Statistical Adequacy and Reliability of Inference in Regression-like Models

Alfredo A Romero

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Aris Spanos, Chair
Sheryl Ball
Richard Ashley
Chris Parmeter
Randall Billingsley

May 4,2010
Blacksburg, Virginia

Statistical Adequacy and Reliability of Inference in Regression-like Models

Alfredo A Romero

(ABSTRACT)

Using theoretical relations as a source of econometric specifications might lead a researcher to models that do not adequately capture the statistical regularities in the data and do not faithfully represent the phenomenon of interest. In addition, the researcher is unable to disentangle the statistical and substantive sources of error and thus incapable of using the statistical evidence to assess whether the theory, and not the statistical model, is wrong. The Probabilistic Reduction Approach puts forward a modeling strategy in which theory can confront data without compromising the credibility of either one of them. This approach explicitly derives testable assumptions that, along with the standardized residuals, help the researcher assess the precision and reliability of statistical models via misspecification testing. It is argued that only when the statistical source of error is ruled out can the researcher reconcile the theory and the data and establish the theoretical and/or external validity of econometric models.

Through the approach, we are able to derive the properties of Beta regression-like models, appropriate when the researcher deals with rates and proportions or any other random variable with finite support; and of Lognormal models, appropriate when the researcher deals with nonnegative data, and specially important of the estimation of demand elasticities.

Acknowledgments

To my parents, Amado and Lourdes, who taught me that sacrifice and hard work have their benefits, if maybe in the long run.

To my brother, Ivan, who is always there when I need anything.

To my uncles Hector and Jesus, and my grandpa Hector, who saw something good in me and were willing to bet money on it.

To the rest of my family and friends, who anchor me to a big part of my life and remind me that some of us, the lucky ones, should pay it forward.

To my advisors, Dr. Billingsley, Dr. Parmeter, Dr. Ashley, Dr. Ball, and the rest of the faculty and staff in the Economics Department at Virginia Tech, who made an economist out of me, working with so little material.

To my chair, Dr. Aris Spanos, whose discipline, dedication, insight, and passion for econometrics have been a constant source of inspiration. This dissertation would not have been possible without his mentorship.

And to my wife, Jennifer, who shared the good times and the bad times, who supported me and my ideas, and from whom I took so many hours in the pursuit of this goal but who always offered me love at the end of a long day.

Contents

List of Figures

List of Tables

Chapter 2

Chapter 3

Chapter 4

1 Introduction

It is undeniable that the technical development of econometrics has been rapid and impressive; driven in part by the ever-increasing processing power of computers and in part by the adoption and acceptance of general results that allow researchers to posit claims based on asymptotic assumptions. Unfortunately, mathematical sophistication and generalization of results, with the intended goal of streamlining economic modeling, has also had the unintended consequence of confusing the role of econometrics in the validation of economics as a science. Instead of acting as the arbiter of economic theories, the role of econometrics has been relegated to the mere quantification of econometric conjectures presupposed correct. It might seem that econometric models no longer need represent the economic phenomena of interest but rather pay lip-service to complicated theoretical models with even more sophisticated stochastic versions.

The success of econometrics as the main contributor of empirical evidence in economics is questionable. Irreconcilable theories, most of the time with opposing results, have been allowed to coexist, rendering the profession incapable of settling economic disputes. To overcome this problem, we adopt in this document the demarcation of an econometric model into two different sources of error: substantive and statistical, paying special attention to the latter: the inability of an econometric model to capture the empirical regularities in the data. We argue that only when the statistical source of error is ruled out, can economic theory confront economic data without compromising the credibility of either one of them. This methodology is henceforth referred to as the Probabilistic Reduction (PR) Approach.

This document is an attempt to illustrate the characteristics of the Probabilistic Reduction Approach and its role in the specification, testing, and respecification of econometric models that warrant statistical adequacy, a measure of the fidelity of the model to capture the empirical regularities in the data. Chapter 2 proposes a methodology that probes for the existence of additional statistical information outside

of the boundaries of a proposed model. The existence of this additional information warns the research about the model's inability to completely describe the stochastic regularities in the data and, supplemented with the PR approach, suggests a line of action for the application of corrective measures. Chapter 3 exemplifies the role of the study of the probabilistic structure of the data in the selection of appropriate econometric models. In particular, the chapter focuses on data with naturally bound ranges: between 0 and 1, as in the case of rates and proportions, or any other kind of proper intervals. Paired with the PR approach, guidelines for the specification of such models and ways to assess their statistical adequacy is provided. Finally, Chapter 4 illustrates the flexibility of the approach in tacking a different kind of data, nonnegative, of particular importance of economics. Similar to the previous chapter, the study of the relevant data, under the PR approach, suggests appropriate specifications and relevant ways to test their usefulness. The modeling strategy is the applied to the determination of the elasticity of gasoline demand.

2 Assessing the Reliability of Inference through Misspecification Testing

2.1 Introduction

2.1.1 Economic Theory and Empirical Evidence

The impressive technical development of econometrics, from its humble beginnings of curve fitting by least-squares between two data series, in the early 20th century, to the estimation and testing of dynamic multi-equation systems, has not been accompanied by an enhancement of the trustworthiness of the resulting empirical evidence. As a result, some critics argued that it has failed as the primary source of reliable empirical evidence in economics, incapable of providing accurate forecasts to settle economic disputes (see Leontief, 1971; Lester, 1983; Eichner, 1983; Johansen, 2007; Spanos, 2006b). It is almost impossible to find theoretical relations that have been abandoned because they were found to be invalid when confronted with real data. Instead, 'irreconcilable theories have been allowed to coexist' (Blaug, 1980) creating a universe of contradictory evidence, so pervasive that even 'so-called' consistent results are to be seen with disdain or simply ignored by theorists' (Lester, 1983).

The accompanying disillusionment with econometrics, caused by the unreliability and imprecision of estimates, somewhat paradoxically, solidified the role of economic theory as the first and only source of information for model specification. Kydland and Prescott (1991) argue that the right specification is not the one that fits the data better but rather 'currently established theory dictates which one is used.' Econometrics' role is re-defined as the mere quantification of economic theories (ever since Johnston, 1963). Under this framework, the theoretical economist feels no obligation to take data into account. Proposed models are driven by an array of motivating factors that include mathematical sophistication and rigor, fecundity, generality, and simplicity rather than the ability of the models to explain or account for empirical

regularities. The theoretical econometrician, on the other hand, devises sophisticated statistical techniques, unconcerned with the appropriateness of these methods to the phenomenon of interest or the precision of the inferences conducted on the results. In the middle of all, 'the applied econometrician stares with esteem at the mathematical dexterity of the other two, but finds himself modeling data from observable economic phenomena which are usually not the result of the ideal circumstances envisaged by the theory, but of an ongoing complete data generation process which shows no respect for *ceteris paribus* clauses, and tramples over individual agent's intentions with no regard for rationality' (Spanos, 2009)[1].

2.1.2 Theory-dominated Modeling

Recently, a number of researchers have made the case that one of the primary reasons for the inability of econometric models to account for empirical regularities is the prevalence of *statistical misspecification* resulting from imposing the theory to the data at the outset (Alston and Chalfant, 1991; D'Agostino et al., 1990; Hoover et al., 2008; Johansen, 2007; Juselius and Franchi, 2007; McGuirk et al., 1993; Spanos: 1995, 2005, 2006a, 2006b). To assess the relevance of such claims, a simple battery of statistical misspecification tests, based on Spanos (2007), was performed on an assortment of estimation examples from three current econometrics textbooks (see Romero, 2009a). The results are presented in Table 2.A. These results provide a snapshot of how widespread the problem of statistical misspecification really is. They also point out that transforming theory-model into a statistical model by adding an error term might not be the best way to statistical model specification.

[1]Leontief (1971) stated: "economic theorists will continue to turn out model after model and theoretical econometricians to devise complicated procedures one after another."

Table 2.A - Empirical studies and their misspecifiction		Source of Misspecification			
Example	Estimation Method	Distrib.	Funct.	Structural	Depend.
Returns to Education[(1)]	Robust$^{(W)}$ OLS	X	X	X	
Engel's Curve[(1)]	OLS	X	X		
Credit Card Expenditure[(2)]	Robust$^{(W)}$ OLS	X	X		
Cobb-Douglas Function[(2)]	LAD		X		
Antidumping[(3)]	OLS	X		X	X
Worker's Compensation[(3)]	Robust$^{(NW)}$ DID	X	X	X	X

(1) Dougherty; (2) Greene; (3) Wooldridge. (W) White's Heteroskedasticity Consistent Estimators (NW) Heteroskedasticity-Autocorrelation Consistent Estimators; DID: Difference in Difference.

It can be argued that, because the entire probabilistic structure of the model is carried by the error, the modeler has no choice but to handle any departures from these assumptions directly in terms of modifying the assumptions and using different estimators, usually invoking large-sample theorems of consistency. Johansen (2007) claims that this deviates the goal of the researcher from the selection of an appropriate statistical model that summarizes the data adequately to a choice of optimal estimators in view of the theory by choosing, from a toolbox, corrective measures to inadequate error assumptions. In a sense, the researcher follows a 'recipe' or 'textbook' approach (TA) to modeling, in which econometrics becomes a showcase for exhibiting theories rather than a tool for testing them. It would appear that untestable and unprovable theoretical models are even more dominant now than they were before econometrics was developed, as predicted by Lester (1983).

2.1.3 Statistical vs. Substantive Information

Attaching an error term to the theory model specification appears to have blurred the separation between inadequate *statistical* models (models that do not accurately capture the statistical regularities in the data) and inadequate *substantive* models

(models that do not adequately represent the phenomenon of interest) in a manner that resembles the Duhemian problem in philosophy of science; see Mayo (1996). This distinction is crucial if theories are to be confronted with the data, as needed by the positivist view of Friedman (1953), where "theory is to be judged by its predictive power for the class of phenomena which is intended to explain ... [since] only factual evidence can show whether [the theory] is 'right' or 'wrong'." Friedman's apparent endorsement of theory driven specifications is at best incomplete and contains the separation between substantive and statistical information only implicitly for no judgment of the predictive power of the theory can be assessed unless the reliability of the tools used in that assessment is established first. The attachment of the error term confounds these two different sources of error, statistical and substantive, and precludes the researcher from making assessments about the phenomenon in question. In this sense, it is not possible to assess whether the theory is wrong (either lacks predictive power or is unable to explain the phenomenon of interest) or if the statistical model is wrong (the probabilistic assumptions of the model are not satisfied), where the inadequacy of the latter completely undermines the credibility of any statistical assessment – however informal – of the former[2].

2.1.4 The Probabilistic Reduction Approach

Under the current modeling paradigm, where the specification is given by the theory and the probabilistic structure by the error, the inability to disentangle the two sources of error has been tolerated under the convention that, in principle, all mod-

[2]Distinguishing between substantive and statistical information and their respective assumptions is not a straightforward endeavor. The problem can be seen in Ireland's DSGE model (2004) where the assumptions invoked are: 1) all structural parameters are constant over time; 2) total factor productivity is driving the system; 3) log-output, consumption, and capital are trend-stationary; 4) labor is stationary; 5) labor augmented technology process follows a linear trend which influences the other variables identically; 6) the observable variables follow a VAR(1) process; 7) the errors are NIID.

els are inherently 'wrong,' since they are simplifications of reality, and that 'slight' departures from model assumptions can only have 'minor' effects on the reliability of inference. Since these effects are minimized asymptotically, whether the data gives empirical support to the theory or not becomes secondary. Even more, there seems to be no need to put in context or provide the appropriate testing framework for claims like 'wrong,' 'slight,' and 'minor.' For the modern economist, both a theorist and an empiricist fond of mathematical rigor and technical skills, the issue of 'whether the data fits the model or not is not resolved by the statistical adequacy of the specification but rather by the degree of confidence that is placed in the economic theory being used' (Kydland and Prescott, 1991). Two questions that suggests themselves are whether these generic robustness claims absolve one from performing misspecification testing and whether the TA approach can provide the appropriate testing framework for econometric theories[3].

