

The College Of
WILLIAM & MARY

Revisiting the Price Elasticity of Gasoline Demand

Alfredo A. Romero*
College of William and Mary

College of William and Mary
Department of Economics
Working Paper Number 63

October 2007

COLLEGE OF WILLIAM AND MARY

* Alfredo Romero is a Visiting Assistant Professor of Economics at the College of William and Mary and a Ph.D. candidate in Economics at Virginia Polytechnic Institute and State University.

DEPARTMENT OF ECONOMICS
WORKING PAPER # 63
October 2007

Revisiting the Price Elasticity of Gasoline Demand

Abstract

In this document, we investigate the evolution of the income elasticity and the price elasticity of the demand for gasoline over the period 1975-2006. By using the Probabilistic Reduction Approach, we were able to model changes in mean heterogeneity and variance heterogeneity directly into the model. This method allowed us to determine the timing and the size of shifts in the elasticities. Our estimates are consistent with the current literature: there has been a shift in the price elasticity of gasoline demand. This shift, not present in the income elasticity, occurred almost at the beginning of the period of study. We use these estimates to compute several welfare measures. We also present a sketch of the relationship between a Monthly Fixed Effect panel data model and a Time Series Model with Monthly Dummy Variables.

JEL Codes: D12, L91, Q31, Q41, C22, C23, C50

Keywords: Price Elasticity, Income Elasticity, Gasoline Demand, Time Heterogeneity, Dummy Variables, Monthly Fixed Effects Panel Data Models, Time Series Models, Compensating Variation, Deadweight Loss.

Alfredo A. Romero
Department of Economics
College of William and Mary
Williamsburg, VA 23187-8795
aromeroa@vt.edu
aaromero@wm.edu

1 Introduction

Consistent estimation of the price elasticity of the demand of gasoline is a perennial need in economic analysis. Most of the research in the area, however, has focused on pre-1980's data. The drastic changes to the oil markets, and its impact on gasoline prices that have occurred since the late 1980's, and that have placed the price of oil at record highs since 2006, make even more important the study of the changes in the demand conditions for the gasoline market. In the present document, we will study the change in the price elasticity of gasoline over the period 1975-2006 by accommodating structural changes in both the conditional mean, and the conditional variance equations. This approach would allow us to pinpoint the timing and the size of shifts in the estimated elasticities. This document attempts to follow and expand the work of Hughes, Knittel, and Sperling (2006) but uses the Probabilistic Reduction Approach Modeling Strategy. The paper is divided as follows, section 2 surveys the nonstationarity of the data, section 3 introduces the Probabilistic Reduction Approach and its use in modeling time heterogeneity, section 4 digresses on the probabilistic assumptions required for a particular type of fixed effects model, the Monthly Fixed Effects model, section 5 presents the empirical results, section 6 applies the results to the computation of some welfare measures, and lastly, section 7 presents the conclusions.

2 Nonstationarity in the Demand for Gasoline

In their 2006 working paper (now 2008 The Energy Journal publication), Hughes, Knittel, and Sperling (HKS), do an excellent job providing evidence that the price elasticity of Gasoline Demand has shifted from the period of 1975-1980 to the 2001-2006 period. In their study, they find that the short-run prices elasticities differ considerably between periods, ranging from -0.034 to -0.077 during 2001 to 2006, versus -0.21 to -0.75 for 1975 to 1980[1]. Their estimation is 'robust' to the inclusion of additional variables, such as unemployment and interest rate, as well as to changes in functional form[2].

With respect to the price elasticity, several issues are not addressed by HKS. One of them is the timing of the shift. A second issue is the uniqueness of the shift. Additionally, questions regarding shifts in the income elasticity are left unresolved for their model did not provide evidence of a shift in the income elasticity. To have an idea of the time consistency of both estimators, we conducted two thought experiments. In the first one, we ran several

[1]They rationale for choosing this two periods in particular is the fact that both periods present similar price movements, allowing a semi-experimental framework for testing the difference in price elasticities.

[2]Their basic model specifies the log of gasoline consumption as a function of the log of price and the log of income in a fixed effects panel data framework. Specifically,

$$\ln G_{jt} = \beta_0 + \beta_1 \ln P_{jt} + \beta_2 \ln Y_{jt} + e_j + e_{jt}$$

where G is per capita gasoline consumption in gallons in month j and year t, P is the real retail price of gasoline in month j and year t, Y is real per capita disposable income in month j and year t, e_j represents unobserved demand factors that vary at the month level and e_{jt} is a mean zero error term. For a more detailed description of the data used, see HKS.

regressions with 60 observations each (following the basic model specification of HKS), using a moving window estimator over the entire sample, obtaining a set of estimators for both, the price elasticity and the income elasticity. These naive window estimators reveal their lack of time stationarity. In the second experiment, we ran cumulative regressions, starting with the first 60 and adding one observation at the time, also using the basic model specification of HKS. This naive procedure also reveals the existence of nonstationarity in the estimators.

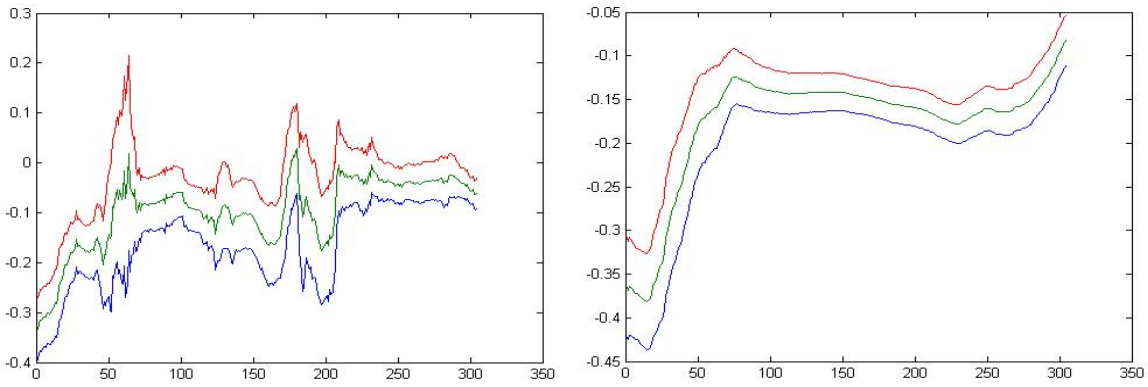


Figure 1: Window Estimators and Cumulative Estimators for Price

The left Panel of Figure 1 reveals the shift in the price elasticity reported in HKS. The first observation, with its 95 percent confidence interval, corresponds to the estimation with the first period data in their analysis. Similarly, the last observation corresponds to the estimation of the second period in their analysis. Clearly, their claim of a shift in the price elasticity is sound since both estimators' 95 percent confidence interval do not overlap. Two additional remarks are in order. As the right panel of Figure 1 suggests, the shift occurred earlier in the sample, at around observation 75 (Jan 1982). Note, however, after the shift, the time consistency of the estimator has not improved, while hovering around -0.05, its

consistency is questionable, as it can be seen from the right panel of the same figure.

