

耐久消費財需要とCensored Modelの改良

経済学科 西山 茂

1 はじめに

経済学の分野において耐久消費財需要の分析は長い歴史を持つ。乗用車に関するモデルチェンジ効果の分析はその一例であるし、エアコン需要と夏期の気温との関係も卒業論文には好適な課題だろう。

耐久消費財は何年にもわたって使われる物であるから、時系列データを用いる場合は、所得、価格といった主要な経済変数に加えて、ストック調整あるいは買い換えサイクルの存在などを考慮することが必要になる。また同一時点に対するクロスセクショナルな情報を利用する場合は、年間収入、世帯人員、年齢、職業（自営業かサラリーマンか無職かという別）、住宅床面積、地域、金融資産、消費者ローン金利、同種財の保有高などの説明変数が考慮される。

バブル経済が華やかになりし頃、消費者の二極分化現象が云々されたことがある。地価と株価上昇の恩恵を享受できた世帯と、そうした資産値上がり益（＝キャピタルゲイン）とは無縁の世帯の間で相当の購買力格差が発生し、その格差が持家購入、耐久消費財あるいは高額な貴金属商品の購買行動に反映されているという指摘である。このような指摘は集計された家計消費にとどまらず、所得格差、資産格差といった分布的側面に研究者の注意をふり向けるものとなった。

幸い筆者は、この数年、総務庁統計局「全国消費実態調査」特別集計結果を用いて個別世帯の世帯構造、所得、消費データを分析する機会を得てきた。本稿では、耐久消費財需要の測定に家計のマイクロデータを用いた場合に、どのような問題点が発生するか、その問題点の把握、解決に際して、本学情報処理センターのソフトウェア群がいかにか有効なものであったかをかいつまんで紹介したい。

2 耐久消費財需要とTobit Model

2.1 ゼロ支出世帯の問題

耐久消費財と一口に言っても多くの商品が存在する。ここでは冷蔵庫、電子レンジなど8種の耐久財を合計した「家事用耐久財」を取り上げよう。

マイクロデータを用いて需要分析を行う際に先ず問題となるのが「ゼロ支出世帯」の存在である。ゼロ支出世帯とは分析対象となっている財貨を購入しなかった世帯のことである。すべての消費統計調査は、ある調査期間を定めて、その期間内に購入した各財貨の金額を記入するようになっている。たとえば、二人以上の普通世帯の場合、「全国消費実態調査」では、9月から11月までの3箇月内に購入した実績を被調査世帯に自記してもらうことになっている。ところが、エアコンなどの耐久消費財は数年にわたって使い続けるものである。たまたま設けられた3箇月内に実際にその世帯がエアコンを買うかどうかは偶然による。実際、1984年9月から11月までに「家事用耐久消費財」に該当するいずれかの財を購入した世帯は、関東東海在住13546世帯の内の僅か12.4%でしかない。残りの87.6%の世帯は総支

出金額の高低にかかわらず家事用耐久消費財購入額はゼロであった。この様子を、本学情報処理センターのcanalに搭載されている「S言語」によって描画したのが図1である。縦軸は総支出額に占めるシェアを示し横軸は総支出額の対数を示している。一見すれば分かるように、実際に購入した世帯(=正值世帯)については負の相関が認められるが、大半の世帯は横軸上に布置されている。ちなみに、総支出金額でクラス分けしたうえ、各クラスに対してゼロ支出世帯の割合を図に描くと図2のようになる。総支出金額が高額な世帯ではやはり耐久財を購入した世帯の割合が高い。

散布図がこうしたパターンを示す時に何が困るかと言えば、(1)全データを用いても、(2)正值データのみを用いても、最小二乗法では線形需要関数の不偏推定値が得られないということである。

2.2 Tobit ModelとP-Tobit Model

本稿の読者には、当該領域の予備知識をもっていない方達も多いと思われる。そこで必要最小限の範囲にしぼってTOBITモデルを要約しておこう。

Tobitモデルは以下のように定式化される。

$$(1) \quad Y_i^* = X_i' \beta + u_i, \quad u_i \sim N(0, \sigma^2)$$

$$(2) \quad Y_i = \begin{cases} y_i^*, & \text{if } y_i^* > 0 \\ 0, & \text{if } y_i^* \leq 0 \end{cases}$$

上の Y_i^* は潜在変数 (Latent Variable) と呼ばれるものであり、たとえば「家事用耐久財」に対する潜在需要を示す。それが負値となる場合、統計調査では全てゼロとして回答され、個別の値は喪失されてしまう。つまり一定条件の下で、情報がcensorされてしまうわけであって、この意味でTobitモデルはCensored Modelの一種である。