The Probabilistic Reduction Approach (PRA) (Spanos 1986, 1999) offers a modeling framework in which the theory can properly confront the data without compromising the credibility of either one of them. This is accomplished by delineating a clear separation between statistical and substantive information and then relating the statistical information to the substantive information through identification. This clear distinction of statistical and substantive information allows the researcher to tie economic theory to the evidence by nesting a stochastic generating process that provides a realization of the phenomenon of interest, the data, and a theoretical

[3]Over the years, several attempts to produce an effective opponent to the TA have been put forward with mixed results (Box-Jenkins (1970, 1976); Sims (1980, 1982), Sargan (1964) & Hendry (2003); Lucas & Sargent (1981); Leamer (1978)). Colander (2009) argues that the TA is currently the preferred methodology not because of its superiority but because it offers the best advancement potential within the existing institutional structure. It is only rational for empirical economists to opt for the theory-first perspective where one only needs to demonstrate technical dexterity in solving, approximating, and calibrating theory-only models. Spanos (2009) argues that the TA preserves economics' status quo, where data has played no role in the specification of models as far back as the 1840's (see Mill, 1844).

model that explains it (Johansen, 2007). Under the approach, the goal is not to find optimal estimators but to develop adequate statistical summaries of the phenomenon of interest (Spanos, 1986).

The PRA gives the data a life of its own and attempts to uncover the statistical mechanism that gave birth to it. All the information contained in the data, the observed variables, is combined into a multivariate stochastic process, $\{\mathbf{Z}_t,\, t\in\mathbb{N}\}$, $\mathbf{Z}_t{:=}(Y_t, \mathbf{X}_t)'$ constrained by several probabilistic assumptions. These assumptions aim to reduce the vector of variables into an estimable model and to orthogonally decompose the information into truth and error, operationalized as:

$$Y_t= \overbrace{E\left(Y_t \mid \mathbf{X}_t{=}\, \mathbf{x}_t\right)}^{\text{Systematic Component}} + \overbrace{u_t}^{\text{Nonsystematic}} \tag{1}$$

The probabilistic assumptions imposed on $\{\mathbf{Z}_t,\, t\in\mathbb{N}{:=}(1,2,...,n,...)\}$ are selected from three broad categories:

$$\text{(D) Distributional, (M) Memory/Dependence, (H) Heterogeneity} \tag{2}$$

thus partitioning the space of all possible statistical models into a family of operational ones (Figure 2.1). The first two conditional moments establish the specification of both the regression line and the skedastic function.

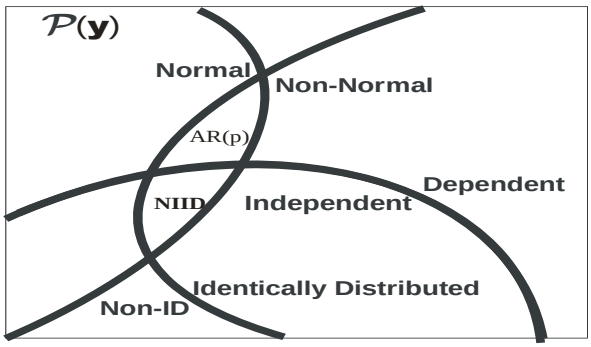


Fig. 2.1 - Specification by partitioning

The resulting specification is supplemented with a set of testable probabilistic assumptions where the observable data is used to qualify the models as statistically

adequate or statistically misspecified[4]. As it turns out, this recasting of specification selection allows the researcher to assess the precision and reliability of inference and provides the testing framework to qualify claims such as 'wrong,' 'slight,' and 'minor.'

2.2 Data and Methodology

2.2.1 Simulated Data

To compare and contrast the TA and the PRA perspective, a series of Monte Carlo experiments were conducted with different sample sizes and probabilistic structures that a typical researcher would find in practice. These experiments resemble usual misspecification problems in econometrics such as functional form, autocorrelation, and heteroskedasticity. Since the aim is to elucidate the kind of problems an empirical modeler will face while attempting to model observational data, the data will be simulated from the probabilistic assumptions of the observable variables rather than by simulating error terms. Unless otherwise indicated, the experiments were conducted using two sample sizes, n=50 and n=100, and N=10000 replications using Matlab 9.

It will be assumed that the modeler obtains some a priori information from the economic theory regarding the relationship between three variables: a response variable Y and a set of two predictors X_1 and X_2. The proposed relationship provided by the theorist is:

$$Y=\beta_0 + \beta_1 X_1 + \beta_2 X_2. \tag{3}$$

It is also assumed that the modeler either receives or collects data germane to (Y, X_1, X_2) and attempts to either corroborate of reject the assumed relationship.

2.2.2 Misspecification (M-S) Testing

A statistically adequate model is a necessary condition for the sound appraisal of a relevant structural model because without it the inference procedures used in such

[4]In contrast with the TA, the statistical properties of the errors are derived rather than assumed.

an appraisal will be unreliable; the actual error probabilities will be different from the actual ones. As such, a very important component of the modeling process should be to determine whether the assumptions of the model are valid vis-a-vis the data. M-S testing probes outside the boundaries of pre-specified models by testing $H_0 : f_0(z) \in \mathcal{M}$ vs. $\overline{H}_0 : f_0(z) \in (\mathcal{P} - \mathcal{M})$, where $\mathcal{P}$ denotes the set of all possible statistical models[5]. Detection of departures from the null in the direction of $\mathcal{P}_1 \subset (\mathcal{P} - \mathcal{M})$ can be considered sufficient to deduce that the null is false but not to deduce that $\mathcal{P}_1$ is true.

Despite its importance for the reliability of inference, McGuirk et al., (1993) assert that M-S testing is not widely appreciated as a crucial aspect of statistical modeling and inference. It can be argued that the lack of a general misspecification methodology and the fact that in practice it may be very difficult to identity the sources of misspecification (Alston and Chalfant, 1991) have dissuaded researchers from testing models. But to the fact of how much to probe[6], an additional argument exists against misspecification testing: to what extent M-S testing involves illegitimate use, or double-use, of data. Spanos (2007) contends that this methodological objection does not arise when the **studentized** estimation residuals are used for M-S testing and makes the case of the use of ancillary regressions to probe for misspecification.

[5]This form of testing differs from Neyman-Pearson (NP) testing in the sense that NP assumes that the pre-specified statistical model class $\mathcal{M}$ includes the true model, and probes within the boundaries of this models using the hypothesis $H_0 : f_0(z) \epsilon \mathcal{M}_0$ vs. $H_1 : f_0(z) \epsilon \mathcal{M}_1$, where $\mathcal{M}_0$ and $\mathcal{M}_1$ form a partition of $\mathcal{M}$.

[6]McGuirk et al. (1993) suggest that, at a minimum, the validity of all testable assumptions should be examined. Additionally, they propose that to improve the probativeness of misspecification testing, all versions of both individual and joint tests should be included in a test regime. The full-fledged misspecification battery, based on Spanos and McGuirk (2001), and used throughout this document, is presented in Appendix B.

If indeed the proposed specification has been able to capture all the systematic information in the data through $g(\mathbf{X}_t)$, then any other function of the conditioning set $h(\mathbf{X}_t)$ will cause the following condition to hold[7]:

$$E([y_t - g(\mathbf{X}_t)]\, h(\mathbf{X}_t)) = 0,\ t \in \mathbb{N}. \tag{4}$$

The condition is referred to as the orthogonality expectation theorem. Using the *studentized* residuals, ancillary regressions of the form:

$$\widehat{v}_t = [y_t - g(\mathbf{X}_t)], \quad t = 1, 2, ..., n, \tag{5}$$

where $\widehat{v}_t = \frac{\sqrt{n}(y_t - \widehat{y}_t)}{\widehat{\sigma}} \sim St(n-1)$, $\sigma = \sqrt{Var(y_t | \mathbf{X}_t = \mathbf{x}_t)}$, can be estimated and used to assess deviations from the statistical model assumptions.

In principle, this transformation also solves the practical issue since it is possible to capture departures from the model assumptions and to probe outside the limits of the model using transformations (functional, structural, dependence) of $\mathbf{X}_t$ and Y_t.

To assess the reliability of residual-based M-S testing and the precision of estimation under the TA and the PRA, a set of ancillary regressions will be used from $\widehat{y}_t = E(y_t | \mathbf{X}_t = \mathbf{x}_t)$, $\widehat{\sigma}^2 = Var(y_t | \mathbf{X}_t = \mathbf{x}_t)$, and $\widehat{u}_t = y_t - \widehat{y}_t$, to test for statistical model departures in the first four conditional moments, as follows:

$$\left.\begin{array}{lcl}
E\left(\frac{u_t}{\sigma}\right)=0 & \Leftrightarrow & \left(\frac{\widehat{u}_t}{\widehat{\sigma}}\right)=\gamma_{10}+\boldsymbol{\gamma}_{11}'\mathbf{X}_t+\boldsymbol{\gamma}_{12}'\boldsymbol{\xi}_t+\boldsymbol{\gamma}_{13}'\mathbf{X}_t^2+\boldsymbol{\gamma}_{14}'\mathbf{X}_{t-1}+\varepsilon_{1t}, \\
E\left(\frac{u_t^2}{\sigma^2}\right)=1 & \Leftrightarrow & \left(\frac{\widehat{u}_t}{\widehat{\sigma}}\right)^2=\gamma_{20}+\boldsymbol{\gamma}_{21}'\mathbf{X}_t+\boldsymbol{\gamma}_{22}'\boldsymbol{\xi}_t+\boldsymbol{\gamma}_{23}'\mathbf{X}_t^2+\boldsymbol{\gamma}_{24}'\mathbf{X}_{t-1}+\varepsilon_{2t}, \\
E\left(\frac{u_t^3}{\sigma^3}\right)=0 & \Leftrightarrow & \left(\frac{\widehat{u}_t}{\widehat{\sigma}}\right)^3=\gamma_{30}+\boldsymbol{\gamma}_{31}'\mathbf{X}_t+\boldsymbol{\gamma}_{32}'\boldsymbol{\xi}_t+\boldsymbol{\gamma}_{33}'\mathbf{X}_t^2+\boldsymbol{\gamma}_{34}'\mathbf{X}_{t-1}+\varepsilon_{3t}, \\
E\left(\frac{u_t^4}{\sigma^4}\right)=3 & \Leftrightarrow & \left(\frac{\widehat{u}_t}{\widehat{\sigma}}\right)^4=\gamma_{40}+\boldsymbol{\gamma}_{41}'\mathbf{X}_t+\boldsymbol{\gamma}_{42}'\boldsymbol{\xi}_t+\boldsymbol{\gamma}_{43}'\mathbf{X}_t^2+\boldsymbol{\gamma}_{44}'\mathbf{X}_{t-1}+\varepsilon_{4t},
\end{array}\right\} t=1,2,...,n \tag{6}$$

where $\mathbf{X}_t$ is the vector of regressors of the original specification, $\boldsymbol{\xi}_t$ is a vector of trends that capture structural change misspecification, $\mathbf{X}_t^2$ is a vector of monotonic transformations of $\mathbf{X}_t$ that allows the conditional standardized moment to have additional sources of nonlinearities, and $\mathbf{X}_{t-1}$ is a vector or lagged values of $\mathbf{X}_t$ and Y_t that allows for temporal of spatial dependence.

[7]For the regression function, $g(\mathbf{X}_t) = E(y_t | \sigma(\mathbf{X}_t))$, where $\sigma(\mathbf{X}_t)$ denotes the σ-field generated by $\mathbf{X}_t$.

To assess the probativeness of the use of the standardized residuals, each equation in the system is tested separately with an F-type test. The null hypotheses are of the form $\boldsymbol{\gamma}'_{.1}=\boldsymbol{\gamma}'_{.2}=\boldsymbol{\gamma}'_{.3}=\boldsymbol{\gamma}'_{.4}=\mathbf{0}$. Distribution assumptions can also be tested for some distributions, like the normal, where the individual tests would incorporate $\boldsymbol{\gamma}'_{.0}=[0\ 1\ 0\ 3]'$. A variation of these tests using the mean-corrected standardized residuals is also evaluated.

The degree of probativeness can be increased by separating the individual components of the equations into the different sources of misspecification: Orthogonality, Structural, Functional, and Dependence. This separation can potentially increase the ability of the researcher to isolate the source of the misspecification.

The previous system of equations can also be tested simultaneously for departures from the model assumptions and normality, that, is, $\boldsymbol{\gamma}_{(.0)}=\mathbf{0}$, and $\boldsymbol{\gamma}'_{.0}=[0\ 1\ 0\ 3]'$ using both the raw standardized residuals and the corrected standardized residuals. This procedure effectively creates a five-dimensional joint misspecification test. The joint test is conducted using a Multivariate Normal Linear Regression Framework (see Appendix 2.C).

