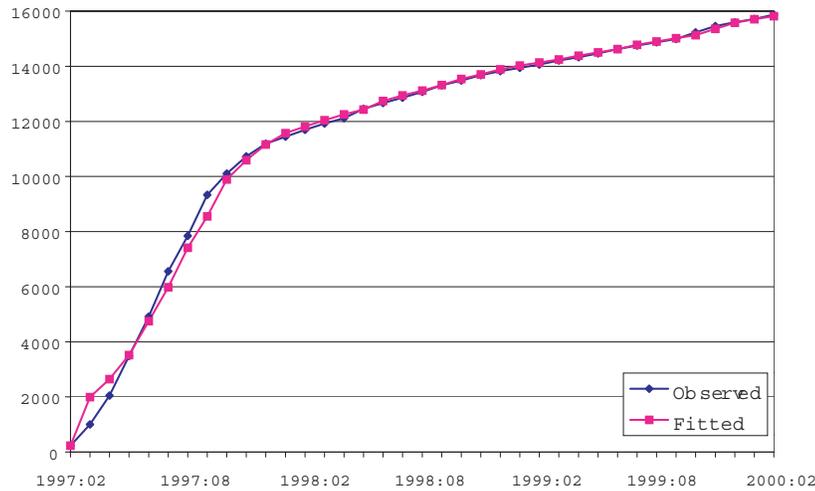


Figure 1: Total accounts in AFOREs

It should be clear that the estimated saturation level was obtained through an exploratory data analysis that was aimed only at obtaining a satisfactory fit to the observed data. This rationale is in line with the statistical requirement of a saturation level mentioned by Meade (1984), although put forward by other authors: “The primary objective should be to fit recent data satisfactorily, rather than forcing the curve to pass through some pre-ordained point in 50 years time”. Moreover, Meade proposed that: “If the estimation of the saturation level is consistent over time, this is satisfactory. However, if the estimated saturation level is infeasible or fluctuates widely, it would seem reasonable to estimate the saturation level as a function of appropriate explanatory variables, given sufficient degrees of freedom”.

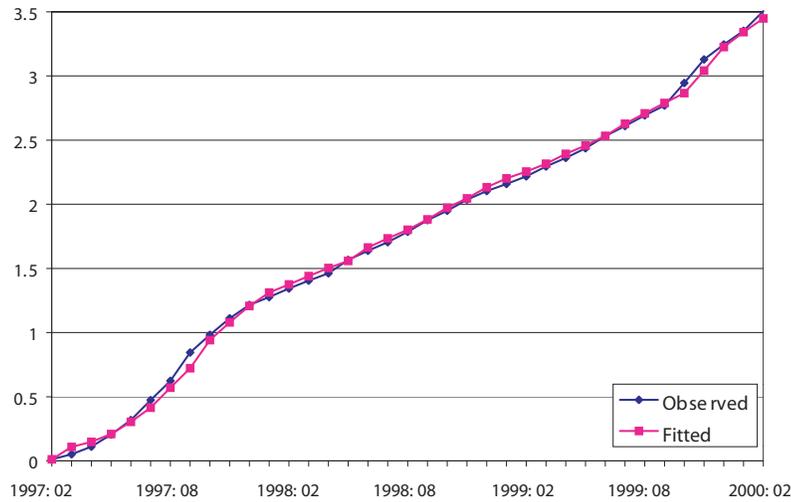


Figure 2: Odds of total accounts

Due to the short length of the time series under study, we considered only two additional estimation exercises of the saturation level, first with data up to January, 2000 and then with data up to December, 1999. In both exercises we obtained the estimated value $\hat{\alpha} = 20200$, which is very close to the aforementioned estimate obtained with all the available data (up to February, 2000). Therefore we did not deem necessary to employ a model with explanatory variables for the saturation level. Besides, we are convinced that the estimated saturation level already takes into account population (formal labor market) growth implicitly, since the data correspond to consecutive time periods with different population sizes. An implicit assumption when estimating the saturation level is that the population growth will remain basically unchanged in the future.