With respect to income elasticity, the results are even more shocking. Note that for the first and the last observations, in the left panel of Figure 2, it is not possible to find, in this first graphical exercise, that both estimators are statistically different, just as expressed by HKS; however, shifts in the income elasticity have occurred in an apparent recurring fashion[3]. The set of estimators indicate that the income elasticity has evolved from 0.4 to 0.0; from 0.3 to 0.0; and from 0.5 to 0.3[4]. The time inconsistency of this estimator is corroborated by the cumulative regressions, shown in the right panel.

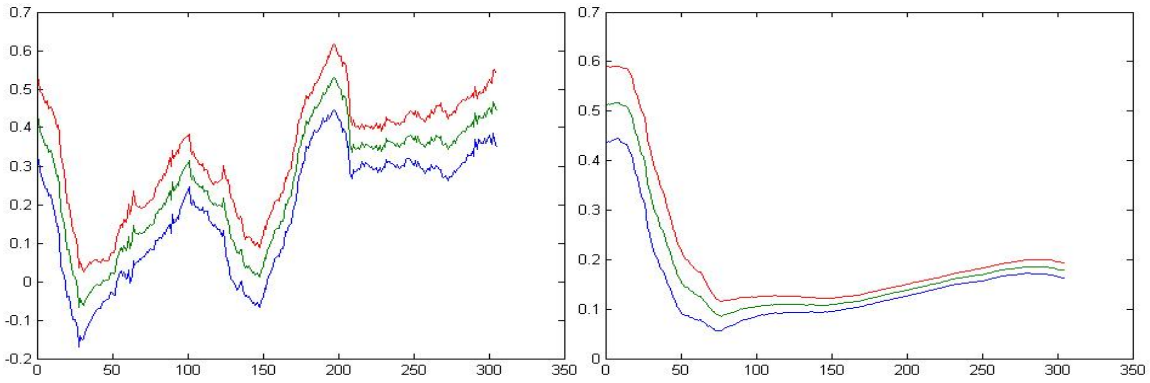


Figure 2: Window Estimators and Cumulative Estimators for Income

In order to produce a time series model for the entire sample, first, we need to deal with the nonstationarity of the estimators. One way to tackle this problem is by differencing the variables; however, we would lose the temporal structure of the data. The approach followed in this document will be that of modeling the heterogeneity of the data directly.

[3]The true nonstationarity of the income elasticity estimator could be masked by the nonstationarity of the price elasticity estimator.

[4]Of course, the precision of these numbers is compromised for exposition's sake.

3 Modeling Strategy

3.1 The Probabilistic Reduction Approach

The Probabilistic Reduction (PR) Approach (Spanos 1986, 1998) aims to construct statistical models by systematically studying "economic phenomena using observed data in the context of a statistical framework" (Spanos, 1986). In a nutshell, the PR approach provides a flexible methodology which can accommodate empirical modeling using both experimental and non-experimental data, enhances the reliability and precision of the inferences drawn and enables the modeler to learn about the phenomenon of interest (Spanos and McGuirk, 2001). The approach is firmly rooted within the frequency tradition of statistical inference founded by Fisher, whom defined the main task of statistics as 'the reduction of a large quantity of data to a few numerical values; a reduction which adequately summarizes all the relevant information in the original data' (Fisher, 1922).

An important element of the PR approach is the formalization of Fisher's notion of the reduction of data. As stated by Spanos and McGuirk, the starting point for this formalization is the joint distribution of all the observable random variables involved; the set of observables has been chosen by some theory in conjunction with what aspect of the phenomenon of interest are measurable. Let all the observables involved be denoted by Z_t (an $m \times 1$ vector). Kolmogorov's extension theorem ensures that the probabilistic structure of an observable vector stochastic process is fully described by the joint distribution $D(Z_1, Z_2, , Z_T; \phi)$, for $T > 1$. This distribution demarcates the relevant statistical informa-

tion because it provides the most general description of the potential information contained in the data. Kolmogorov's theorem warrants the existence of not only the process itself but also the 'few numerical values', the parameters ϕ. How few these parameters can be depends crucially on the invariant structure of the process. This reduction would provide the modeler with a statistical representation of the stochastic process, an approximation to the Statistical Generating Mechanism (SGM), the true mathematical/statistical representation of the stochastic process that gave origin to the data. If the actual SGM gives rise to an ever-changing observable processes, as it is the case with nonstationary processes, its reduction potential is very limited. Thus, for the reduction of the data to give rise to applicable models, the observable process should enjoy a certain degree of invariance over $t \in T$. This should be interpreted as requiring that certain measurable aspects of the phenomenon of interest remain invariant or that we know how they are changing with t.

By using the reduction of the data, the PR approach aims to capture the systematic features of the observable phenomenon of interest by modeling the systematic (recurring) information in the observed data. Concisely, it is the systematic study of the data that would drive the 'appropriate' model specification. To reach this ultimate probabilistic model, the PR approach is integrated in four stages: specification, misspecification, respecification, and identification. Whereas the role of the first three stages is the establishment of the relationship between the observable data and the assumptions of the model, it is the fourth stage that would engage the resulting probability model to the underlying economic theory. The resulting model will then be assessed as 'Statistically Adequate' if the data not only satisfies

the probabilistic assumptions but also the economic assumptions of the model.

A statistical model, in this context, would be defined as a set of probabilistically consistent set of assumptions that depicts the observable random variables of the data[5]. Modeling, under the PR approach is driven by an interchange of information between the probabilistic characteristics of the data and a set of consistent probabilistic assumptions.

The starting point of the PR approach modeling strategy is the set of all possible statistical models that could have given rise to the observed data. We can define this set as $Z = (z_1, z_2, ..., z_T)$, where z_t is an $m \times 1$ vector that includes all the observable variables of the model at time t. Given that a complete description of the probabilistic structure of the vector stochastic process $(Z_t, t \in T)$ can be provided by the joint distribution $D(Z_1, Z_2, ..., Z_T; \phi)$, one can characterize all the possible statistical model in relations to this distribution by imposing probabilistic reduction assumptions. These reductions give rise to the statistical model in question. The reduction assumptions come from three broad categories:

$$(D)Distribution \quad (M)Dependence \quad (H)Heterogeneity$$

By combining different reduction assumption one can generate a wealth of statistical models that would have been impossible to construct otherwise.

[5]From the outset we can discern two important differences with the modeling strategy of the error specification approach. One, the specification is not theory driven. Two, the probabilistic assumptions of the model are warranted by the observable variables (independent and dependent) instead of the unobservable error term.