The results of applying this M-S testing methodology will help assess the reliability and precision of inference under both econometric modeling approaches, the TA and the PRA.

2.3 Empirical Results

To perform estimation and to conduct statistical inference, it is necessary that the theoretical model (3) be embedded into a statistical one. Under the TA, the process consists on endowing the theory specification with a stochastic term at the end of the equation. This term will hold all the relevant statistical information of the relationship between Y and X_1 and X_2. The probabilistic structure of the error also determines the sampling properties of inference procedures conducted in the quantified relation.

The statistical model from the TA is:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + e_t,\ t\in\mathbb{N}. \tag{7}$$

Under the PRA, the process consists on embedding the relevant data (y_t, x_{1t}, x_{2t}) into a vector stochastic process $\{\mathbf{Z}_t{:=}(y_t, x_{1t}, x_{2t})\,,\ t\in\mathbb{N}\}$, whose probabilistic structure determines the statistical relationships between Y and X_1 and X_2 (Billingsley, 1986) and whose probabilistic reduction determines the specification of the model.

2.3.1 Experiment 1 - Normal/Linear Regression Model

For this particular case, the probabilistic reduction of $\mathbf{Z}_t$ takes the form,

$$\begin{aligned} f(\mathbf{Z}_1, ...\mathbf{Z}_n;\boldsymbol{\phi}) &\overset{\mathsf{I}}{=} \prod_{t=1}^{n} f_t(\mathbf{Z}_t;\boldsymbol{\varphi}_t) \overset{\mathsf{IID}}{=} \prod_{t=1}^{n} f(\mathbf{Z}_t;\boldsymbol{\varphi}) = \prod_{t=1}^{n} f(Y_t, X_{1t}, X_{2t};\boldsymbol{\varphi}) = \\ &= \prod_{t=1}^{n} f(Y_t|X_{1t}, X_{2t};\boldsymbol{\varphi}_1)\, f(X_{1t}, X_{2t};\boldsymbol{\varphi}_2) \overset{\mathsf{NIID}}{=} \prod_{t=1}^{n} f(Y_t|X_{1t}, X_{2t};\boldsymbol{\varphi}_1) \end{aligned} \tag{8}$$

where it is possible to ignore the marginal distribution $f(X_{1t}, X_{2t};\varphi_2)$ by imposing normality. The reduction assumptions then imply NIID. The model assumptions are given in Table 2.B.

Table 2.B: The Normal Linear Regression Model (NLR)

	$y_t{=}\beta_0 + \boldsymbol{\beta}_1^{\intercal}\mathbf{x}_t + u_t$
[1] Normality	$(y_t\|\mathbf{X}_t{=}\,\mathbf{x}_t) \sim N(.,.)$
[2] Linearity	$E(y_t\|\mathbf{X}_t{=}\,\mathbf{x}_t){=}\beta_0 + \boldsymbol{\beta}_1^{\intercal}\mathbf{x}_t$
[3] Homoskedasticity	$Var(y_t\|\mathbf{X}_t{=}\,\mathbf{x}_t){=}\sigma_0^2$
[4] Independence	$\{(y_t\|\mathbf{X}_t{=}\,\mathbf{x}_t)\,,\ t\in\mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi}_1 {:=}(\beta_0,\ \boldsymbol{\beta}_1^{\intercal},\ \sigma_0^2)$ do not change with t
where $\beta_0{=}\mu_1 - \boldsymbol{\beta}_1^{\intercal}\boldsymbol{\mu}_2$, $\boldsymbol{\beta}_1^{\intercal}{=}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$, $\sigma_0^2{=}\sigma_{11} - \boldsymbol{\sigma}_{21}^{\intercal}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$	

The data $\{(y_t, x_{1t}, x_{2t})\,,\ t{=}1, 2, \ldots n\}$ is generated via a three-variate, identical, and independently distributed normal process[8],

$$\begin{pmatrix} y_t \\ x_{1t} \\ x_{2t} \end{pmatrix} \sim \mathsf{N}\left[\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1.2 & .7 & -.4 \\ .7 & 1 & .2 \\ -.4 & .2 & 1 \end{pmatrix}\right] \tag{9}$$

The statistical information contained in (9) is used to derive the normal/linear regression model (NLR), embedding statistically (3). The resulting true statistical model is then,

$$y_t = 1.0625 + .8125x_{1t} - .5625x_{2t} + u_t, \tag{10}$$

where $\sigma_0^2{=}\sigma^2_{y_t|\mathbf{x}_t}{=}.4062$ and $\mathcal{R}^2{=}1 - \frac{\sigma_0^2}{Var(Y_t)}{=}.6614$, for $t{\in}\mathbb{N}$.

For this case, both the TA and the PRA specification and estimation should coincide as well as the statistical properties of the errors. Estimation with OLS yields the results presented in Table 2.1A (see Appendix 2.A). The results from this table will be used as the benchmark for subsequent experiments. The statistics reported include t-statistics of the estimated parameters against the true parameters and a joint F-test to establish whether all the estimated coefficients are equal to the true coefficients simultaneously. Both sets of statistics are used to asses the precision of the estimation via the percentage of rejections at a pre-specified significance level. For the estimation to be statistically *close* to the true values, the percentage of rejections has to be of equal magnitude to the nominal significance level (α) for all estimated parameters, 5 percent throughout this analysis.

Table 2.1A shows that, in general, the nominal and the actual error probabilities are in check. The t-tests and the F-test do not present significant deviations in the percentage of rejections. To asses the reliability of inferences, the table also includes a typical set of misspecification tests usually reported automatically in TA modeling; tests that would include the Durbin-Watson test for autocorrelation, a test

[8]Simulating from the joint distribution increases the control over the statistical properties of the model than simulating from the error (see Romero, 2009b).

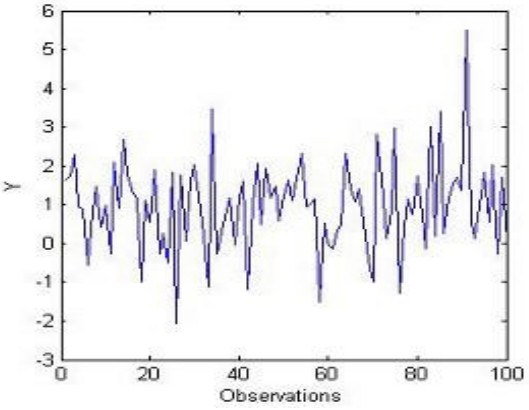
for heteroskedasticity (either the Breusch-Pagan test or the White test), and a test for Normality (either the Shapiro–Wilk test or the Jarque-Bera test). This default misspecification battery does not seem to indicate any major departures from the model assumptions. Actual and nominal error probabilities are in check (notice the relatively lower power of the Durbin-Watson test). Based on all this, inference can be reliably conducted.

To arrive at a specification, the PRA's first step is to examine t-plots and scatterplots of $\{\mathbf{Z}_t := (y_t, x_{1t}, x_{2t})\,,\ t=1,2,\ldots n\}$ in an attempt to assess the marginal and joint distributions as well as the sampling properties of the data. The goal is to establish whether the model assumptions and the reduction assumptions imposed on $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$, i.e., NIID, hold.

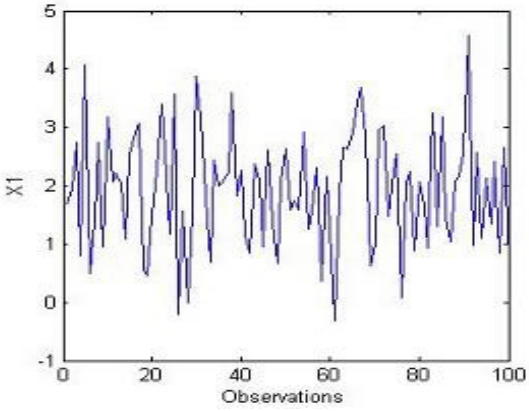
From the t-plots in Panel 2.1, it is possible to imply that both the mean and the variance of y_t and x_{it} appear to be constant over the index t, that is, the processes exhibit spatial of temporal independence. From the scatters in the same panel, it also seems that the marginal distributions appear to be bell-shaped symmetric around a constant mean. These same figures help asses the elliptically-shaped scatter between the three pairs of variables. A positive principal axis for the case of y_t and x_{1t}, and a negative principal axis for the case of y_t and x_{2t}. The ensuing model specification is that the variables are NIID and that the NLR model is in order. As suspected above, the estimated regression model using the Monte Carlo simulated data then coincides with the results obtain in Table 2.1A with the TA.

For the PR modeler, a battery of misspecification tests needs to encompass a set of individual and joint tests of all testable assumptions. Table 2.1B presents a full-fledged misspecification battery that combines individual as well as joint misspecification tests (see Appendix 2.B). Similar to the TA case, no major departures from the model assumptions can be detected from the results. The nominal and actual error probabilities are of equal magnitudes. With these results in hand, it is then possible to perform statistical testing and to conduct reliable inferences.

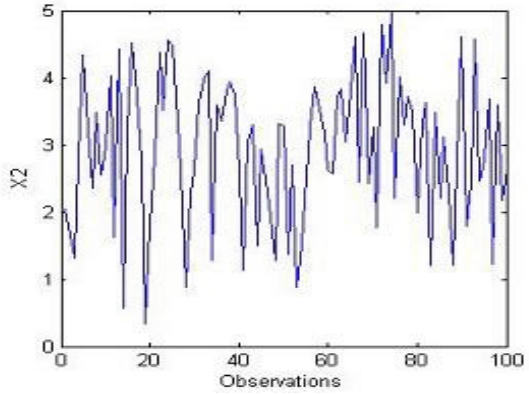
Panel 2.1: t-plots and scatter plots of (y_t, x_{1t}, x_{2t}) from Experiment 1



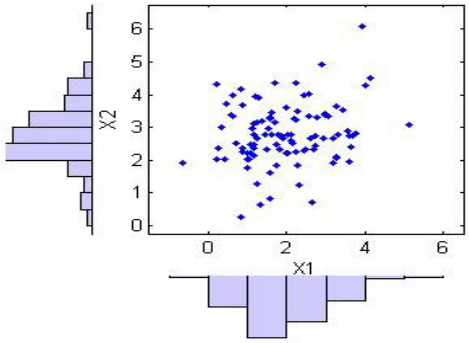
t-plot of y_t



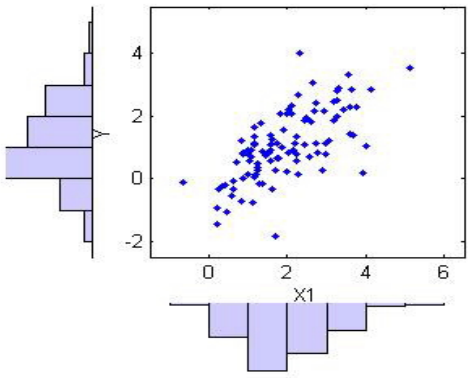
t-plot of x_{1t}



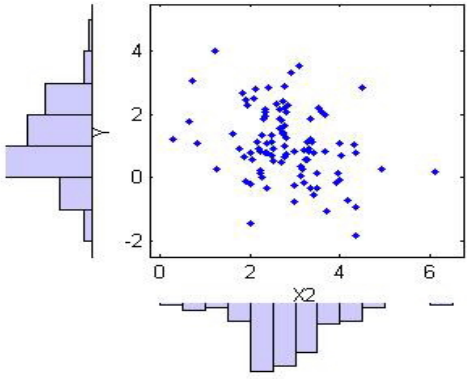
t-plot of x_{2t}



Scatter-plot of (x_{1t}, x_{2t})



Scatter-plot of (y_t, x_{1t})



Scatter-plot of (y_t, x_{2t})

2.3.2 Experiment 2 - Heterogeneous NLR Model

For the second experiment, the identical distribution assumption is allowed to fail. For the three variables, the marginal means have become heterogeneous and linearly related to the index, representing spatial or temporal dependence. The variance-covariance matrix is still stationary. The reduction takes the form:

$$f\left(\mathbf{Z}_1,...,\mathbf{Z}_n;\boldsymbol{\phi}\right)\overset{\mathsf{I}}{=}\prod_{t=1}^{n}f_t\left(\mathbf{Z}_t;\boldsymbol{\varphi}_t\right)\overset{\mathsf{I}}{=}\prod_{t=1}^{n}f\left(Y_t|X_{1t},X_{2t};\boldsymbol{\varphi}_{1t}\right)f\left(X_{1t},X_{2t};\boldsymbol{\varphi}_{2t}\right)\overset{\mathsf{NI}}{=}\prod_{t=1}^{n}f\left(Y_t|X_{1t},X_{2t};\boldsymbol{\varphi}_{1t}\right)$$

where, unless it is determined how the parameter set changes with t, no additional reductions are possible and no estimation could be performed. Letting $\mu_1^*\left(t\right)=\mu_1+\delta_1 t$, $\boldsymbol{\mu}_2^*\left(t\right)=\boldsymbol{\mu}_2+\boldsymbol{\delta}_2 t$ and $\boldsymbol{\Sigma}\left(t\right)=\boldsymbol{\Sigma}$, the NLR with a trend model is (Table 2.C).