In Table 2 we present some estimation results of the extended logistic growth curve introduced in the previous section. Let us recall that $\hat{\beta}_{OLS}$ and $\hat{\beta}_{DIF}$ are given by equations (2.10) and (2.11) and that the estimated intercept $\hat{\alpha}$ is only required in association with $\hat{\beta}_{OLS}$.

Just for the sake of completeness, we also included $\lambda = 0$ in Table 2, because that value leads to the use of an ordinary Logistic growth curve, which is evidently not recommended in the present situation, due to its R^2 value and corresponding saturation level. Thus, in this case there is some gain in goodness of fit by using the extended logistic curve.

In order to make the results of the growth curve fitting for the different AFOREs comparable with each other, it was decided to use the same power

Table 2: Extended logistic estimation results for total accounts ($\tilde{a} = 21900$)

Saturation Level	Index of transformation	R^2 values	
		$\hat{\beta}_{OLS}$	$\hat{\beta}_{DIF}$
21900	$\lambda_{Med} = 1.02$	0.9782	0.9981
20400	$\lambda = 1$	0.9867	0.9983
16400	$\lambda = 0$	0.8566	0.9868

transformation found earlier ($\lambda = 1$) for all the AFOREs. The odds were fitted by growth curves with their respective optimal saturation level. In all cases, the largest coefficient of determination was found with $\hat{\beta}_{DIF}$. Table 3 shows the saturation levels, estimated slopes, R^2 values, dates at which the median level (50% of the saturation level) was attained and percentage of the saturation level reached in February, 2000, which is the end-of-the-sample month.

Table 3: Estimation results for the growth curves of the AFOREs

AFORE	Saturation level (\hat{a})	$\hat{\beta}_{DIF}$ for odds	R^2	Median level at	End-of-sample level(%)
Banamex	2600	0.07873	0.9940	Sep 97	74.11
Bancomer	3100	0.12488	0.9948	Aug 97	81.87
Banorte	2200	0.04726	0.9976	Aug 98	63.05
Bital	2500	0.05209	0.9980	Oct 98	65.26
Garante	2100	0.12684	0.9978	Oct 97	82.04
Profuturo	2200	0.33875	0.9985	Aug 97	92.43
Santander	2500	0.21966	0.9977	Sep 97	88.79
Others	3400	0.06794	0.9975	Jun 98	70.98
Total	20400	0.09709	0.9983	Nov 97	77.81

The estimated coefficients $\hat{\beta}_{DIF}$ in Table 3 are directly linked to the end-of-sample level reached, since the larger the estimated slope, the larger the percentage level reached in February, 2000. Thus, in accordance with these values we have the following categories for the AFOREs.