As an illustration, let us consider the reduction in the case of the Normal/Linear Regression Model. The vector of observables in this case is $Z_t = (y_t, X_t)$ and the reductions assumptions imposed on the process $(Z_t, t \in T)$ are *(D) Normal*, *(M) Independent*, *(H) Identically Distributed*. This takes the form

$$D(Z_1, Z_2, ...Z_T; \phi) =^{I} \Pi_{t=1}^{T} D_t(Z_t; \phi_t) =^{IID} \Pi_{t=1}^{T} D(Z_t; \phi) =^{IID} \Pi_{t=1}^{T} D(y_t|X_t; \phi_1) \times D(X_t; \phi_2),,$$

$$\forall (x_t, y_t) \in \Re_X^k \times \Re_Y$$

Note the role of each of the reduction assumptions and the reparameterization/restriction from primary parameters ϕ to the model parameters ϕ_1, ϕ_2. In order to be able to disregard the marginal distribution $D(X_t; \phi_2)$ and concentrate exclusively on $\Pi_{t=1}^{T} D(y_t|X_t; \phi_1)$, we need the reduction assumption of normality for $(Z_t, t \in T)$[6].

3.1.1 Modeling from Left to Right

As stated above, the PR approach modeling strategy differs from the error specification approach with respect to the direction of the modeling. Modeling in econometrics is sometimes characterized by a process flowing from right to left. That is, a particular set of data (observed independent variables) is used to explain another set of data (observed dependent variables) by specifying a functional form of the independent variables that, along with some observational error, gave birth to the dependent variable in question. To make this relationship statistically meaningful, additional statistical assumptions are attributed to the

[6]The efficiency of partitioning the set of all possible models should be contrasted with the traditional way of statistical model specification which attempts to exhaust it using ad-hoc modifications.

error. In a bi-variate linear relationship, for instance, the dependent variable y_t is the result of adding $f(x_t)$ and u_t, where the former is a linear function on x_t. Thus, the relationship flows as follows,

$$y_t \leftarrow f(x_t) + u_t$$

Note that for this model to be statistically meaningful, probabilistic assumptions regarding (u_t) will be transferred onto y_t. If the assumptions for u_t are those of a normal, independent, and identically distributed process, the normal linear regression model with two variables arises, i.e.,

Table 1: OLS Error Assumptions

SGM: $y_t = \beta_0 + \beta_1 x_t + u_t$

(1) Normality $(u_t|X_t = x_t) \sim N(.,.)$

(2) Zero mean $E(u_t|X_t = x_t) = 0$

(3) Homoskedasticity $E(u^2|X_y = x_t) = \sigma^2$

(4) No autocorrelation $E(u_t u_s|X_t = x_t) = 0$ for $t \neq s$

The PR approach modeling strategy, that would also give rise to the equivalent assumptions about the error, is the result of modeling from left to right. That is, by starting with the dependent variable and by supplying in it with equivalent probabilistic assumptions, it is possible to decompose the observed data (independent variable) into two orthogonal components. This decomposition, analogous to the orthogonal projection theorem (Luenberger 1969), relates orthogonal projections and conditional expectations following the framework that Kolmogorov set while extending the work of Wold (1938). In the orthogonal decomposition modeling strategy, obtaining the SGM is the result of relating the first stochastic conditional moment of a response random variable Y relative to the information set D. In this fashion, it is possible to orthogonally decompose the observed dependent variable as a first component that can be explained entirely by its conditioning information set, in this case, $X = x$, called the systematic component, and a second component that contains no statistical information regarding Y. That is,

$$Y = E(Y|D) + u$$

where $E(Y|D)$ represents the systematic component and $u = (Y - E(Y|D))$ represents the non-systematic component. For this decomposition to be orthogonal, the following assumptions must be satisfied[7],

[7]Proofs of these conditions are shown in Appendix I.

(i) $E(u|D) = 0$

(ii) $E(u^2|D) = Var(Y|D) < \infty$

(iii) $E(u \times [E(Y|D)]) = 0$

Given that $E(u^2|D) = Var(Y|D) < \infty$, the conditional variance of Y_t can also be specified as an orthogonal decomposition, that is, letting $u^2 = Var(u|D) = Var(Y|D)$, then

$$u^2 = Var(Y|D) + v$$

Hence, a regression model of the process Y based on its first two conditional moments is given by,

$Y_t = E(Y_t|D) + u_t$

$u_t^2 = Var(Y_t|D) + v_t$

Now it is evident that in order to attain a specific linear specification, the probabilistic assumption would fall into the observables (dependent and independent variables) rather than into the non-systematic component (error). Here is where the equivalence statements for assumptions (1)-(4) would take place. In particular, continuing with our bivariate model, letting the information set be $D := X_t = x_t$, we can consider the following probabilistic assumption regarding y_t, x_t. Let

$$\begin{pmatrix} \mathrm{y}_t \\ \mathrm{x}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right), \forall t \in T$$

where the marginal and conditional distributions of x and y are also normally distributed. That being the case, then $E(y_t|X_t = x_t) = \frac{\sigma_{12}}{\sigma_{22}}x_t + \mu_1 - \frac{\sigma_{12}}{\sigma_{22}}\mu_2$ and $Var(y_t|X_t = x_t) = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$. By reparameterizing this conditional moments, it is possible to obtain the linear representation of the conditional mean in the normal linear regression model, that is, letting $\beta_1 = \frac{\sigma_{12}}{\sigma_{22}}$, $\beta_0 = \mu_1 - \beta_1\mu_2$, and $\sigma^2 = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$, we have,

$$E(y_t|X_t = x_t) = \beta_0 + \beta_1 x_t \text{ and } Var(y_t|X_t = x_t) = \sigma^2.$$

Thus, given the orthogonal decomposition $Y = E(Y|D) + u$, and the fact that $E(y_t|X_t = x_t) = \beta_0 + \beta_1 x_t$, we get the normal linear specification, with the implied assumptions[8],

---

Table 2: Normal Linear Regression Model Assumptions

SGM: $y_t = \beta_0 + \beta_1 x_t + u_t$

[1] Normality $(y_t|X_t = x_t) \sim N(.,.)$

[2] Linearity $E(y_t|X_t = x_t) = \beta_0 + \beta_1 x_t$

[3] Homoskedasticity $Var(y_t|X_t = x_t) = \sigma^2$

[4] No autocorrelation $Cov(y_t y_s|X_t = x_t) = 0$ for $t \neq s$

[8]The derivation of these equivalence of the assumptions is straightforward. For [3] see proof of (ii) in the appendix.