Table 2.C: The Normal Linear Regression Model with a Trend (NLR-trend)

$$y_t=\beta_0+\delta t+\boldsymbol{\beta}_1^{\intercal}\mathbf{x}_t+u_t$$

[1] Normality	$\left(y_t\right.\vert\ \mathbf{X}_t=\mathbf{x}_t)\sim\mathsf{N}(.,.)$
[2] Linearity	$E\left(y_t\right.\vert\ \mathbf{X}_t=\mathbf{x}_t)=\beta_0+\delta t+\boldsymbol{\beta}_1^{\intercal}\mathbf{x}_t$
[3] Homoskedasticity	$Var\left(y_t\right.\vert\ \mathbf{X}_t=\mathbf{x}_t)=\sigma_0^2$
[4] Independence	$\{\left(y_t\right.\vert\ \mathbf{X}_t=\mathbf{x}_t),\ t\in\mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi}_1:=(\beta_0,\ \delta,\ \boldsymbol{\beta}_1^{\intercal},\ \sigma_0^2)$ do not change with t

where $\beta_0=\mu_1-\boldsymbol{\beta}_1^{\intercal}\boldsymbol{\mu}_2$, $\delta=\delta_1-\boldsymbol{\beta}_1^{\intercal}\boldsymbol{\delta}_2$, $\boldsymbol{\beta}_1^{\intercal}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$, $\sigma_0^2=\sigma_{11}-\boldsymbol{\sigma}_{21}^{\intercal}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$

The data $\{\left(y_t,x_{1t},x_{2t}\right),\ t=1,2,\ldots n\}$ is generated with the following three-variate, non-identical, and independently distributed normal process:

$$\begin{pmatrix}y_t\\x_{1t}\\x_{2t}\end{pmatrix}\sim\mathsf{N}\left[\begin{pmatrix}1+0.2t\\2+0.3t\\3+0.4t\end{pmatrix},\ \begin{pmatrix}1.2&0.7&-0.4\\0.7&1.0&0.2\\-0.4&0.2&1.0\end{pmatrix}\right]\tag{11}$$

With the inclusion of mean heterogeneity, the true regression model becomes,

$$y_t=1.0623+0.1812t+0.8125x_{1t}-0.5625x_{2t}+u_t,\tag{12}$$

where $\sigma_0^2=0.4062$ and $\mathbb{R}^2=0.6614$, for $t\in\mathbb{N}$. Notice that the parameterization of (11) allows the marginal responses of X_1 and X_2 to coincide with those of (10).

The question at hand is whether this is a 'slight' model departure that would have 'minor' effects on estimation and inference. From the TA approach, there would be no reason to add a trend since none is justified by (3). The results of estimating (3), presented in Table 2.2A, would be typical of the results obtained through the TA. From the automatic diagnostic measures (Autocorrelation, Heteroskedasticity, and Normality), it is clear that a researcher would feel confident about the reliability of the estimation since no major model departures can be detected.

Undoubtedly, this experiment, though trivial in nature, exemplifies the danger of taking at face value the results from misspecification tests without paying careful attention to the data. Even if the research were to pry at the 'residuals,' where the model assumptions are corroborated in the TA approach, she would find no evidence of the existence of a linear trend for none is perceivable with the naked eye. Figure 2.2 and Figure 2.3 illustrate this point. The figures represent the t-plots of two typical realizations of the residuals of estimating **(7)** instead of **(12)** at the two sample sizes. Note that, at plain sight, it seems impossible to detect the existence of any trends. This might lead the researcher to believe that a slight departure brings no major consequences to the inference.

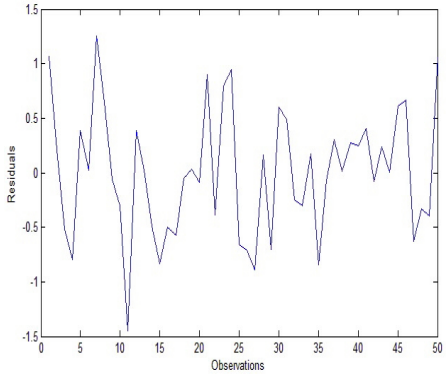


Fig. 2.2: t-plot of $\widehat{e}_t$ for T=50

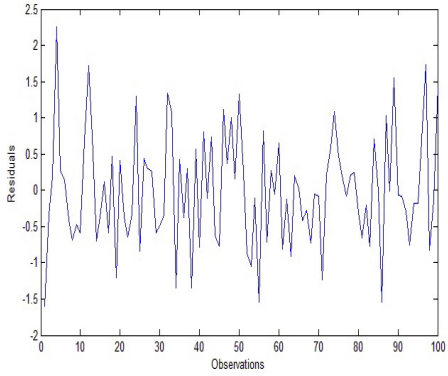


Fig. 2.3: t-plot of $\widehat{e}_t$ for T=100

Exacerbating the situation, the estimated coefficients are all individually and simultaneously different from zero, as indicated by the t and the F-tests, also in Table 2.2A; and R^2 is relatively high as well (although significantly off from the true value). As it stands, the model proposed and estimated with the TA seems reasonably 'good': The response variable y_t is statistically related to x_1 and x_2. How 'good' the results can be confirmed by looking at a plot of the fitted values versus the actual data, Figure 2.3.

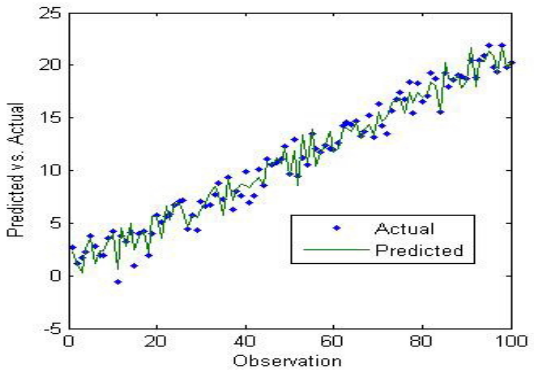


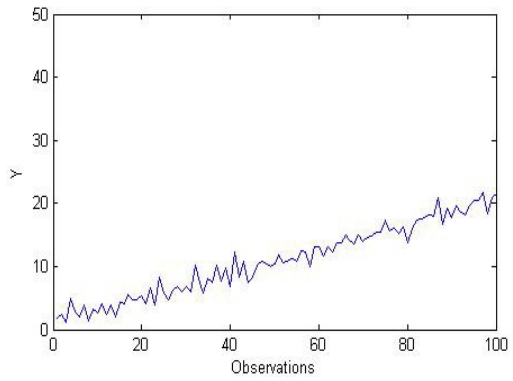
Fig. 2.3: t-plot of $(y_t, \widehat{y}_t)$

Albeit the fit, tests of whether the estimated parameters differ from the true parameters elicits the extent of the misleading results and the danger in which a researcher will be if conducting inference with this estimated model. For n=50, the modeler would accept a biased relationship between X_1 and Y as true almost 48 percent of the time instead of the nominal 5 percent of the time. This probability would increase to almost 81 percent of the time at n=100. Additionally, for the relationship between X_2 and Y, the respective actual error probabilities will be off almost 97 percent of the time when n=50 increasing to almost 99 percent of the time when n=100, instead of the nominal 5 percent that the researcher believes holds true. A similar conclusion will be reached when looking at the F-test of the estimated parameters versus the true parameters. This implies that the modeler

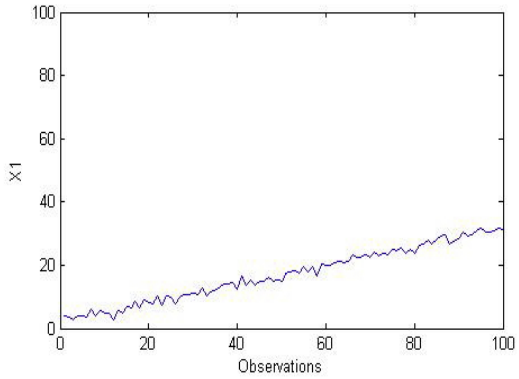
will feel confident that the estimated parameters are indeed different from the true parameters despite the knowledge to the contrary.

Applying the full-fledged battery of misspecification tests reveals the existence of a trend in the conditional mean (Table 2.2B). Whether the trend is the result of individual trends in the explanatory variables, in the dependent variable, or in all of the involved variables simultaneously becomes irrelevant since in either case the trend will be carried onto the conditioned variable and its conditional distribution. This is particularly evident in both the individual test for trends in mean and the joint tests including trends in the conditional mean. As argued by Spanos and McGuirk (2001), the use of joint and individual tests of heterogeneity facilitates the detection of departures of parameter homogeneity from the sampling assumption. It seems then that the rejection of at least one the previous misspecification tests invalidates any reliable inference drawn from the model above.

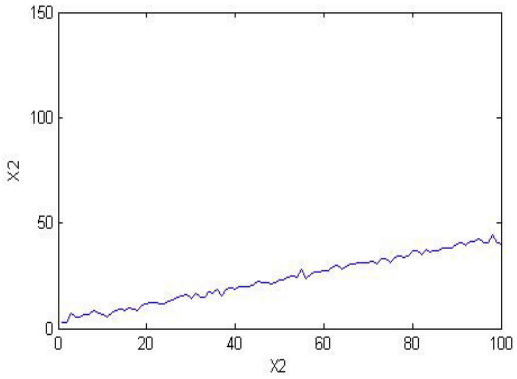
Panel 2.2: t-plots of (y_t, x_{1t}, x_{2t}) from Experiment 2



t-plot of y_t



t-plot of x_{1t}



t-plot of x_{2t}

In contrast, under the PRA, the researcher attempts to produce an initial specification warranted by the probabilistic structure of the data. By looking at the t-plots of the process (Panel 2.2), it is clear that the unconditional mean of all the variables is trending, changing with the index. It is also apparent that the processes seem to maintain the same variance homogeneity throughout the index. Unfortunately, the marginal distributions of the variables cannot be assessed in this stage unless the data is detrended and dememorized (see Spanos, 1999). Only then, it would be possible

to assert that the joint (marginal) distribution(s) are indeed normally distributed. Normality and the heterogeneity in the conditional mean observed in the data would suggest the use of a NLR model with a trend; a first approximation to the statistical modeling of the process.

Even if the detection of the trends is not possible through the t-plots, either because the trends are too subtle or because the modeler decided not look at them, the existence of the trends and its potential solution would have been detected by the full-fledged battery of misspecification tests. In particular, the presence of the trend in the conditional mean would be detected at least 78 percent of the time when n=50 and almost 99 percent of the time when n=100.

The specification of a normal linear model with a linear trend, warranted by the PRA, allows the researcher to capture the entirety of the trend heterogeneity present in the three variables. The results, shown in Tables 2-2C and 2-2D, indicate that this is indeed the case and that no major departures from the augmented model assumptions exist. In this case, the statistical adequacy of the econometric model is warranted and the actual and nominal error probabilities are in check. Thus, it seems that one has to be suspicious of the reliability of the default misspecification battery. It might lead to imprecise and unreliable but seemingly statistically adequate inferences.

2.3.3 Experiment 3 - Dynamic Normal/Linear Regression Model

For this experiment, a first-order Markov dependent variance-covariance matrix is proposed while maintaining joint normality. The probabilistic reduction yields:

$$f\left(\mathbf{Z}_1,...\mathbf{Z}_n;\boldsymbol{\phi}\right)\overset{\mathsf{M\&S}}{=}\prod_{t=1}^{n}f\left(\mathbf{Z}_t|\mathbf{Z}_{t-1}^0;\boldsymbol{\varphi}\right)=\prod_{t=1}^{n}f\left(Y_t|\mathbf{Z}_{t-1}^0,\mathbf{X}_t;\boldsymbol{\varphi}_1\right)\cdot f\left(\mathbf{X}|\mathbf{Z}_{t-1}^0;\boldsymbol{\varphi}_2\right)\overset{\mathsf{N}}{=}\prod_{t=1}^{n}f\left(Y_t|\mathbf{Z}_{t-1}^0,\mathbf{X}_t;\boldsymbol{\varphi}_1\right),$$

where $\mathbf{Z}_{t-1}^0$:=$(\mathbf{Z}_{t-1},\mathbf{Z}_{t-2},...,\mathbf{Z}_1)$ denotes the past history of $\mathbf{Z}_t$:=$(y_t,\mathbf{X}_t')'$. The reduction produces the Dynamic Linear Regression Model specification (Table 2.D).