1. Accelerated growth (characterized by a slope larger than 0.2 and end-of-

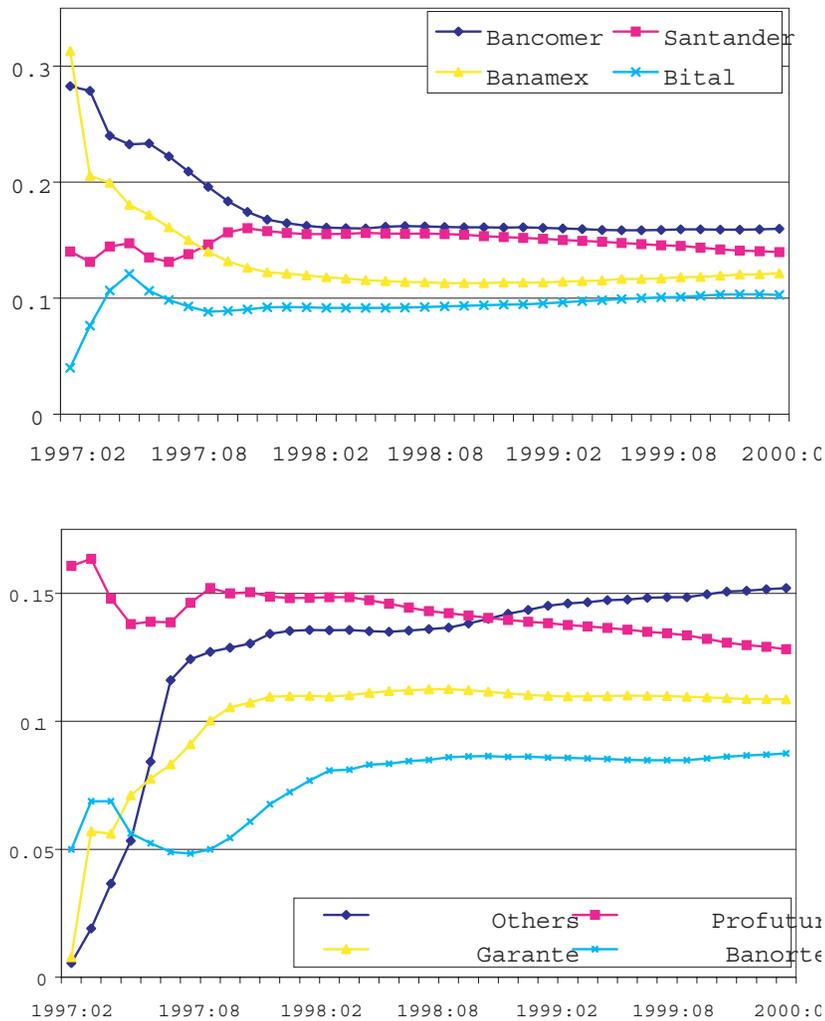


Figure 3: AFOREs as proportions of total accounts

sample level around 90%): Profuturo and Santander.

2. Fast growth (slope between 0.1 and 0.2, and end-of-sample level around 82%): Garante and Bancomer.

3. Moderate growth (slope between 0.06 and 0.1, with end-of-sample level around 72%): Banamex and Others.

4. Slow growth (slope smaller than 0.06 and end-of-sample level around 65%): Bital and Banorte.

According to this classification, the Total number of accounts can then be classified as having either Fast or Moderate growth.