Extending the bivariate linear regression model to the multivariate linear regression model, we have,

$$\begin{pmatrix} \mathrm{y}_t \\ \boldsymbol{x_t} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \boldsymbol{\mu_2} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma_{12}} \\ \boldsymbol{\sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{pmatrix}\right), \forall t \in T$$

resulting in the following SGM: $y_t = \boldsymbol{\beta^T x_t} + u_t$,

where the parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \sigma^2)$ are time invariant. The extension of the reparameterization to the multivariate case is straightforward,

$\boldsymbol{\beta} \equiv \boldsymbol{\Sigma_{22}^{-1}\sigma_{21}}$, and $\sigma^2 = \sigma_{11} - \boldsymbol{\sigma_{12}\Sigma_{22}^{-1}\sigma_{21}}$. Note that the identical distribution assumption warrants these coefficients to be independent of t. Thus, the joint distribution of the process, assuming normality, is, $\boldsymbol{Z_t} \sim N(\boldsymbol{m}, \boldsymbol{\Sigma})$.

3.2 Nonstationary of the Regressors: Time Heterogeneity

The existence of nonstationary implies that the joint distribution of the underlying process is not constant, it changes with t in either a stochastic or a deterministic fashion. That is, $\boldsymbol{Z_t} \sim N(\boldsymbol{m(t)}, \boldsymbol{\Sigma(t)})$.

In the multivariate linear regression framework this implies,

$$\begin{pmatrix} \mathrm{y}_t \\ \boldsymbol{x_t} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1(t) \\ \boldsymbol{\mu_2(t)} \end{pmatrix}, \begin{pmatrix} \sigma_{11}(t) & \boldsymbol{\sigma_{12}(t)} \\ \boldsymbol{\sigma_{21}(t)} & \boldsymbol{\Sigma_{22}(t)} \end{pmatrix}\right), \forall t \in T$$

with $E(y_t|\boldsymbol{X_t} = \boldsymbol{x_t}) = \boldsymbol{\beta_t^T x_t}$, and $Var(y_t|\boldsymbol{X_t} = \boldsymbol{x_t}) = \sigma_t^2$.

These parameters cannot be estimated since they change with t. In particular,

$$\beta_{(0)t} = \mu_1(t) - \boldsymbol{\sigma_{12}(t)\Sigma_{22}^{-1}(t)\sigma_{21}(t)\mu_2(t)}$$

$$\beta_{(1)t} = \boldsymbol{\Sigma_{22}^{-1}(t)\sigma_{21}(t)}$$

$$\sigma_t^2 = \sigma_{11}(t) - \boldsymbol{\sigma_{12}(t)\Sigma_{22}^{-1}(t)\sigma_{21}(t)}.$$

It is clear that, unless we have information regarding the source of the heterogeneity, the degree of summarization of the data performed by the PR approach, or any other economic approach for that matter, would be negligible. Assuming that some information can be drawn from the data (see below), we can derive two ways to deal with the problem of time heterogeneity. The first one, some (or all) of the parameters are changing with the index, will lead to the specification of time trend models. The second one, some (or all) of the parameters are changing in temporal clusters, will lead to the specification of dummy variables models. Note that, so far, nothing has been said regarding the source of the heterogeneity. In the present normal linear regression model, it can occur in the conditional mean specification, in the conditional variance specification, or in both.

3.2.1 Trends in the Conditional Mean and the Conditional Variance

The addition of a linear trend in a normal linear regression model has been treated as the addition of another conditioning variable to the model. Johnston and Dinardo (1997) argue that "the time variable may be treated as a fixed regressor."

Another approach, also potentially misleading, is that of Wooldridge's (2001) conditioning set for the model $y_{it} = c_i + g_i t + x_{it}\beta + u_{it}$, that is

$$E(y_{it}|x_{i1}, \ldots, x_{iT}, c_i, g_i) = E(y_{it}|x_{it}, c_i, g_i) = c_i + g_i + x_{it}\beta,$$

clearly, defining the conditioning set by the set of observables (the x's), and the parameters to be estimated c_i and g_i. Unless Bayesian approach is assumed, the previous conditioning does not have any meaning in statistical grounds. We will attempt to show that the existence of time trends and their estimators is not the result of conditioning on a time trend or conditioning on the time trend parameters. The existence of time trends is granted by the joint distribution of the observables, the x's and the specification of their moments. If the expected value of y_t given x_t is changing with t, that must be the result of the expected value of the marginal distribution of y_t being a function of t, the expected value of the marginal distribution of x_t being a function of t, or both. With the previous framework, it is possible to shed some light on this fact. For illustration purposes, consider again the bivariate normal linear regression model and let the expected value of the marginal distribution of y_t be a linear function of t, that is, $E(y_t) = \mu_1 + \delta t$. Thus, the joint distribution of (y_t, x_t) can be represented by,

$$\begin{pmatrix} \mathrm{y}_t \\ \mathrm{x}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 + \delta t \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right), \forall t \in T$$

The addition of the mean heterogeneity renders the following conditional mean and conditional variance, $E(y_t|X_t = x_t) = \frac{\sigma_{12}}{\sigma_{22}}x_t + \mu_1 + \delta t - \frac{\sigma_{12}}{\sigma_{22}}\mu_2$ and $Var(y_t|X_t = x_t) = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$.

Note that, as stated before, the heterogeneity is present in the conditional mean but not in the conditional variance. We can reparameterize the model to attain,

$\beta_1 = \frac{\sigma_{12}}{\sigma_{22}}$, $\beta_0 = \mu_1 - \beta_1\mu_2$, $\beta_2 = \delta$, and $\sigma^2 = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$

yielding the normal linear model with a linear trend specification,

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 t + e_t,$$

as required, with the model assumption given in table 1. Given that there is no source of heterogeneity in the conditional variance, statistical inference would be valid, granted that the other assumptions of the normal/linear regression model hold.

The previous discussion reveals one of the hidden dangers of the error specification approach, the fact that the probabilistic assumption of the error does not rule out the fact that the parameters may be changing with t. To illustrate, let the source of the time heterogeneity be present in the variance-covariance matrix of the joint probability distribution of y_t,x_t. That is,

$$\begin{pmatrix} \mathrm{y}_t \\ \mathrm{x}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11}\delta t & \sigma_{12}\delta t \\ \sigma_{21}\delta t & \sigma_{22}\delta t \end{pmatrix}\right), \forall t \in T$$

This would result in

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

where, as before, $\beta_1 = \dfrac{\sigma_{12}}{\sigma_{22}}$ and $\beta_0 = \mu_1 - \beta_1 \mu_2$, and with the model assumption given in table 1.

Thus, a researcher using OLS would find no evidence of any linear trend being present in the regression equation and would conduct invalid statistical inferences in the parameters. This would be the result of not taking into consideration the fact that the conditional variance has an additional parameter to be estimated, δ, and that the variance-covariance matrix is changing with t, that is, $\sigma^2 = \delta t \left(\sigma_{11} - \dfrac{\sigma_{12}^2}{\sigma_{22}} \right)$[9]. The appropriate statistical inference procedure would entail $b \sim N \left(\beta, \delta t \left(\sigma_{11} - \dfrac{\sigma_{12}^2}{\sigma_{22}} \right) (X'X) \right)$.