Tobitモデルを最尤推定する際の対数尤度関数は以下のようなになる。

$$(3) \quad \ln L = \sum_0 \ln \Phi\left(\frac{-X_i' \beta}{\sigma}\right) + \sum_+ \ln \left\{ \sigma^{-1} \phi\left(\frac{Y_i - X_i' \beta}{\sigma}\right) \right\}$$

X_i は第 i 番目の世帯に関する総支出、各年齢層の世帯人員などの説明変数行列である。また Φ は標準正規分布関数、 ϕ は標準正規密度関数を意味する。最尤推定を未知母数 β に関して行う時には上の $\ln L$ を最大化すればよい。

しかしながら、耐久消費財需要の分析にはもう一つの問題が存在する。それは消費統計調査結果の示す過小性 (Underreporting) である。この問題は統計実務家の間では共通の印象として周知のこととされてきたが、この点を詳細に分析した研究例はこれまでのところ非常に少ない。Deaton-Irish (1984) は現代の需要理論の下で「消費データの過小性」までを一挙に解明しようとした点において大きな価値を有する文献といえる。彼らが採用したモデルはMisreporting Modelと呼ばれ、実際に財貨を購入した世帯が支出金額を正しく記入するかどうかは、二項確率変数 (W) が値1をとる時に限られると想定するものである。購入したにもかかわらず $W=0$ となった世帯は支出金額を記入しない、つまりmisreportingが生じるというのが彼らの発想の核心である。したがってDeaton-Irishの定義した対数尤度関数は以下のようなになる。但し、 p は確率変数 W が値1をとる確率であり、全世界帯について共通の母数としている。Deaton-Irishは $p=1$ を帰無仮説として統計的仮説検定にうったえる手法を採用したのである。

$$(4) \quad \ln L = \sum_0 \ln \left\{ 1 - p \Phi \left(\frac{-X'_i \beta}{\sigma} \right) \right\} - \frac{n_+}{2} \ln \sigma^2 + n_+ \ln p + \sum_+ \ln \phi \left(\frac{Y_i - X'_i \beta}{\sigma} \right)$$

実は、Deaton-Irishの試みは失敗に帰している。というのはpの最尤推定値が1を超えるからである。pは変数Wが1となる確率であるから1を超えることはありえない。換言するとp=1が最尤推定値となるわけで、Misreportingという行動をモデルによって説明することには失敗したわけである。

2.3 P - Tobit Modelの改善

今回、採用したPi-Tobitモデルは次の尤度関数を用いる。

$$(5) \quad \ln L = \sum_0 \ln \left\{ 1 - \Phi(X'_{2i} \gamma) \Phi \left(\frac{-X'_{1i} \beta}{\sigma} \right) \right\} - \frac{n_+}{2} \ln \sigma^2 + \sum_+ \ln \Phi(X'_{2i} \gamma) + \sum_+ \ln \phi \left(\frac{Y_i - X'_i \beta}{\sigma} \right)$$

つまりDeaton-Irishは全世帯に共通のP値を想定したが、我々は各世帯の属性により記入確率が異なると考えた。推定の詳細は別に譲るとして、各世帯の記入確率の推定値 $\Phi(X'_{2i} \gamma)$ をプロットしたのが図3である。

総支出金額が非常に高い世帯と低い世帯では記入確率が低下している。このことは統計実務家の間で頻繁に指摘される事柄と一致する。

その全世帯の平均値をとると0.640となる。以前、Maki-Nishiyama(1993)で示された耐久消費財の過小性の計測結果は0.672であった。全く違った研究手法を採用したにもかかわらず双方の結果は予想以上に一致している。「家事用耐久消費財」ではなく「教養娯楽用耐久財」、「紳士婦人用被服」で試みても結果は同様に整合的である点を付け加えておこう。

こうした結果をみると、Pi-Tobitモデルは、Deaton-Irishの示した原モデルの改良になっているものと判断される。

3 計算作業で遭遇した問題点と感想

本稿で紹介した研究を進める際に筆者が遭遇した問題点と感想を最後に要約しておこう。これが本学情報処理センターの一層の充実を期するうえで何らかの材料を提供するならば幸いである。

理工学系の研究では実験観測を通じてデータは自ら収集することが多い。他方、社会科学系の研究では対象が社会現象である。データ収集のための制御された実験は不可能であり、統計調査が主たる方法となる。その統計調査は政策策定上の必要性から主として政府の統計部局が実施している。プライバシー保護の観点もあって研究者自らが詳細な情報を収集することは一般に困難である。そこで先ずデータ処理上の問題となるのは、「現在、最も流通性のあるファイル媒体とファイル形式は何か」である。