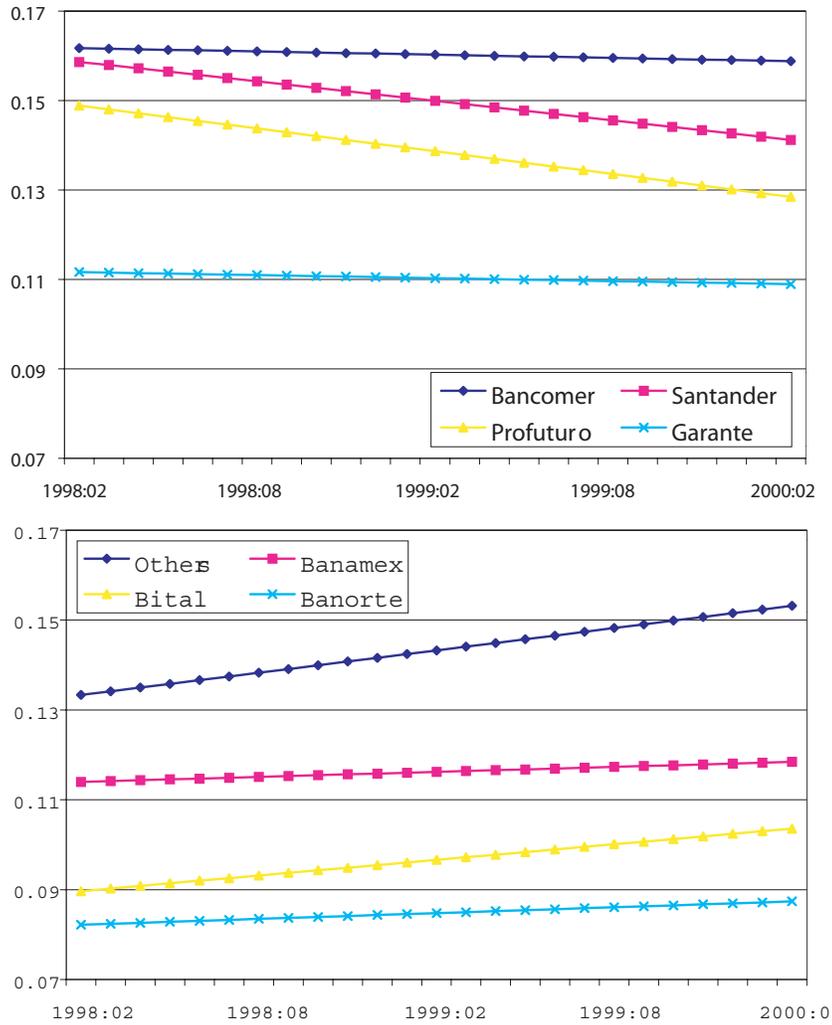


Figure 4: Fitted proportions of AFOREs

Next, if we define that a market reaches maturity when its level equals or is larger than 50% of its saturation level, then we may say that this happened in November, 1997 for Total Accounts. For each individual AFORE, its maturity started as early as August, 1997 (Bancomer and Profuturo) and no later than October, 1998 (Bital). On the other hand, it is interesting to notice that the sum of the individual saturation levels is 20600. This figure is close to the estimated saturation level of Total Accounts. Thus we are led to believe that if a joint estimation of the whole set of growth curves had been carried out, taking into account the restriction that the sum of the saturation levels be equal to 20400,

we would have obtained similar results to those reported in Table 3. However, carrying out that kind of joint estimation is beyond the scope of this paper.

Another type of descriptive analysis consists of studying the proportions of values of each AFORE with respect to Total Accounts. In Figure 3 we present such proportions. There we can observe that each proportion has its own trend, but all the trends are very close to linear from the beginning of 1998. Thus, it is natural to fit a linear trend to each proportion of AFORE to represent its behavior once it became a mature product. That is, the fit for all the proportions started at the average time of maturity (the average of August 1997, October 1998 and November 1997) which is February 1998.

The fitted trend lines for each proportion of AFORE are shown in Figure 4. There we can see that four AFOREs have proportions trending downward (Bancomer, Santander, Profuturo and Garante). While Banamex, Bital, Banorte and Others trend upwards. Of course this empirical regularity is valid only during the sample period and it should not be expected to hold true in the long-run.

One particular application of the previous analysis is that of predicting the future number of accounts in each AFORE. To do that we can use the estimated slope for the odds of the AFORE, $Z_t^* = Z_t/(\hat{a} - Z_t)$, and add it to the most recent odds in order to get a forecast of Z_{t+1}^* . Then we must bring this forecast to the original scale of the variable by means of the formula $Z_{t+1} = \hat{a}Z_{t+1}^*/(1 + Z_{t+1}^*)$.

Just for illustrative purposes, since in our case there was no real need to get forecasts, we calculated three monthly forecasts for Banamex. Since the most recent observation of the AFORE accounts in Banamex was that of February, 2000 ($Z_{Feb} = 1927$) we first calculated the odds

$$Z_{Feb}^* = Z_{Feb}/(2600 - Z_{Feb}) = 2.8633$$

Then we obtained forecasts of the odds

$$Z_{Mar}^* = Z_{Feb}^* + \hat{\beta}_{DIF} = 2.9420, Z_{Apr}^* = Z_{Mar}^* + \hat{\beta}_{DIF} = 3.0208$$