Table 2.D: The Dynamic Linear Regression Model [DLR(m)]	
$y_t{=}\alpha_0+\boldsymbol{\beta}_0^{\intercal}\mathbf{x}_t+\sum_{k=1}^{m}\left[\alpha_k y_{t-k}+\boldsymbol{\beta}_k^{\intercal}\mathbf{x}_{t-k}\right]+u_t,\ t{\in}\mathbb{N}.$	
[1] Normality	$\left\{\left(y_t\vert\mathbf{X}_t{=}\,\mathbf{x}_t,\sigma\left(\mathbf{Z}_{t-1}^0\right);\boldsymbol{\varphi}_1\right),\ \mathbf{Z}_{t-1}^0{=}\left(\mathbf{Z}_{t-1},\mathbf{Z}_{t-2},...,\mathbf{Z}_1\right)\right\}\sim\mathsf{N}(.,.)$
[2] Linearity	$E\left(y_t\vert\mathbf{X}_t{=}\,\mathbf{x}_t,\sigma\left(\mathbf{Z}_{t-1}^0\right);\boldsymbol{\varphi}_1\right){=}\alpha_0{+}\boldsymbol{\beta}_0^{\intercal}\mathbf{x}_t{+}\sum_{k=1}^{m}\left[\alpha_k y_{t-k}+\boldsymbol{\beta}_k^{\intercal}\mathbf{x}_{t-k}\right]{+}u_t$
[3] Homoskedasticity	$Var\left(y_t\vert\mathbf{X}_t{=}\,\mathbf{x}_t,\sigma\left(\mathbf{Z}_{t-1}^0\right);\boldsymbol{\varphi}_1\right){=}\sigma_0^2$ is free of $\left(\mathbf{x}_t,\mathbf{Z}_{t-1}^0\right)$
[4] Independence	$\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$ is a Markov(m) process
[5] t-homogeneity	$\boldsymbol{\varphi}_1{:=}(\alpha_k,\ \boldsymbol{\beta}_k,\ k{=}0,1,\ldots,m,\ \sigma_0^2)$ do not change with t
where $\boldsymbol{\alpha}{=}\mu_1-\boldsymbol{\beta}^{\intercal}\boldsymbol{\mu}_2$, $\boldsymbol{\beta}^{\intercal}{=}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$, $\sigma_0^2{=}\sigma_{11}-\boldsymbol{\sigma}_{21}^{\intercal}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$	

The data, generated from the following joint distribution:

$$\begin{pmatrix} y_t \\ x_{1t} \\ x_{2t} \\ y_{t-1} \\ x_{1t-1} \\ x_{2t-1} \end{pmatrix} \sim \mathsf{N}\left[\begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1.20 & 0.70 & -0.40 & 0.80 & 0.50 & -0.38 \\ 0.70 & 1.00 & 0.20 & 0.50 & 0.70 & 0.10 \\ -0.40 & 0.20 & 1.00 & -0.38 & 0.10 & 0.75 \\ 0.80 & 0.50 & -0.38 & 1.20 & 0.70 & -0.40 \\ 0.50 & 0.70 & 0.10 & 0.70 & 1.00 & 0.20 \\ -0.38 & 0.10 & 0.75 & -0.40 & 0.20 & 1.00 \end{pmatrix}\right] \tag{13}$$

gives rise to the following Dynamic Linear Regression Model (DLR(1)):

$$y_t{=}.654+.8026x_{1t}{-}.4531x_{2t}{+}.4239y_{t-1}{-}.3365x_{1t-1}{+}.1164x_{2t-1}+u_t,\ t{\in}\mathbb{N}, \tag{14}$$

where $\sigma_0^2{=}.3303$, $y_0\sim\mathsf{N}(1,1.2)$, $x_{10}\sim\mathsf{N}(2,1)$, and $x_{20}\sim\mathsf{N}(3,1)$.

Once more, the question at hand is whether the departure in the sampling process significantly affects the precision and reliability of inferences in (3) when estimation is conducted under the TA. The researcher, similarly to the previous example, would need not incorporate a dynamic structure in the specification since none is suggested by the theory. The results are presented in Table 2.3A. From the results it is clear that the test for temporal dependence will detect the presence of the autocorrelated process often enough. Additionally, notice that the power of the Durbin-Watson test increases with the sample size, as should be expected. Similarly, the full-fledged misspecification battery, Table 2.3B, elucidates that the existence of temporal dependence will be the first concern for the statistical adequacy of this model. Although there are some 'yellow flags', in particular, Normality, Homoskedasticity, and Linearity, it is evident that the 'red flags' are raised by the temporal dependence tests: Durbin-Watson, $AC(1)$, $AC(2)$, $\widehat{u}_{t-1}$ in mean, and $(y_{t-1}, \mathbf{x}_{t-1})$ in mean.

Under the TA, the natural course of action is to correct the problem by adopting an alternative to the NLR model that captures the autocorrelation present in the errors. This is accomplished by modifying their sampling properties and delineating an autoregressive structure. The caveat of this procedure is the imposition of common factor restrictions, which would have to be tested whether the researcher is aware of them or not.

Table 2.3C shows the result of estimating an AR(1) model using a 2-step Cochrane-Orcutt correction[9]. Very interesting results arise from this model. At face value, it seems that the modification does a great job capturing the temporal dependence of the data. The estimators seem to be statistically different from zero and the default misspecification tests do not seem to indicate any major departures from the model assumptions. Even more, the t-tests of the estimated parameters versus the true coefficients seem to indicate that, with the data at hand, the estimators are in the 'ball-park.' Only with the joint significance test it is possible to discern the

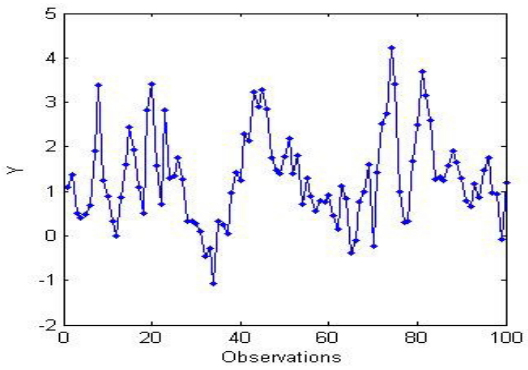
[9]Similar results were obtained when estimating the model via Maximum Likelihood.

discrepancy between the true and the estimated coefficients. All in all, it seems then that artificially imposing the common factor restrictions would do a good job modeling the data even when they are invalid since the default misspecification battery does not explicitly test for them.

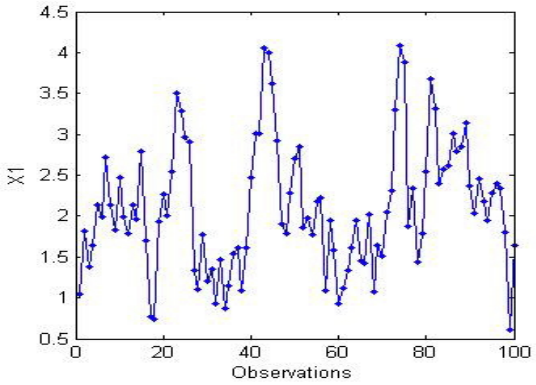
The presence of the common factor restrictions is revealed only through the supplemented battery of misspecification tests, presented in Table 2.3D. Of course, the power of the common factor test is a function of the magnitude of the true coefficients and/or the sample size. For this particular example, it is clear that although the actual error probability is greater than the nominal error probability at n=50, this does not seem to be more than a 'yellow flag.' At the larger samples, n=100, and above (see Table 2.3G), it is clear that the 'yellow flag' becomes a 'red flag.'

Under the PRA, the modeler would attempt to derive the specification from the assessment of the t-plots and scatter-plots. At first sight, both explanatory variables show evidence of temporal dependence (Panel 2.3). The degree of dependence for Y_t becomes more complicated to identify due to the inter-temporal relationship with the X's but it seems to reveal that first degree temporal dependence exists between Y_t and Y_{t-1}. However, no apparent time heterogeneity is present in the data; the processes seem to fluctuate within a homogeneous band throughout the index. After subtracting the temporal dependence from each series, the resulting t-plots resemble those of the NLR model. The PR modeler's first specification is a DLR model.

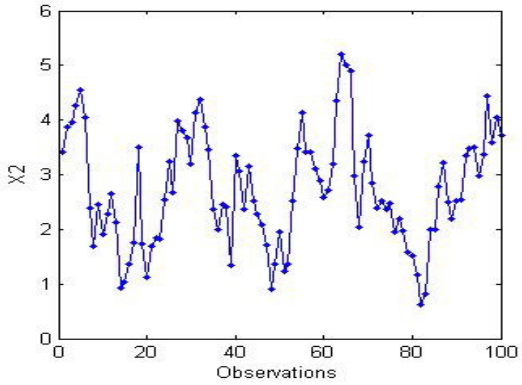
Panel 2.3: t-plots of (y_t, x_{1t}, x_{2t}) from Experiment 3



t-plot of y_t



t-plot of x_{1t}



t-plot of x_{2t}

Note that, even if the modeler does not look at the data, it would be possible to spot the source of the problem by looking at the results of the augmented battery of misspecification tests on the NLR, Table 2.3B. Testing the model assumptions would reveal the lack of statistical adequacy of the initial model and would hint the direction of correction for the modeler towards the DLR model.

The appropriateness of this specification is revealed by Tables 2-3E and 2-3F. No major discrepancies can be detected in this model (Table 2.3F). The actual and nom-

inal error probabilities are in check and the modeler can undertake reliable inference given the statistical adequacy of the specification.

It is important to mention the fact that, even at n=100, the joint F-test of the estimated versus the true coefficients has an actual error probability which is higher than the nominal. This is mainly due to the sample size and the amount of noise embedded into the specification by the proposed variance-covariance structure[10]. Table 2.3G shows the performance of the F-test when the sample size is increased. As expected, the actual error probabilities approximate the nominal error probabilities for the DLR while the contrary happens to the R-DLR. Notice also that the increased sample size increases the power of the Wald test to discriminate between the restricted AR(1) model and the unrestricted DLR model.

2.3.4 Experiment 4 - Heteroskedastic, Non-Linear Regr. Model

For this particular case, the probabilistic reduction resembles that of the NLR model except that $\mathbf{Z}_t^*$ refers to $\ln(\mathbf{Z}_t)$, that is:

$$\begin{aligned} f(\mathbf{Z}_1^*,...\mathbf{Z}_n^*;\boldsymbol{\phi}) &\overset{\mathsf{I}}{=} \prod_{t=1}^{n} f_t(\mathbf{Z}_t^*;\boldsymbol{\varphi}_t) \overset{\mathsf{IID}}{=} \prod_{t=1}^{n} f(\mathbf{Z}_t^*;\boldsymbol{\varphi}) = \prod_{t=1}^{n} f(\ln(Y_t),\ln(\mathbf{X}_t);\boldsymbol{\varphi}) = \\ &= \prod_{t=1}^{n} f(\ln(Y_t)\,|\,\ln(\mathbf{X}_t);\varphi_1)\cdot f(\ln(\mathbf{X}_t);\boldsymbol{\varphi}_2) \overset{\mathsf{NIID}}{=} \prod_{t=1}^{n} f(\ln(Y_t)\,|\,\ln(\mathbf{X}_t);\boldsymbol{\varphi}_1) \end{aligned} \tag{15}$$

where it is possible to ignore the marginal distribution of $\ln(\mathbf{X}_t)$ when imposing normality. The reduction gives raise to the Lognormal Regression Model, Table 2.F.

[10]Note that $x_{1t} = 0.6 + 0.7x_{1t-1} + \nu_{1t}$ with $Var(v_{1t}) = 0.51$ and $x_{2t} = 0.75 + 0.75x_{2t-1} + \nu_{2t}$ with $Var(\nu_{2t}) = 0.4375$.