Thus, in order to guarantee significance of the inference procedure, it is necessary to add an additional assumption in the specification of the normal linear regression model: time invariance of the parameters. Clearly, both sets of assumption can be equivalently stated as,

[9] In this particular case, $t > 1$.

Table 3

(1) Normality $(u_t\|X_t = x_t) \sim N(.,.)$	[1] Normality $(y_t\|X_t = x_t) \sim N(.,.)$
(2) Zero mean $E(u_t\|X_t = x_t) = 0$	[2] Linearity $E(y_t\|X_t = x_t) = \beta_0 + \beta_1 x_t$
(3) Homosk. $E(u^2\|X_y = x_t) = \sigma^2$	[3] Homosk. $Var(y_t\|X_t = x_t) = \sigma^2$
(4) No Autocorr. $E(u_t u_s\|X_t = x_t) = 0$ for $t \neq s$	[4] No Autocorr. $Cov(y_t y_s\|X_t = x_t) = 0$ for $t \neq s$
	[5] t-Invar. $\theta = (\beta_0, \beta_1, \sigma^2)$ not changing with t

3.2.2 Dummy Variables in the Conditional Mean and the Conditional Variance

The usual way to compensate for a shift in the mean of a stochastic process has been to accommodate a dummy variable that would capture that shift. This would render the following specification $y_t = \beta_0 + \beta_1 x_t + \beta_2 D + e_t$. If the shift occurred exclusively in the mean of the process, and all the other assumptions of the normal linear regression model are satisfied, then usual t-tests and, in general, inferences regarding the estimators will be valid. This is so because the addition of the dummy variable in the conditional mean does not impose any changes in the specification of the conditional variance of the joint distribution.

To be precise, in our bivariate normal linear regression model, such specification would arise from letting the mean of either x_t, or y_t, or both to shift. For illustration purposes, we would allow the shift to occur in the mean of y_t. In this case, we would have,

$$\begin{pmatrix} \mathrm{y}_t \\ \mathrm{x}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 + \delta D \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right), \forall t \in T$$

The addition of the mean heterogeneity renders the following conditional mean and conditional variance, $E(y_t|X_t = x_t) = \frac{\sigma_{12}}{\sigma_{22}}x_t + \mu_1 + \delta D - \frac{\sigma_{12}}{\sigma_{22}}\mu_2$ and $Var(y_t|X_t = x_t) = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$. Note that, as stated before, the heterogeneity is present in the conditional mean but not in the conditional variance. We can reparameterize the model to attain,

$\beta_1 = \frac{\sigma_{12}}{\sigma_{22}}$, $\beta_0 = \mu_1 - \beta_1\mu_2$, $\beta_2 = \delta$, and $\sigma^2 = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$.

The specification would be

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 D + e_t,$$

as required, with the model assumption given in the table 3. Given that there is no source of heterogeneity in the conditional variance, the t-test would be valid, granted that the other assumptions of the normal/linear regression model are valid.

Unfortunately, it could be the case that the addition of the dummy variable in the conditional mean is not the result of a shift in the mean of the process but rather a shift in the covariance of x_t and y_t. This different source of heterogeneity would cause the conditional mean to have multiplicative interaction between the regressors and the dummy variables while affecting the conditional variance of the process. For instance, consider the case where the source of heterogeneity occurs in the covariance of x_t and y_t,

$$\begin{pmatrix} \mathrm{y}_t \\ \mathrm{x}_t \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} + \delta D \\ \sigma_{21} + \delta D & \sigma_{22} \end{pmatrix}\right), \forall t \in T.$$

In this case, $E(y_t|X_t = x_t) = \frac{\sigma_{12}}{\sigma_{22}}x_t + \frac{\delta}{\sigma_{22}}Dx_t + \mu_1 - \frac{\sigma_{12}}{\sigma 22}\mu_2 - \frac{\delta D}{\sigma_{22}}\mu_2$, and $Var(y_t|X_t = x_t) = \sigma_{11} - \frac{(\sigma_{12} + \delta D)^2}{\sigma_{22}}$. Reparameterizing the model, we get the dummy variable with interaction specification

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 Dx_t,$$

with the model assumptions given in table 3.

The distribution of the coefficients is still normal; however, their variance would change as the dummy variable shifts, that is, $b \sim N(\beta, \sigma(X'X))$, where $\sigma = \sigma_{11} - \frac{(\sigma_{12} + \delta D)^2}{\sigma_{22}}$.

4 Digression: Where does the Monthly Fixed Effects model come from?

As stated above, HKS utilized a Monthly Fixed Effect model for their analysis, a model that has wide acceptance in the current econometric literature. The basic premise is that the fixed effects components of the specification captures heterogeneity across the $i's$ but remains constant along t. One may want to shed some light on what is the set of probabilistic assumptions that can give birth to this specification. We will present, following Spanos, that the Monthly Fixed Effects model arises from a multivariate normal distribution with mean heterogeneity, similar to the mean heterogeneity discussed above. Furthermore, we will show that, for the case of the Monthly Fixed Effects model, the model can alternatively be specified

as a time series regression model with dummy variables for each month.

We can specify a general panel data model as,

$$\begin{pmatrix} \mathrm{y}_{it} \\ \boldsymbol{X_{it}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1(i,t) \\ \boldsymbol{\mu_2}(i,t) \end{pmatrix}, \begin{pmatrix} \sigma_{11}(i,t) & \boldsymbol{\sigma_{12}}(i,t) \\ \boldsymbol{\sigma_{12}}(i,t) & \boldsymbol{\Sigma_{22}}(i,t) \end{pmatrix}\right), \forall t \in T, i \in N$$

As it stands, this specification renders a non-operational model since the number of model parameters increases with both N and T. By imposing probabilistic conditions, it is possible to reduce the number of parameters to render operational models. Consider the vector stochastic process $Z_{it}, i \in N, t \in T$, where $Z_{it} = (y_{it}, X_{it}^T)^T$. The joint distribution of this process is given by $D(Z_{11}, Z_{21}, ..., Z_{NT}; \phi)$. Depending on the restrictions to this stochastic process, different model specifications will arise. The pooled panel data model would be the result of imposing Independence, Identical Distribution, and Normality to the joint distribution. That is,

$D(Z_{11}, Z_{21}, ..., Z_{NT}; \phi) =^I \Pi_{i=1}^N \Pi_{t=1}^T D(Z_{it}; \varphi(i,t)) =^{ID} = \Pi_{i=1}^N \Pi_{t=1}^T D(y_{it}|x_{it}; \varphi_1) \cdot D(X_{it}; \varphi_2) =^N = D(y_{it}|x_{it}\varphi_1)$

rendering,

$$\begin{pmatrix} \mathrm{y}_{it} \\ \boldsymbol{X_{it}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \boldsymbol{\mu_2} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma_{12}} \\ \boldsymbol{\sigma_{12}} & \boldsymbol{\Sigma_{22}} \end{pmatrix}\right), \forall t \in T, i \in N$$

where the model parameters, the pooled panel data parameters, are given by $\boldsymbol{\beta} \equiv \boldsymbol{\Sigma_{22}^{-1}\sigma_{21}}$, and $\sigma^2 = \sigma_{11} - \boldsymbol{\sigma_{12}\Sigma_{22}^{-1}\sigma_{21}}$.