$$Z_{May}^* = Z_{Apr}^* + \hat{\beta}_{DIF} = 3.0995$$

from which the following forecasts of the number of accounts were obtained

$$\hat{Z}_{Mar} = 2600Z_{Mar}^*/(1 + Z_{Mar}^*) = 1940.4, \dots,$$

$$\hat{Z}_{Apr} = 2600Z_{Apr}^*/(1 + Z_{Apr}^*) = 1953.4$$

$$\hat{Z}_{May} = 2600Z_{May}^*/(1 + Z_{May}^*) = 1965.8$$

4. Conclusions

We present two statistical descriptions of the data on the AFOREs. (1) The first one was based on an extended logistic model for market penetration. We were able to classify the AFOREs objectively, according to its observed behavior, without any preconception of the factors that we should take into consideration. For example, some early observers claimed that the fund Siglo XXI will win a large share of the market simply because it has the backing of the IMSS. It did not turn out that way. The very data led us to a classification scheme that emphasized the speed of growth of market penetration as well as product maturity.

The second descriptive analysis focused on the relative behavior of each AFORE with respect to the Total number of accounts. A simple fitting of regression lines, from the point of maturity onwards, allowed us to see the trend of each AFORE's market penetration. We believe that these trend lines could be used by the business managers of the AFOREs as a reference, in order to make plans about future investments, management fees and marketing in other markets. It is critical to emphasize that once the "first phase" of growth of market share goes through, there is not much that can be done to change the market share as a saturated market can only offer additional workers for an AFORE only at the cost of cannibalizing other AFOREs.

The analysis has important implications for business strategy for pension funds. In many countries around the world, the social security systems are being privatized. A company planning to enter the market needs to know early in the game whether they should stay (when they have initially captured a small market share) or they should cut their losses and leave the market. Our analysis shows that the first twelve months is the most important time frame (the first phase). If a company has not made significant inroads into the market by then, they are likely never to get a significant market share.

There are several caveats to our analyses here. (i) Market shares can change dramatically as consolidation takes place. In fact, when we analyzed the data, the Spanish bank BBV had its own AFORE called Probursa. However, in June 2000, BBV took over the second largest bank in Mexico, Bancomer. Bancomer also had the largest AFORE. Thus, BBV had to sell off its own holding of the AFORE (to Profuturo). In early 2002, another AFORE Genesis was sold to Principal. Thus, there was a realignment of market sizes. (ii) The point forecasts derived from the second descriptive analysis are only a first approximation to obtaining optimal forecasts. This is so because there is no measure of uncertainty associated to the forecasts and no optimality criterion was used to obtain them. However, if the need for optimal forecasts arises, then the point forecasts can be used as a starting point to improve upon them.

Acknowledgements

The authors gratefully acknowledge the financial support provided by Asociacion Mexicana de Cultura, A.C. to carry out this research project. They also thank the editor of this journal as well as an anonymous referee for several comments and suggestions which helped to improve our exposition.

References

- CONSAR (1999). Boletín Informativo SAR, Num. 6, Año 4, México City.
- Deo, R. S. and Hurvich, C. M. (1998). Linear trend with fractionally integrated errors. *Journal of Time Series Analysis* **19**, 379-398.
- Espinosa, M. and Sinha, T. (2000). A primer of the Mexican pension system. *Economic Review of the Federal Reserve Bank of Atlanta*, 1-23.
- Guerrero, V. M. (2000). Selecting a linearizing power transformation for time series. *Journal of Applied Statistics* **27**, 185-195.
- Kvålseth, T. O. (1985) Cautionary note about R^2 . *The American Statistician* **39**, 279-285.
- Meade, N. (1984). The use of growth curves in forecasting market development – a review and appraisal. *Journal of Forecasting* **3**, 429- 451.
- Packard, T. (2002). Are there positive incentives from privatizing social security? A panel analysis of pension reform in Latin America. *Journal of Pension Economics and Finance* **1**, 89-109.
- Sinha, T. (2000). *Privatization of Social Security in Latin America*. Kluwer Academic Publishers.
- Sinha, T. (2002). *Retrospective and Prospective Analysis of the Privatized Mandatory Pension System in Mexico*. Society of Actuaries Monograph, Society of Actuaries, Chicago, USA.
- Young, P. and Ord, J. K. (1989). Model selection and estimation for technological growth curves. *International Journal of Forecasting* **5**, 501-513.

