**Econometrics: Case Study 1**

**Multiple Linear Regression for Bank Revenue**

**Background**

A bank wants to understand how customer banking habits contribute to revenues and profitability. The bank has customer age and bank account information, e.g., whether the customer has a savings account, whether the customer has received bank loans, and other indicators of account activity.

**The Task**

We want to build a model that allows the bank to predict profitability for a given customer. A surrogate for customer profitability available in our data set is the **Total Revenue** a customer generates through their accounts and transactions. The resulting model will be used to forecast bank revenues and guide the bank in future marketing campaigns.

1. Write the Analytical Framework
2. Perform Descriptive Analysis
3. Process the data for missing and outliers
4. Prepare the data for modeling
5. Develop a statistical model
6. Validate the model on In-time sample
7. Document the outputs and your findings
8. Predict profitability for all customers

**The Data** BankRevenue.sas7bdat  The data set contains information on 7,420 bank customers:

**Rev\_Total** Total revenue generated by the customer over a 6-month period.

**Bal\_Tota** Total of all account balances, across all accounts held by the customer.

**Offer** An indicator of whether the customer has received a special promotional offer in the previous one-month period. Offer=1 if the offer was received, Offer=0 if it was not.

**AGE** The customer’s age.

**CHQ** Indicator of debit card account activity. CHQ=0 is low (or zero) account activity,

CHQ=1 is greater account activity.

**CARD** Indicator of credit card account activity. CARD=0 is low or zero account activity, CARD=1 is greater account activity.

**SAV1** Indicator of primary savings account activity. SAV1=0 is low or zero account activity, SAV1=1 is greater activity.

**LOAN** Indicator of personal loan account activity. LOAN=0 is low or zero account activity, LOAN=1 is greater activity.

**MORT** Indicator of mortgage account tier. MORT=0 is lower tier and less important to the bank’s portfolio. MORT=1 is higher tier and indicates the account is more important to the bank’s portfolio.

**INSUR** Indicator of insurance account activity. INSUR=0 is low or zero account activity, INSUR=1 is greater activity.

**PENS** Indicator or retirement savings (pension) account tier. PENS=0 is lower balance and less important to bank’s portfolio. PENS=1 is higher tier and of more importance to the bank’s portfolio.

**Check** Indicator of checking account activity. Check=0 is low or zero account activity, Check=1 is greater activity.

**CD** Indicator of certificate of deposit account tier. CD=0 is lower tier and of less importance to the bank’s portfolio. CD=1 is higher tier and of more importance to the bank’s portfolio.

**MM** Indicator of money market account activity. MM=0 is low or zero account activity, MM=1 is greater activity.

**Savings** Indicator of savings accounts (other than primary) activity. Savings=0 is low or zero account activity, Savings=1 is greater activity.

**AccountAge** Number of years as a customer of the bank.