For the case of the Monthly Fixed Effects Model, we would impose Normality and Independence, but the distribution of the stochastic process would not be identically distributed. In fact, the distribution would be homogeneous with respect to t but heterogeneous with respect to i. The reduction is as follows,

$D(Z_{11}, Z_{21}, ..., Z_{NT}; \phi) =^{I} \Pi_{i=1}^{N}\Pi_{t=1}^{T} D(Z_{it}; \varphi(i,t)) =^{T-Homogeneity}= \Pi_{i=1}^{N}\Pi_{t=1}^{T} D(y_{it}|x_{it}\varphi_1(i)) \cdot D(X_{it}; \varphi_2(i)) =^{N}= D(y_{it}|x_{it}(i)\varphi_1)$

rendering,

$$\begin{pmatrix} \mathrm{y}_{it} \\ \boldsymbol{X_{it}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1(i)) \\ \boldsymbol{\mu_2}(i) \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma_{12}} \\ \boldsymbol{\sigma_{12}} & \boldsymbol{\Sigma_{22}} \end{pmatrix}\right), \forall t \in T, i \in N$$

In particular, if the mean heterogeneity occurs in y in the form of monthly seasonal patterns, we would have $\mu_1(i) = \delta_1 Jan + \delta_2 Feb + \delta_3 Mar \ldots \delta_{11} Nov$, $\mu_2(i) = \mu_X$, and the following model parameters,

$\boldsymbol{\beta_1 \equiv \Sigma_{22}^{-1}\sigma_{21}}$ and $\beta_0 = \delta_1 \cdot Jan + \delta_2 \cdot Feb + \ldots + \delta_{11} Nov + \mu_2^T \beta_1$.

The preceding parameterization would render the following conditional mean, $E(y_{it}|X_{it} = x_{it}) = c_0 + \delta_1 \cdot Jan + \delta_2 \cdot Feb + \ldots + \delta_{11} Nov + x_{it}^T \beta$, and thus, the following SGM,

$$y_{it} = c_0 + \delta_1 \cdot Jan + \delta_2 \cdot Feb + \ldots + \delta_{11} Nov + x_{it}^T \beta + u_{it}$$

Note, however, that the β parameter is nothing but the panel data pooled parameter. Through simple vectorization, this model is statistically equivalent to a time series model with monthly dummy variables.

The resulting model is,

$$y_t = c_0 + \delta_1 \cdot Jan + \delta_2 \cdot Feb + \ldots + \delta_{11} Nov + x_t^T \beta + u_t$$

with the model assumption given in table 3.

This specification would be the one adopted in this document to conduct the empirical analysis.

5 Empirical Results

The empirical analysis was conducted in three stages. In the first one, the base model of HKS was applied over the entire sample, and the estimated errors were calculated. In the second stage, the residuals will be used for the detection of changes of the heterogeneity of the process. In the final stage, the time heterogeneity in the overall sample will be captured with dummy variables and time trends. The result of this procedure will be the capture of the time heterogeneity of the errors as well as the satisfaction of the regression conditions in the model.

To start stage one, and following our Time Series Model with Monthly Dummy Variables, we conducted the estimation of the base model over the entire sample. This procedure reveals the presence of heterogeneity in the conditional errors, Figure 3. These residuals would serve as our base for the detection of time heterogeneity in the sample.

Visual inspection of the residuals signals the existence of at least four structural changes in the data[10].

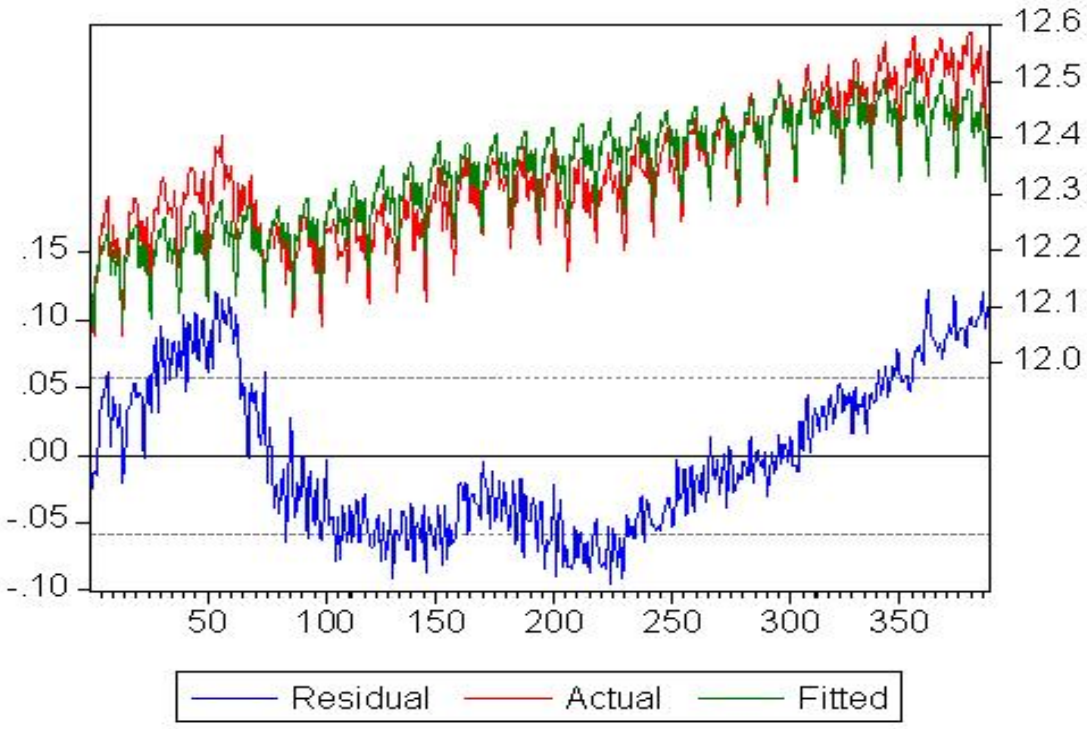


Figure 3: Base model over the entire sample

To detect the breaks in the process, we used three standard stability tests, widely used in the current econometric literature: Recursive Residuals Test, CUSUM of Squares Test, and One-Step Forecast Tests. However, in order to draw valid conclusions from these tests, it was necessary for the competing models to be Statistically Adequate, as defined above. For instance, the first statistically adequate model corresponded to the period 23-60, and contains the logarithm of price and income, the dummy variables for the months[11], and four

[10]The true number of changes in the heterogeneity will be obtained through statistical testing.