Table 2.F: The Lognormal Regression Model [LogNR]	
	$y_t = \alpha_0 \prod_{k=1}^{K} \mathbf{X}_{kt}^{\beta_k} + u_t$
[1] Lognormality	$(y_t \mid \mathbf{X}_t = \mathbf{x}_t; \boldsymbol{\varphi}_1) \sim \mathsf{LN}(.,.)$
[2] Exponential Growth	$E(y_t \mid \mathbf{X}_t = \mathbf{x}_t; \boldsymbol{\varphi}_1) = \alpha_0 \prod_{k=1}^{K} \mathbf{X}_{kt}^{\boldsymbol{\beta}_k}$
[3] Heteroskedasticity	$Var(y_t \mid \mathbf{X}_t = \mathbf{x}_t; \boldsymbol{\varphi}_1) = \delta_0 \prod_{k=1}^{K} \mathbf{X}_{kt}^{2\boldsymbol{\beta}_k}$
[4] Independence	$\{(y_t \mid \mathbf{X}_t = \mathbf{x}_t),\ t \in \mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi}_1 := (\alpha_0,\ \delta_0,\ \boldsymbol{\beta}_k,\ k=0,1,\ldots,K,\ \sigma_0^2)$ do not change with t
$\alpha_0 = \exp\left\{\beta_0 + \frac{\sigma_0^2}{2}\right\}$, $\delta_0 = \alpha_0^2\left(e^{\sigma_0^2} - 1\right)$, $\beta_0 = \mu_1 - \boldsymbol{\beta}_1^{\intercal}\boldsymbol{\mu}_2$, $\boldsymbol{\beta}_k^{\intercal} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$, $\sigma_0^2 = \sigma_{11} - \boldsymbol{\sigma}_{21}^{\intercal}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$	

The data $\{(\ln(y_t), \ln(x_{1t}), \ln(x_{2t})),\ t=1,2,\ldots n\}$ is then generated via a three-variate, identical, and independently distributed normal process:

$$\begin{pmatrix} \ln(y_t) \\ \ln(x_{1t}) \\ \ln(x_{2t}) \end{pmatrix} \sim \mathsf{N}\left[\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1.2 & .7 & -.4 \\ .7 & 1 & .2 \\ -.4 & .2 & 1 \end{pmatrix}\right] \tag{16}$$

The parameterization of the previous probabilistic reduction is inherently nonlinear and heteroskedastic. The true model becomes:

$$Y_t = 3.545\,3(X_1^{.812\,5}/X_2^{.562\,5}) + v_t, \quad Var(Y_t|D_t) = 6.2981(X_1^{1.625}/X_2^{1.125}). \tag{17}$$

From the previous equation, it is possible to derive the average marginal response of Y to changes in X_1 and X_2. Since $\ln(X_{1t}) \sim \mathsf{N}(2,1)$, then $E(X_t) = 12.182\,4$. Similarly, for X_{2t}, $E(X_{2t}) = 33.115$. Thus, the average marginal response of Y to a change in X_1 is $\frac{\partial}{\partial X_1}(Y_t) = \frac{2.881}{X_1^{0.1875} X_2^{0.5625}}\Big|_{\bar{X}_1 = 12.18,\ \bar{X}_2 = 33.11} = 0.252$, and similarly for X_2 is $\frac{\partial}{\partial X_2}(Y_t) = -1.994\,2\ \frac{X_1^{0.812\,5}}{X_2^{1.562\,5}}\Big|_{\bar{X}_1 = 12.18,\ \bar{X}_2 = 33.11} = -.0641$.

Under the TA, the researcher would estimate (3), which in this case is tantamount to estimating the average marginal response for Y. The reliability of inference will be assessed through the default misspecification battery. The results, presented in Table 2.4A, suggest that heteroskedasticity and non-normality are a problem. This is

confirmed by the results of the full-fledged battery of misspecifications in Table 2.4B. The TA suggests two ways to correct the problem and obtain reliable estimators of $\boldsymbol{\beta}$. The first method involves the use of heteroskedastic consistent estimators ('Robust'), such as White or Newey-West, that warrant consistency and robustness of the estimators. The results, presented in Table 2.4C, indicate that the use of heteroskedasticity consistent standard errors does not alleviate the severity of the misspecification, as it can be seen in Table 2.4D.

It can be argued that, even though the model is misspecified, the estimates are relatively 'close' to the true average marginal responses. Thus, although reliability is compromised, precision is not. If this is indeed the case, the slight departure leads to minor effects in the estimation. This conclusion, however, may be both dangerous and deceiving. First, from Table 2.4C, the actual and the nominal probabilities differ considerably. The researcher would make a Type I error 5 times more often for β_1 and 6.5 times more often for β_2. But suppose, for the sake of the argument, that the 'Robust' OLS estimators are exactly equal to the true average marginal response. This would only allow the researcher to conduct inference in the 'vicinity' of the mean values of the regressors, the 'ball-park' argument. The researcher, however, does not know the size of that park nor she has anyway of finding out. Outside that range, within-sample prediction might be inaccurate, being the degree of inaccuracy a function of the curvature of the true regression line and the true form of the skedastic function (Fig. 2-4).

A second method to 'deal' with heteroskedasticity is to take the logarithm on both sides of (3), as if estimating elasticities. This variables' transformation will lead to meaningful and statistically adequate inference in the transformed data, that is:

$$\ln(y_t) = \beta_0 + \beta_1 \ln(x_{1t}) + \beta_2 \ln(x_{2t}) + u_t, \ t \in \mathbb{N}, \tag{18}$$

where the β's would be those obtained in Table 2.1A.

If the goal of this endeavor is the computation and statistical inference on the elasticities, the modeler would have done a good job capturing all the statistical

information. If, however, the goal is still to describe Y and to produce values for the marginal effects of the regressors on it, failing to take into consideration the change in the probabilistic structure of the data will lead the modeler to biased marginal effects and inaccurate predicted values. To see this, notice that simply taking the antilogarithm of the estimated elasticities from equation (18), using the results from Table 2.1A, yields, $E\left(Y_t|\mathbf{X}_t\right)$=$2.894(X_1^{.8125}/X_2^{.5625})$, different from (17). Even if the researcher obtains estimators 'close' to the true values, the average marginal effect for X_1 will be:

$$\frac{\partial}{\partial X_1}\left(Y\right) = 2.351[X_1^{0.1875}X_2^{0.5625}]^{-1} = .2054,$$

which is different from the true average marginal effect of .251. The conclusion will be off by almost 19 percent. Similarly, the average marginal effect of X_2:

$$\frac{\partial}{\partial X_2}\left(Y\right) = -\,1.627(X_1^{.8125}/X_2^{.5625}) = -\,.0523,$$

compared to the true value of $-.0641$, will also be off by almost 19 percent. In fact, the regression line obtained by taking the antilogarithm of (18) will effectively be modeling the conditional median instead of the conditional mean (Figure 2.5).

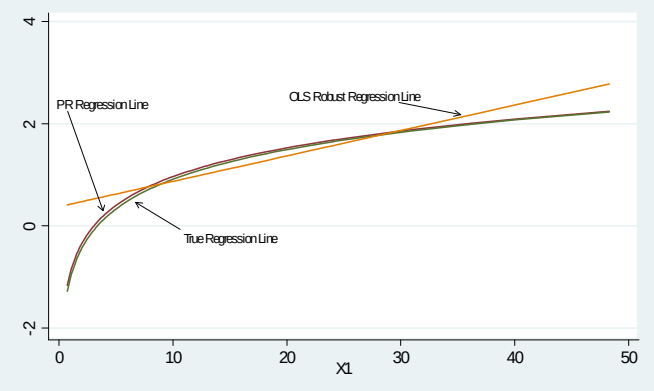


Fig. 2.4: True Regression vs. Robust Regression

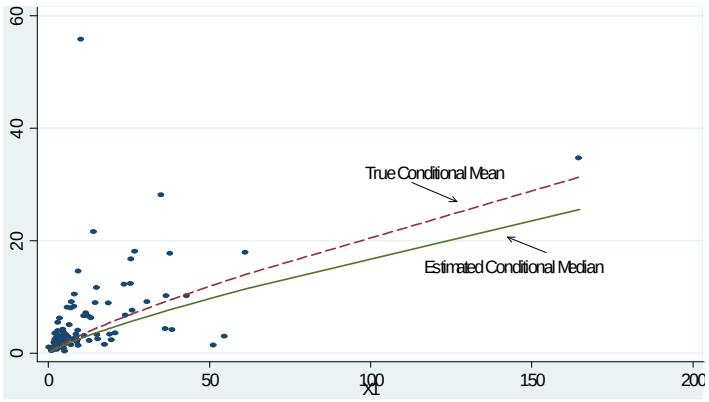
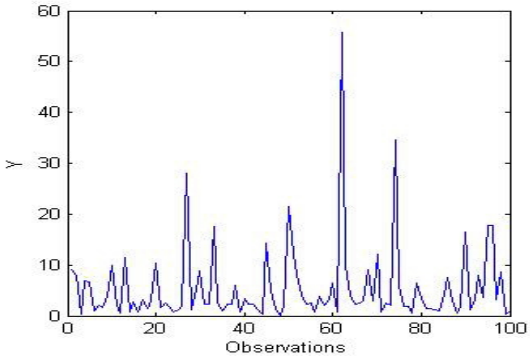


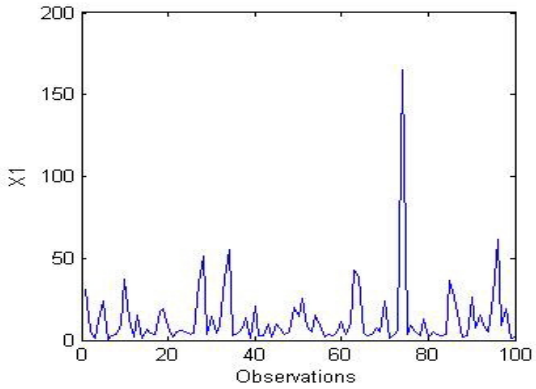
Fig. 2.5: Conditional Mean vs. Conditional Median

The PRA will, once again, rely on graphical techniques to inspect the data and assess the underlying probability distribution of the vector process $\mathbf{Z}_t$. The t-plots and the scatter-plots (Panel 2.4) reveal a series of strictly positive data with apparent index-homogeneity in both the mean and the variance and non-elliptical dependence. Suspecting lognormality, a simple Normality test of the joint process $\ln(\mathbf{Z}_t)$ and its marginal components will reveal that indeed the appropriate model is a lognormal regression model.

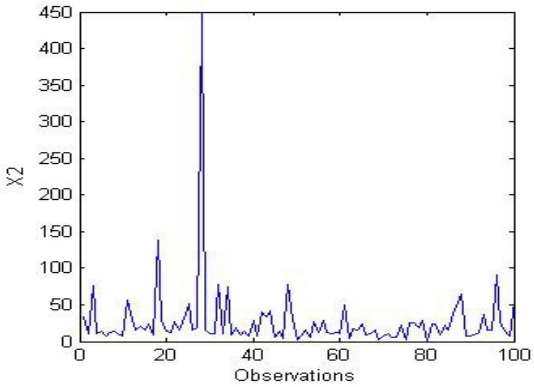
Panel 2.4: t-plots and scatter plots of (y_t, x_{1t}, x_{2t}) from Experiment 4



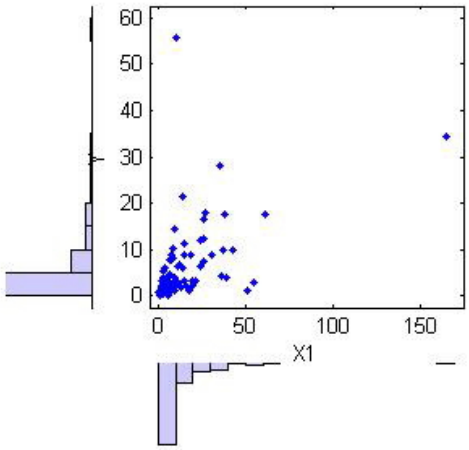
t-plot of y_t



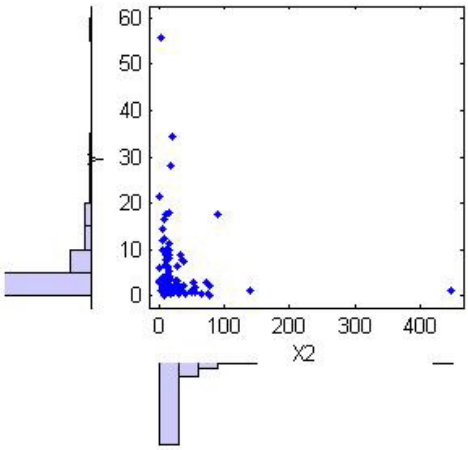
t-plot of x_{1t}



t-plot of x_{2t}



Scatter-plot of (y_t, x_{1t})



Scatter-plot of (y_t, x_{2t})

2.3.5 Joint Misspecification Testing

The result of applying the joint misspecification test to the raw and the corrected residuals are presented in Tables 2-5A and 2-5B, respectively. Both tables incorporate a normality test, the application of tests in the equations of system (6) individually, and the simultaneous joint test. From the outset, the increase in computational efficiency is evident for the first two rows [the normality test and the joint conditional mean test (M_1)] already account for 80 percent of the tests conducted on the full-fledged misspecification battery, with comparable results. Indeed, the first three rows of both tables contain enough information to assess the reliability of any estimation relying on the first two conditional moments, as it is the usual case in econometrics. The most computational efficiency is obtained with the simultaneous joint misspecification coefficient (last row in each table), which includes 100 percent of the tests conducted in the full-fledged misspecification battery, including distribution, with similar results.