多くの政府部局では、徐々にWork Station、Personal Computerが浸透してきてはいるが、主な業務は依然として大型汎用機で行っている。データ保管とデータ交換には、いわゆる2分の1インチ磁気テープが使われることが多く「全国消費実態調査」特別集計結果も、この媒体が使われている。ファイル形式はF形式、FB形式、VB形式が混在している。また、複数のボリュームにまたがるファイルも存在した。最初の課題は、これらの汎用機上で作成されたファイルを本学情報処理センターの電算機でいかに処理するか、であった。

残念ながら「ファイルの切出し」作業は北大大型計算機センターで進めることになった。本学内の施設で処理することが不可能とは言えないまでも、門外漢の筆者にとって、本計算に至るまでの時間を節約することが何よりも優先された。具体的に言うと、北大センター内のHITACM800/210でSASにより必要な情報のみを原ファイルから切り出し、それをネットワーク経由で本学のgeminiに転送した。北大センターでもテープドライバーをデバイス指定できないとか、占有ディスク容量制限があったりとかで、ジョブクラス「Y」を事前申請する必要が伴う。このあたり、官公庁の情報処理施設が時代に遅れているとも言えるし、大学側の施設が先に進み過ぎているとも感じられる点である。

最初に行ったことは、北海道在住約3000世帯について、既述のTobitモデルを最尤推定することであった。このような作業では、まずMATLAB、TSP、GAUSSなどの数値計算パッケージを利用して部分サンプルを対象に解を求めておくと便利だからである。計算時間の上からそれらが「使い物」になるか、Fortranプログラムを自前で作る必要があるかについて概ね見当がつく。筆者の研究室にはMacintosh Quadra 840AVがある。統計解析用パッケージとしてはTSP等が利用でき、ABSOF社製FORTRANコンパイラが入っている。そこでTSPを用いてTobitモデルの最尤推定を進めると同時にFortranプログラムを組むことにした。同一のソース・プログラムをパソコン、gemini、北大センターの大型汎用機等で動作させれば計算時間の大体の目安がつく。

非線形最適化のFortranルーティンとしてはH. Press他による“Numerical Recipes in Fortran”所載のPowell's Direction Set法を利用した。これは計算速度の点でそれほどの利点はない。また目的関数の定義は紙数の制約から掲載を控えるが、各世帯による財の購入が正值かどうかを条件判断したうえで、対数尤度の寄与を計算するというやり方をとった。これも思い付きやすい組み方だが、計算時間を大きく消費する原因となる。

実際に、北海道在住約3000世帯を対象にして計算すると、Fortranによる計算時間がTSPのそれを上回った。双方の計算結果は一致する。それにしても、TSPは二次微係数まで計算していること、対数関数の実引数が負になった時の処理など諸々のハウスキーピング的作業までソフトウェア側で行っているにもかかわらず、計算本体のルーティンしか記述していないFortranプログラムよりも高速であることには驚く人も多いのではなかろうか。加えて、サブルーティン・ライブラリが想定していないケースに対応するため、提供されているFortranソースを修正するという作業を強いられてもいるのである。ここから得られる教訓の第1は、

『素人が下手なプログラムを組むほど時間の浪費はない』

である。

一体、計算機の水準に応じて、どの程度の計算時間の差が生じるものだろうか。同じプログラムをMacintosh、gemini、北大センターの汎用機（M800/210）とスーパーコンピューター（HITAC S-3800）で動作させてみた。結果は、Quadra 840AVが約3時間、geminiが20分程度、大型汎用機が3分強、スーパーコンが2分余であった。メモリ上で反復計算しているだけであるからディスク間転送効率などは関係ない。それにしても数値計算は不得手と言われるMacintoshだけあって、その非効率性は歴然としている。また、汎用機の10倍と言われるSuper Computerの能力を十分に引き出すためには、プログラムの組み方を工夫しないと駄目であることもよく分かる。教訓の第2は

『大型といわれるコンピューターは評判だけのことはある』

であろうか。

それでは最も高速な大型機を日常的に利用したい気になるかと言えば、決してそうではない。問題はプログラム作成環境にある。UNIX上のNEmacs、f77、make、dbxの組み合わせは慣れると非常に効率的である。しかも同一の環境がUNIX機ではほぼ例外なく利用できるのである。大型汎用機は高速であるが、編集環境は機種固有のものである。F社とH社の統合環境エディターも似ているようでいて操作法は非常に異なる。この辺りがユーザーが減少する一因になっているのではあるまいか。またスーパーコンではUNIXベースのOSを採用することが最近は多いときだが、北大センターのhopについていえば、Fortran77コンパイラが標準仕様そのものであり、拡張機能はないという点が使いづらかった。Fortran90が搭載されると事情は多少異なると思うが（既にサポートされているのだろうか...）。