[11]These variables would be referred to as the Base Model Variables.

lags of the dependent variable that capture fourth order autocorrelation. The same model was then estimated again over the entire sample and tested for time stability. Failure of the tests indicated the start point of a different time structure for the model. The preceding algorithm identified seven separate breaks in the conditional mean.

Table 4: Structural Breaks Detected

Period	Observation	Dates
I	23-60	Nov 1975 - Dec 1978
II	60-109	Dec 1978 - Jan 1983
III	112-158	Apr 1983 - Feb 1987
IV	158-188	Feb 1987 - Aug 1989
V	188-232	Aug 1989 - Apr 1993
VI	232-310	Apr 1993 - Oct 1999
VII	315-387	Mar 2000 - Mar 2006

The statistically adequate model included all the base model variables, four lags of the dependent variable, and a set of dummy variables for the different time periods interacting with price, income, and a linear time trend, respectively. The resulting model fits the following SGM,

$$g_t = \Sigma_{i=1}^{4}\lambda_i g_{t-1} + \alpha_0 p_t + \Sigma_{i=1}^{6}\alpha_i p_t D_i + \beta_0 y_t + \Sigma_{i=1}^{6}\beta_i y_t D_i + \Sigma_{i}^{11}\gamma_i M_i + \delta_0 t\Sigma_{i=1}^{6} + \delta_i t D_i + e_t.$$

Figure 4 shows how the former specification captures most of the time heterogeneity of the data by adding dummy variables in the conditional variance and thus generating interaction

variables in the conditional mean. These errors, not yet standardized, present no evidence of time dependence of serious problems of serial correlation.

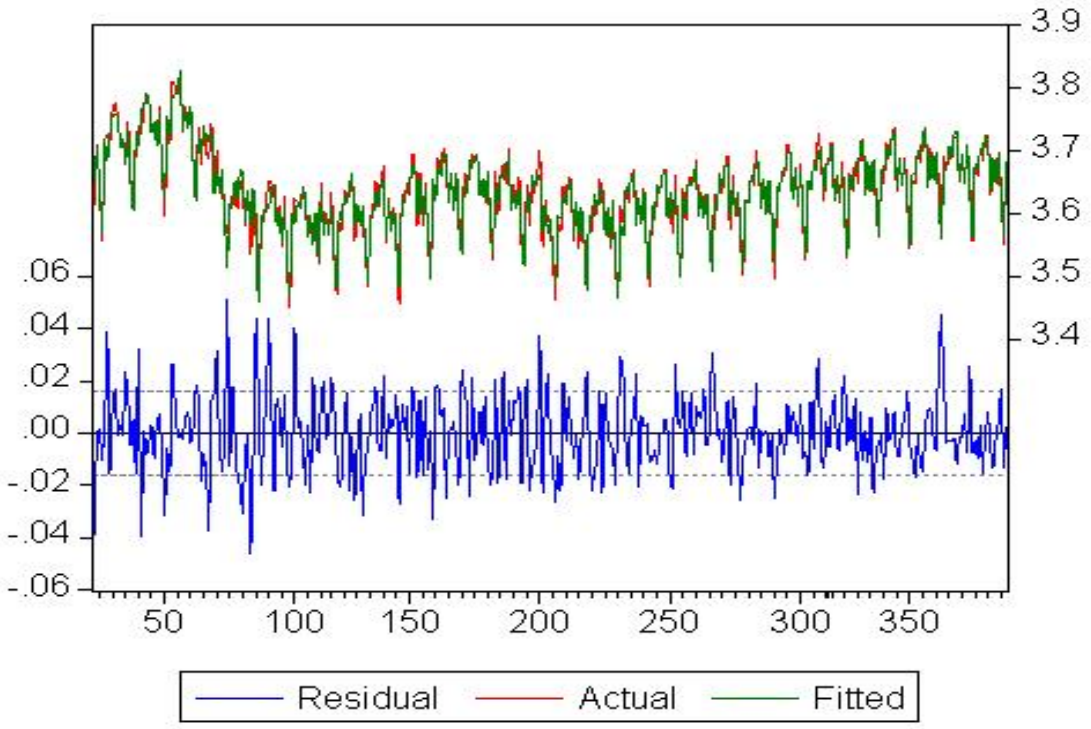


Figure 4: Statistically Adequate Model

The complete results of this regression are stated in Appendix 2. The most important conclusion are reproduced in the following table,

Table 5: Estimates for Price Elasticity and Income Elasticity

Period	Price Elasticity	Income Elasticity
I	0.1641	0.4142
II	-0.1809	0.4049
III	-0.0637	0.3869
IV	-0.0637	0.3906
V	-0.0637	0.3887
VI	-0.0083	0.3617
VII	-0.0637	0.3450

Note how, after capturing the time heterogeneity of the data, it is only the price elasticity what has changed trough time, whereas the income elasticity has remained relatively constant over the entire period.

6 Application

For an application, we would use our estimated demand function to compute the Compensating Variation and the Deadweight Loss resulting from the imposition of a tax. Hausman (1981) has provided a link between the Marshallian Demand of a good and the expenditure function using Roy's identity and the fact that $\frac{\partial e(p,\bar{u})}{\partial p_j} = h_j(p,\bar{u})$. He demonstrates that it is possible to obtain exact measures of the Compensating Variation and the Dead Weight Loss of tax impositions utilizing solely observable market data. Given that

in equilibrium the following condition holds, $u^0 = v(p^0, y^0)$, and the fact that $x_j(p, y) = -\partial v(p, y)/\partial p_j / \partial v(p, y)/\partial y$, Hausman will derive the expenditure function by integrating out the resulting Hicksian Demand.