Interestingly, Table 2.5A reveals that when statistical misspecification is an issue, the reliability of inference is compromised in every subsequent conditional moment. If the source of misspecification exists in the first conditional moment, as it would be the case under the NLR-trend and the Dyn-NLR models, the consequences of misspecification are transferred to the conditional skewness and kurtosis moments. This puts in doubt the idea that these higher moments would remain constant under the presence of trends or additional dynamics in the regression equation but ignored in the proposed specification. The implication that can be drawn from this fact is that any test that assumes constancy of the third and fourth standardized moments would be effectively invalid.

It seems that not correcting the residuals exaggerates the effects of model departures in the separate testing while the probativeness of the simultaneous test is increased when using the corrected residuals. In Table 2.5A, for instance, in the NLR-trend model, the misspecification of the conditional mean is transferred to the

testing of the third conditional moment, and its power seems to increase with the sample size. A similar argument can be made about the Dyn-NLR model, where the contamination seems to trickle through the higher three conditional moments. This should not be the case since the departures belong only to the specification of the conditional mean in these two cases. The case for the lognormal regression is different since departures in the first and second moments are expected. From Table 2.5B, it is clear that 'purging' the residuals at every stage isolates the source of the misspecification to a particular central moment for no major additional sources of misspecification are present in the three higher moments of the NLR-trend model and the Dyn-NLR model. The misspecification in the second moment for the lognormal models persists even after 'correcting' the residuals, as it should be expected.

The simultaneous joint misspecification test performs well capturing the discrepancy between the actual and the nominal error probabilities in both residual regimes. The power of the test increases with the sample size but its power is reduced using the corrected residuals under the lognormal model. It is important to realize that, although an excellent approximation to the degree of misspecification of the models, it does not shed much light on the source of the misspecification. The same can be said about the individual equations in the system of auxiliary regression.

The degree of probativeness can be increased by separating the sources of misspecification in each individual test. The results from testing the four sources of misspecification in the conditional mean (orthogonality, structural, functional, and dependence) are shown in Table 2.6. The orthogonality test, the first indication of an incomplete systematic component, seems to perform relatively well in the NLR-trend and the LogNR model but not on the Dyn-NLR. It cannot be considered a sufficient test for misspecification. The other components perform as it should be expected. In the NLR-trend model, the separate testing is able to detect structural misspecification accurately, as well as in the LogNR model. Functional misspecification is adequately hinted in the LogNR model.

It can be argued that these joint misspecification tests, although useful, cannot be considered in isolation without a systematic analysis of the marginal and joint distributions of the observed variables, but would prove crucial for directing the researcher to the likely sources of misspecification; case in point, the Dyn-NLR model, where the t-plots would be able to reveal the existence of dependence and tilt the scale towards dependence misspecification.

2.4 Conclusion

Misspecification testing, an important component in econometric modeling, does not seem to have the recognition it deserves. One part of the problem is the fact that no general guideline for probing model departures has been established. The other part is the fact that the usual econometric specifications are not supplemented, explicitly or implicitly, with testable model assumptions. The PRA offers the researcher a guideline for misspecification testing through the use of auxiliary regressions and sheds light on the explicit and implicit assumptions of the proposed models.

The current methodology for model specification, more often than not, precludes the researcher from confronting the theory with the data in a credible and reliable manner. As long as the error term, and not the observed variables, determines the statistical properties of the model, the researcher is unable to separate the two sources of error: substantive and statistical. This separation is crucial since the data's own ontology allows the modeler to create complete summaries of the data, independently from the theory, that can be qualified as 'good' or 'bad'. The PRA considers the standalone nature of the data and provides a comprehensive modeling framework to create adequate summaries of the observed variables.

Theory first specifications carry unto the models their caveats. First, theories are vague, they are unable to relate the idealized representation of the variables involved in the model and the observable variables presented to the econometrician. Second, theories are incomplete, they almost never specify under what conditions the *ceteris*

paribus clause hold. What secondary variables should be held constant in order to isolate the primary effects? When attempting to establish a relation between interest rate and investment, do changes in the GNP, the level of unemployment, technology and foreign competition have to be taken into account? When relating education and earnings, is it necessary to add IQ, work effort, occupational choice, or family background? Third, theories are deterministic, they do not account for the stochastic nature of observational data and, as a result, they have no information regarding the distributional or sampling properties of their arguments.

It appears that theories are inadequate sources for model specification and usually lead to both substantive and statistically misspecified models. When Ireland's (2004) DSGE model was tested for misspecification by Juselius and Franchi (2007), it failed miserably. This example and other has raised suspicion of the current state of econometrics to even non-econometricians. Robert M. Solow's (2007) reflection:

> "The DSGE model is logically incorrect, but because it does not pass the judgment test; it is simply beyond belief that with all the assumptions the DSGE model must make to arrive at a formal model, that that model shed much light on the type of short-run problems that the macro economy often experiences. It simply does not meet the common sense test, so unless there are other arguments for using it, it is not an approach to policy that anyone other than someone who has been taught it is the only correct theory would use as the sole approach for thinking about macroeconomic policy." (p. 235)

What is then the role of theory? Arguably, to provide an explanation for the phenomenon of interest, but it is only when a description of the evidence becomes statistically adequate that one can establish a dialogue between theory and data. At this point, the theorist can not only deal with issues of theoretical and external validity but also with matters of inaccurate data or incongruous measurements (Spanos, 2005). Economic theory has to make room for the distinction between and account for the discrepancies between a theoretical and an estimable model.

2.5 Appendix 2.A: Tables

Table 2.1A: True: NLR / / Estimated: NLR				
	n=50		n=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.063	.3216	1.061	.2241
$\hat{\beta}_1$	.8121	.0948	.8128	.0660
$\hat{\beta}_2$	$-.5630$	.0948	$-.5623$	.0660
$\hat{\sigma}$	.4071	.0839	.4062	.0585
R^2	.6666	.0789	.6640	.0559
t-statistics	Mean	% reject(.05)	Mean	% reject(.05)
$\tau_{\beta_0}=\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	.0038	.0546	$-.0047$	.0535
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	$-.0049$	.0502	.0064	.0502
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$	$-.0069$	.0534	.0031	.0512
F-statistic	Mean	% reject(.05)	Mean	% reject(.05)
$\phi_{\beta_0,\beta_1,\beta_2}$	1.059	.0524	1.026	.0510
Default Misspecification Battery (% reject .05)				
SW Normality	.9752	.0613	.9862	.0605
Durbin Watson	1.997	.0897	2.001	.0723
White's Homosked.	4.867	.0506	4.883	.0546

Top row: true underlying model and estimated model.

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.1B: True: NLR / / Estimated: NLR				
	n=50		n=100	
Misspecification Test	Mean	% reject(.05)	Mean	% reject(.05)
D'AP Normality	1.904	.0443	1.889	.0391
SW Normality	.9752	.0613	.9862	.0605
Durbin Watson	1.997	.0897	2.001	.0723
AC Test (1):	1.003	.0443	.9862	.0467
AC Test (2):	1.036	.0477	1.010	.0489
White's Homosked.	4.867	.0506	4.883	.0546
BP Homosked.	1.834	.0410	1.915	.0441
Reset (2)	1.028	.0477	1.029	.0524
Joint Mean (A)	1.053	.0510	1.016	.0447
Trend in mean	−.0045	.0548	.0077	.0519
Reset (2) linearity	.0103	.0487	.0124	.0525
$\widehat{u}_{t-1}$ in mean (1)	−.2833	.0493	−.2079	.0483
Joint Mean (B)	1.051	.0525	1.017	.0494
trend in mean	.0044	.0548	.0086	.0527
Reset (2) linearity	.0109	.0507	.0091	.0523
y_{t-1}, x_{t-1} in mean	1.049	.0483	1.021	.0500
Joint Variance	.9853	.0407	.9833	.0468
trend in variance	.0322	.0506	.0077	.0500
Reset 2 Homosk.	−.0548	.0434	−.0454	.0472
ARCH(1)	−.2641	.0284	−.1889	.0371

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.2A: True: NLR-trend / / Estimated: NLR [No Trend]				
	n=50		n=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	−.1230	.2623	−.1892	.1684
$\hat{\beta}_1$	1.003	.0974	1.008	.0678
$\hat{\beta}_2$	−.2621	.0737	−.2586	.0510
$\hat{\sigma}$	.5588	.1150	.5617	.0804
R^2	.9441	.0122	.9841	.0023
t-statistics	Mean	% reject(.05)	Mean	% reject(.05)
$\tau^*_{\beta_0}=\frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	−.4653	.0688	−1.126	.1983
$\tau^*_{\beta_1}=\frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	10.52	1.000	15.01	1.000
$\tau^*_{\beta_2}=\frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}}$	−3.629	.9256	−5.124	.9974
$\tau_{\beta_0}=\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	−4.562	.9955	−7.486	1.000
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	2.001	.4874	2.915	.8111
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$	4.160	.9702	6.018	.9996
F-statistic	Mean	% reject(.05)	Mean	% reject(.05)
F at zero	1438	1.000	9521	1.000
$\phi_{\beta_0,\beta_1,\beta_2}$	874.0	1.000	6724	1.000
Default Misspecification Battery (% reject .05)				
Normality	.9754	.0587	.9861	.0623
D-W	2.025	.0782	2.015	.0674
Homosk.	5.130	.0511	5.206	.0575

Top row: true underlying model and estimated model.

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.2B: True: NLR-trend / / Estimated: NLR [No Trend]

	n=50		n=100	
Misspecification Test	Statistic	% reject(.05)	Statistic	% reject(.05)
D'AP Normality	1.890	.0415	1.889	.0408
SW Normality	.9754	.0587	.9861	.0623
Durbin Watson	2.025	.0782	2.015	.0674
AC Test (1):	1.005	.0445	1.013	.0474
AC Test (2):	1.006	.0454	1.011	.0454
White's Homosked.	5.130	.0511	5.206	.0575
BP Homosked.	1.844	.0413	1.940	.0462
Reset (2)	1.056	.0525	.9836	.0468
Joint Mean (A)	7.194	.9166	13.58	.9994
Trend in mean	4.153	.9702	6.047	1.000
Reset (2) linearity	−.0039	.0522	−.0070	.0492
$\widehat{u}_{t-1}$ in mean (1)	−.3080	.0492	−.2373	.0519
Joint Mean (B)	4.747	.8567	8.578	.9978
trend in mean	2.865	.7837	4.066	.9786
Reset (2) linearity	−.0418	.0517	−.0132	.0507
y_{t-1}, x_{t-1} in mean	1.056	.0507	1.044	.0538
Joint Variance	1.020	.0457	1.010	.0469
trend in variance	.0597	.0590	.0711	.0558
Reset 2 Homosk.	−.0641	.0578	−.0784	.0540
ARCH(1)	−.3044	.0335	−.2562	.0342

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.2C: True: NLR-trend / / Estimated: NLR-trend				
	n=50		n=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.060	.3635	1.062	.2520
$\hat{\beta}_1$	.8113	.0959	.8115	.0665
$\hat{\beta}_2$	$-.5602$	.0960	$-.5615$	.0664
$\hat{\delta}$	.1805	.0435	.1811	.0299
$\hat{\sigma}$	.4067	.0845	.4069	.0588
R^2	.9602	.0089	.9886	.0016
$\tilde{R}^2$	.6666	.0789	.6640	.0559
t-statistics	Mean	% reject(.05)	Mean	% reject(.05)
$\tau^*_{\beta_0}=\frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	2.978	.8105	4.257	.9841
$\tau^*_{\beta_1}=\frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	8.645	1.000	12.33	1.000
$\tau^*_{\beta_2}=\frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}}$	-5.963	.9993	-8.541	1.000
$\tau^*_{\delta}=\frac{\hat{\delta}}{\hat{\sigma}_{\delta_2}}$	4.239	.9752	6.109	1.000
$\tau_{\beta_0}=\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	$-.0052$	.0443	.0000	.0498
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	$-.0114$	.0499	$-.0132$	.049
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$	.0236	.0526	.0154	.05
$\tau_{\delta}=\frac{\hat{\delta}-\delta}{\hat{\sigma}_{\delta}}$	$-.0128$	.0491	$-.0040$	.0509
F-statistic	Mean	% reject(.05)	Mean	% reject(.05)
F at zero	1488	1.000	9872.8	1.000
$\phi_{\beta_0,\beta_1,\beta_2}$	1.035	.0456	1.015	.0515

Top row: true underlying model and estimated model.