本稿の図1から図3はcanalに搭載されている「S」を利用して描画したものである。その際、まずSからPostscriptコードをファイルに出力し、それをファイル変換ののちMacintoshから手元のPostscriptプリンターにダウンロードしたものである。流通性のあるファイル形式にはこうした利点もある。そこで教訓の第3。

『コンピューターの世界でも、寄らば大樹の蔭、が妥当する』。

参考文献

[1] William H. Press, Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. "NUMERICAL RECIPES in FORTRAN, SECOND EDITION", Cambridge U.P., 1992

[2] Angus Deaton, Irish, M. "Statistical Models for Zero Expenditures in Household Budgets", Journal of Public Economics, 23, pp.59-77, 1984

[3] Atsushi Maki, Nishiyama, S. "Consistency Between Macro- and Micro-data Sets in the Japanese Household Sector", Review of Income and Wealth, 39-2, pp.195-207, 1993

Figure 1. Share of Electric Appliances

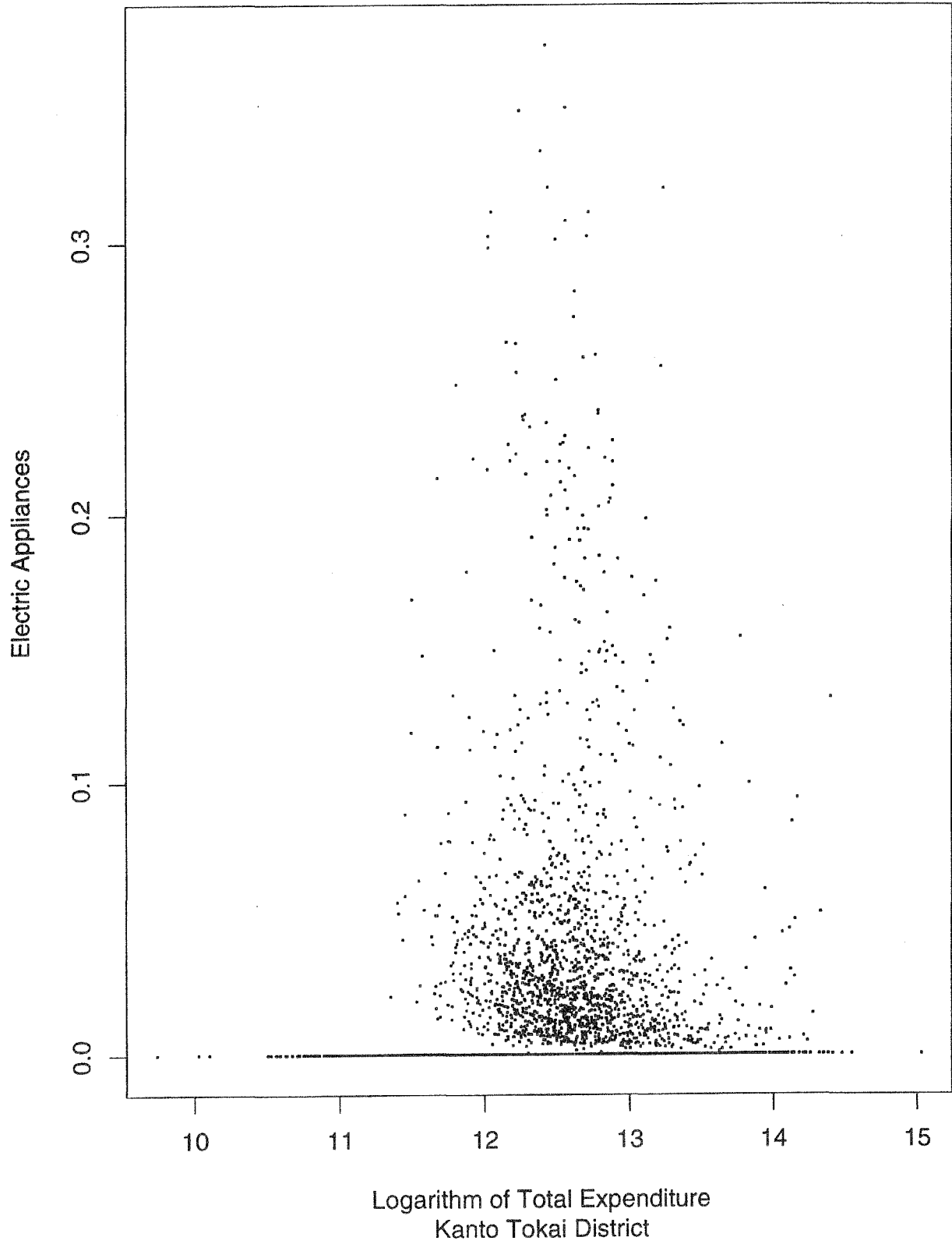


Figure 2. Zero Probability of Electric Appliances

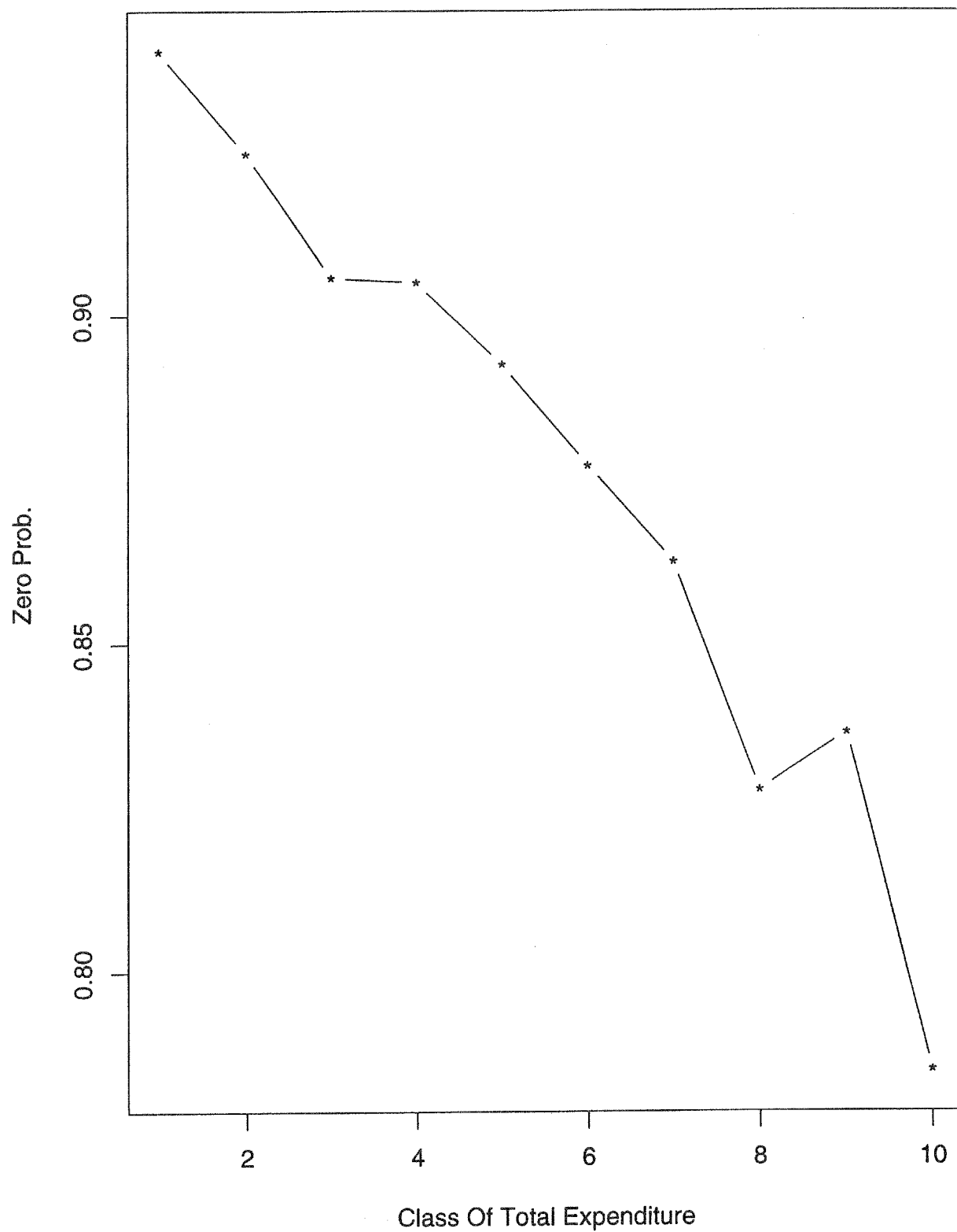


Figure 3. Total Expenditure and P Factor