As the price of a good changes, to remain in the same indifference curve, a compensating variation of income must takes place. That is, $v(p(t), y(t)) = u_0$, where t indicates a price path. Along the price path, we have

$$\frac{\partial v(p(t), y(t))}{\partial p_1(t)} \frac{dp_1(t)}{dt} + \frac{\partial v(p(t), y(t))}{\partial y(t)} \frac{dy(t)}{dt} = 0$$

Using the implicit function theorem and Roy's identity, we can solve for the change in income necessary to maintain the consumer in the same indifference curve,

$$\frac{dy(p)}{dp} = x$$

Depending on the shape of the demand function, we will have different solutions for this differential equation. For our particular case, the Marshallian demand is given by $x = e^{z\gamma} p^{\alpha} y^{\delta}$. The solution to this functional form is provided by Hausman (Eq. 21). Treating the integrating factor as the initial level of utility, and then solving for this level of utility, we find the indirect utility function,

$$v(p, y) = c = -e^{z\gamma} \cdot \frac{p^1 + \alpha}{1 + \alpha} + \frac{y^{1-\delta}}{1 - \delta}$$

With this indirect utility and the first Slutsky condition, it is possible to derive the expenditure function,

$$e(p,\bar{u}) = \left[(1-\delta)\left(\bar{u} + e^{z\gamma}\frac{p^{1+\alpha}}{1+\alpha}\right)\right]^{1/1-\delta}$$

The compensating variation will be defined as $CV(p^0,p^1,y^0) = e(p^1,u^0) - e(p^0,u^0)$. For this particular case,

$$CV(p^0,p^1,y^0) = \left\{\frac{(1-\delta)}{(1+\alpha)y^{0\delta}}\left[p^1x^1(p^1,y^0) - p^0x^0(p^0,y^0)\right] + y^{0(1-\delta)}\right\}^{1/1-\delta} - y^0.$$

From this, the welfare loss can be defined, in a straightforward manner, as the difference between the CV and the tax revenue collected. That is, $DWL = CV(p^0,p^1,y^0) - x(p^1,y^0)\cdot(p^1-p^0)$.

For our model, we utilized the estimators from the last homogeneous period in the sample to compute the CV and the DWL that take place if we were to impose a 10 percent increase tax in the price of gasoline. Notice that our model for the demand, after taking antilogarithms, would be of the form $x = e^{z\gamma}p^{\alpha}y^{\delta}$. We will compute these measures for an increase 10 percent in the price of gas from 2.0 to 2.2, keeping a certain level of income, say 30,000. The results are the following[12],

[12]The estimated demand parameters are $\alpha = -0.043624$ and $\delta = 0.560642$.

Table 6: Computation of Welfare Measures

Tax Imposed	10%
CV	$1,396.86
Tax Revenue	$8.40
DWL	$1,388.45
DWL%	99.39%

7 Conclusion

In this paper, we estimated the price elastic and income elasticity of the demand for gasoline over the period 1975-2006 using the Probabilistic Reduction Approach to capture the time heterogeneity in the data. The approach allowed us to model shifts in both the conditional mean and the conditional variance. We also sketched the probabilistic assumptions needed to create a Fixed Effects Panel Data Model to be used as our estimation framework. Finally, the computation of some welfare measures, such as Deadweight Loss and Compensating Variation were also undertaken.

8 Appendix 1

Proof of (i): Consider

$Y = E(Y|D) + u$

Applying the Law of Iterated Expectation (LIE) we get,

$E(Y|D) = E(E(Y|D))|D) + E(u|D)$

$E(Y|D) = E(Y|D) + E(u|D)$

$E(u|D) = 0$

Proof of (ii): From above, consider[13]

$E(Y|D) = E(Y|D) + E(u|D)$

Now $Var(Y|D) = Var(E(Y|D)) + Var(u|D)$

Note that $Var(E(Y|D)) = E[[E(Y|D) - E[E(Y|D)]]2] = 0$, then $Var(Y|D) = Var(u|D)$.

Proof of (iii): First, note that $E[E[u \times E(Y|D)]|D] = E(u \times [E(Y|D)])$, thus

$E[E[u \times E(Y|D)]|D] = E(Y|D) \times E(u|D) = 0.$

[13]Note that ii also implies that $E(u^2|D) = Var(u|D)$.

9 Appendix 2

Variable	Coefficient	Std. Error	Variable	Coefficient	Std. Error
G1(-1)	-0.1345	0.0527	JAN	-0.0796	0.0064
G1(-3)	0.2405	0.0512	FEB	-0.1380	0.0052
G1(-4)	-0.0904	0.0489	MAR	-0.0454	0.0063
G1(-4)*DM1	-0.1116	0.0519	APR	-0.0163	0.0053
P	-0.0637	0.0127	MAY	0.0306	0.0083
P*DM1	0.2278	0.1677	JUN	-0.0068	0.0104
P*DM2	-0.1172	0.0307	JUL	0.0345	0.0058
P*DM3	-0.0326	0.0292	AUG	0.0309	0.0076
P*DM4	-0.0428	0.0692	SEP	-0.0412	0.0064
P*DM5	-0.0315	0.0395	OCT	-0.0165	0.0054
P*DM6	0.0554	0.0213	NOV	-0.0569	0.0067
Y	0.3450	0.0302	T	0.0004	0.0001
Y*DM1	0.0692	0.0190	T*DM1	0.0009	0.0004
Y*DM2	0.0600	0.0066	T*DM2	-0.0029	0.0004
Y*DM3	0.0420	0.0078	T*DM3	-0.0018	0.0004
Y*DM4	0.0456	0.0092	T*DM4	-0.0018	0.0005
Y*DM5	0.0437	0.0069	T*DM5	-0.0017	0.0003
Y*DM6	0.0167	0.0044	T*DM6	-0.0006	0.0001

Variable	Coefficient
R^2	0.9456
σ_1^2	0.0004
σ_2^2	0.0007
σ_3^2	0.0003
σ_4^2	0.0004
σ_5^2	0.0004
σ_6^2	0.0002
σ_7^2	0.0001

10 References

1. Fisher, R.A. (1922) "On the Mathematical Foundations of Theoretical Statistics." Philos. Trans. Royal Soc. A, 222.

2. Johnston, Jack and John Dinardo (1996),"Econometric Methods," McGraw-Hill/Irwin; 4 edition.

3. Hausman, Jerry A. (1981), "Exact Consumer's Surplus and Deadweight Loss," The American Economic Review, Vol. 71, No. 4.

4. Hughes, Jonathan; Christopher R. Knittel; Daniel Sperling (2006), "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand," NBER Working Paper Series 12530.

5. Luenberger, G.M. (1969) "Optimization by Vector Space Methods," Wiley, New York.

6. Spanos, Aris (1986), "Statistical Foundations of Econometric Modeling," Cambridge University Press, Cambridge.

7. Spanos, Aris (1999), "Probability Theory and Statistical Inference: Econometric Modeling with Observational Data," Cambridge University Press, Cambridge.

8. Spanos, Aris, and Anya McGuirk. "The Model Specification Problem from a Probabilistic Reduction Perspective." American Journal of Agricultural Economics, 83:5 (2001): 1168-1196.

9. Wold (1938) "A Study in the Anlysis of Stationary Time Series," Almquist and Wicksell, Uppsala.

10. Wooldridge, Jeffrey M. (2001), "Econometric Analysis of Cross Section and Panel Data," The MIT Press; 1 edition.