Appendix

Number of accounts in AFOREs (000's)

Company	(a)	(b)	(c)	(d)	(e)	(f)	(g)	Others	Total
Feb '97	71	64	11	9	2	36	32	1	226
Mar	206	280	69	77	57	164	132	19	1003
Apr	408	491	141	218	115	302	296	75	2044
May	629	810	196	421	248	481	513	186	3484
Jun	845	1148	258	523	381	683	665	414	4918
Jul	1057	1457	321	645	546	909	861	761	6557
Aug	1177	1641	379	728	713	1147	1081	974	7841
Sep	1307	1830	466	824	936	1419	1367	1186	9335
Oct	1333	1854	551	900	1066	1516	1586	1301	10107
Nov	1355	1871	652	970	1151	1615	1720	1399	10732
Dec	1368	1875	757	1030	1226	1664	1767	1501	11188
Jan '98	1386	1884	828	1057	1256	1694	1788	1548	11442
Feb	1398	1898	898	1077	1285	1734	1815	1587	11692
Mar	1407	1918	962	1092	1306	1769	1849	1616	11919
Apr	1416	1943	983	1109	1335	1799	1885	1643	12112
May	1438	1994	1034	1140	1383	1835	1945	1683	12451
Jun	1453	2048	1056	1162	1416	1848	1973	1710	12666
Jul	1467	2084	1086	1183	1442	1857	1998	1741	12857
Aug	1486	2115	1110	1209	1470	1871	2035	1778	13074
Sep	1504	2146	1144	1238	1498	1892	2067	1817	13306
Oct	1523	2173	1163	1258	1511	1905	2085	1863	13483
Nov	1546	2203	1182	1286	1526	1921	2099	1916	13680
Dec	1569	2226	1191	1305	1533	1930	2110	1964	13828
Jan '99	1581	2247	1202	1318	1538	1936	2117	2000	13939
Feb 1	598	2259	1208	1341	1546	1945	2126	2040	14062
Mar	1625	2274	1217	1371	1559	1956	2133	2075	14210
Apr	1646	2286	1225	1398	1573	1964	2142	2099	14332
May	1668	2298	1233	1423	1588	1974	2149	2130	14463
Jun	1703	2318	1242	1450	1609	1985	2157	2158	14622
Jul	1722	2340	1251	1476	1622	1992	2162	2186	14751
Aug	1743	2364	1261	1500	1634	1998	2168	2208	14875
Sep	1771	2387	1271	1516	1643	2002	2174	2225	14989
Oct	1804	2428	1301	1553	1665	2014	2186	2279	15230
Nov	1845	2461	1333	1592	1685	2021	2194	2328	15460
Dec	1876	2481	1351	1611	1695	2024	2200	2355	15595
Jan '00	1895	2505	1366	1623	1707	2028	2206	2381	15712
Feb	1927	2538	1387	1632	1723	2033	2220	2413	15873

Company names: (a)= Banamex, (b)=Bancomer, (c)= Banorte (d)=Bital, (e)= Garante, (f)=Profuturo, (g)= Santander.

Received October 8, 2002; accepted April 1, 2003.

Victor M. Guerrero
Statistics Department
Instituto Tecnológico Autónomo de México (ITAM)
México 01000, D.F., México
guerrero@itam.mx

Tapen Sinha
Actuarial Science Department
Instituto Tecnológico Autónomo de México (ITAM)
México 01000, D. F., México
tapen@itam.mx