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.2D: True: NLR-trend / / Estimated: NLR-trend				
	n=50		n=100	
Misspecification Test	Statistic	% reject(.05)	Statistic	% reject(.05)
D'AP Normality	1.907	.0444	1.890	.0408
SW Normality	.9752	.0625	.9862	.0594
Durbin Watson	2.042	.0724	2.021	.0613
AC Test (1):	1.053	.0521	1.031	.0514
AC Test (2):	1.015	.0446	1.021	.0500
White's Homosked.	4.979	.0449	5.031	.0486
BP Homosked.	1.844	.0413	1.940	.0462
Reset (2)	.8245	.0285	.9237	.0400
Joint Mean (A)	.7184	.0225	.6855	.0209
Trend in mean	.0010	.0009	.0024	.0000
Reset (2) linearity	−.0019	.0558	−.0088	.0506
$\widehat{u}_{t-1}$ in mean (1)	−.3829	.0560	−.2896	.0535
Joint Mean (B)	.8465	.0261	.8259	.0260
trend in mean	.1381	.0159	.1148	.0119
Reset (2) linearity	−.0163	.0546	−.0157	.0502
y_{t-1}, x_{t-1} in mean	1.058	.0494	1.045	.0539
Joint Variance	1.009	.0442	1.001	.0449
trend in variance	.0431	.0506	.0620	.0498
Reset 2 Homosk.	−.0512	.0494	−.0661	.0496
ARCH(1)	−.3158	.0306	−.2594	.0358

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.3A: True: Dyn-NLR / / Estimated: NLR [Static]				
	n=50		n=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	.9565	.3915	.9767	.2582
$\hat{\beta}_1$	.7830	.1030	.7934	.0684
$\hat{\beta}_2$	−.5080	.1054	−.5212	.0691
$\hat{\sigma}$	.4215	.1158	.4157	.0765
R^2	.6490	.1199	.6715	.0812
t-statistics	Mean	% reject(.05)	Mean	% reject(.05)
$\tau^*_{\beta_0}=\frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	2.590	.6360	3.889	.8947
$\tau^*_{\beta_1}=\frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	7.962	.9993	11.85	1.000
$\tau^*_{\beta_2}=\frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}}$	−5.092	.9628	−7.738	.9999
$\tau_{\beta_0}=\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	.8272	.2186	1.290	.3171
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	−.1803	.1488	−.1299	.1515
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$	−.5714	.1879	−1.025	.2618
F-statistic	Mean	% reject(.05)	Mean	% reject(.05)
F at zero	82.504	1.000	160.82	1.000
$\phi_{\beta_0,\beta_1,\beta_2}$	2.8483	.3798	3.4739	.5069
Normality	.9692	.1458	.9825	.1430
Default Misspecification Battery (% reject .05)				
D-W	1.2428	.9128	1.1937	.9946
Homosk.	5.8098	.0983	6.0339	.1022

Top row: true underlying model and estimated model.

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.3B: True: Dyn-NLR / / Estimated: NLR [Static]				
	n=50		n=100	
Misspecification Test	Statistic	% reject(.05)	Statistic	% reject(.05)
D'AP Normality	3.412	.1409	3.559	.1291
SW Normality	.9692	.1458	.9825	.1430
Durbin Watson	1.242	.9128	1.193	.9946
AC Test (1):	7.537	.6698	18.14	.9685
AC Test (2):	3.855	.5873	7.612	.9415
White's Homosked.	5.809	.0983	6.033	.1022
BP Homosked.	2.840	.1096	2.931	.1086
Reset (2)	1.531	.1017	1.514	.1033
Joint Mean (A)	3.500	.5312	6.882	.9165
Trend in mean	−.0262	.1066	.0103	.0818
Reset (2) linearity	.0029	.0708	−.0052	.0642
$\widehat{u}_{t-1}$ in mean (1)	2.115	.5284	3.801	.9464
Joint Mean (B)	3.124	.5726	5.214	.9160
trend in mean	−.0518	.0965	−.0130	.0759
Reset (2) linearity	.0035	.0674	−.0052	.0591
y_{t-1}, x_{t-1} in mean	3.499	.5298	7.023	.9193
Joint Variance	1.403	.1117	1.795	.1973
trend in variance	−.4291	.1041	−.4109	.0890
Reset 2 Homosk.	−.0279	.0670	−.0259	.0607
ARCH(1)	.1322	.0619	.9070	.1846

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.3C: True: Dyn-NLR / / Estimated: Dyn-NLR [Restricted]				
	n=50		n=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	.8663	.4349	.8741	.3028
$\hat{\beta}_1$	.8044	.1099	.8039	.0766
$\hat{\beta}_2$	−.4922	.1153	−.4940	.0800
$\hat{\rho}$	.3635	.1428	.3968	.0972
$\hat{\sigma}$	.3255	.1053	.3305	.0481
R^2	.6201	.1053	.6128	.0730
t-statistics	Mean	% reject(.05)	Mean	% reject(.05)
$\tau^*_{\beta_0}=\frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	2.105	.5073	2.962	.7953
$\tau^*_{\beta_1}=\frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	7.568	1.000	10.64	1.000
$\tau^*_{\beta_2}=\frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}}$	−4.456	.9633	−6.301	1.000
$\tau_{\beta_0}=\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	.5338	.1091	.7589	.1321
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	.0191	.0742	.0185	.0624
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$	−.3759	.097	−.5379	.1004
F-statistic	Mean	% reject(.05)	Mean	% reject(.05)
F at zero	104.9	1.000	198.6	1.000
$\phi_{\beta_0,\beta_1,\beta_2}$	3.132	.4123	3.645	.5082
Default Misspecification Battery (% reject 0.05)				
Normality	.9746	.0669	.9858	.0645
Durbin-h	1.936	.0202	1.963	.0078
Homosk.	5.076	.0573	5.031	.0561

Top row: true underlying model and estimated model.

The percentage of rejections represents the actual error probabilites

versusthe nominal error probabilities (5%).

Table 2.3D: True: Dyn-NLR / / Estimated: Dyn-NLR [Restricted]				
	n=50		n=100	
Misspecification Test	Statistic	% reject(.05)	Statistic	% reject(.05)
D'AP Normality	1.969	.0496	1.960	.0466
SW Normality	.9746	.0669	.9858	.0645
CFR	1.073	.0817	2.072	.1291
AC Test (1):	.2598	.0018	.2247	.0003
AC Test (2):	.8549	.0336	.9829	.0558
White's Homosked.	5.076	.0573	5.031	.0561
BP Homosked.	2.042	.0535	2.259	.0689
Reset (2)	1.064	.0522	1.033	.0504
Joint Mean (A)	.8333	.0282	.7786	.0243
Trend in mean	−.0231	.0648	.0039	.0534
Reset (2) linearity	−.1434	.0527	−.2210	.0495
$\widehat{u}_{t-2}$ in mean (1)	−.1143	.0044	−.0644	.0008
Joint Mean (B)	1.002	.0471	1.028	.0578
trend in mean	−.0200	.0807	.0055	.0637
Reset (2) linearity	−.0004	.0551	−.0022	.0490
y_{t-2}, x_{t-2} in mean	.9449	.0428	1.017	.0609
Joint Variance	.9977	.0411	.9816	.0409
trend in variance	.0148	.0569	−.0274	.0540
Reset 2 Homosk.	−.0744	.0442	−.0667	.0466
ARCH(1)	−.4059	.0312	−.2854	.0327

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.3E: True: Dyn-NLR / / Estimated: Dyn-NLR				
	n=50		n=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	.7377	.4085	.6987	.2677
$\hat{\beta}_1$	.7837	.1220	.7904	.0832
$\hat{\beta}_2$	−.4396	.1265	−.4435	.0872
$\hat{\rho}$	.3434	.1358	.3838	.0930
$\hat{\sigma}$	.3493	.0810	.3428	.0535
R^2	.7255	.0930	.7366	.0638
t-statistics	Mean	%reject(.05)	Mean	%reject(.05)
$\tau^*_{\beta_0}=\frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	1.855	.4400	2.630	.7163
$\tau^*_{\beta_1}=\frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	6.587	.9997	9.607	1.000
$\tau^*_{\beta_2}=\frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}}$	−3.561	.8858	−5.139	.9944
$\tau_{\beta_0}=\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	.1788	.1018	.1382	.0766
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	−.1572	.0699	−.1470	.0689
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$	.1211	.1009	.1165	.0918
F-statistic				
F at zero	51.03	1.000	100.5	1.000
$\phi_{\beta_0,\beta_1,\beta_2}$	1.407	.1476	1.231	.1060
Default Misspecification Battery (% reject 0.05)				
Normality	.9745	.0728	.9846	.1025
Durbin-h	1.913	.0476	1.938	.0340
Homosk.	5.143	.0727	5.221	.0693

Top row: true underlying model and estimated model.

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.3F: True: Dyn-NLR / / Estimated: Dyn-NLR				
	$n = 50$		$n = 100$	
Misspecification Test	Statistic	% reject(.05)	Statistic	% reject(.05)
D'AP Normality	2.106	.0605	2.584	.0889
SW Normality	.9745	.0728	.9846	.0525
CFR	1.073	.0404	1.058	.0164
AC Test (1):	.2804	.0008	.2583	.0002
AC Test (2):	.0000	.0000	.0000	.0000
White's Homosked.	5.073	.0607	5.221	.0693
BP Homosked.	5.143	.0724	5.947	.0444
Reset (2)	1.088	.0536	1.024	.0489
Joint Mean (A)	1.285	.0893	1.194	.0847
Trend in mean	−.0608	.0973	−.0200	.0787
Reset (2) linearity	.0066	.0555	−.0033	.0541
$\widehat{u}_{t-1}$ in mean (1)	.2505	.0728	.2813	.0818
Joint Mean (B)	.5160	.0125	.4565	.0078
trend in mean	−.0506	.0993	−.0146	.0757
Reset (2) linearity	.0026	.0560	−.0072	.0534
y_{t-2}, x_{t-2} in mean	.1870	.0001	.1122	.0001
Joint Variance	1.055	.0457	1.057	.0515
trend in variance	−.1719	.0702	−.2875	.0735
Reset 2 Homosk.	−.0740	.0478	−.0623	.0487
ARCH(1)	−.5175	.0358	−.4112	.0366

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.3G: F-test & CFR-test Type I Error						
	DLR		R-DLR		CFR	
	F-stat.	% .05	F-stat.	% .05	Wald-stat.	% .05
50	1.407	.1476	3.132	.4123	1.073	.0817
100	1.231	.1060	3.645	.5082	2.072	.1291
250	1.091	.0736	5.371	.7114	4.670	.3158
500	1.051	.0633	8.359	.8835	8.504	.6150
1000	1.041	.0600	14.25	.9840	15.77	0.913
10000	.9945	.0340	122.9	1.000	139.0	1.000

F-test of the coefficients taking simultaneously their true values and Wald Test of the Common Factor Restrictions

Table 2.4A: True: LogNR / / Estimated: NLR				
	n=50		n=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	3.163	1.142	3.177	.8316
$\hat{\beta}_1$	.2955	.0563	.2821	.0395
$\hat{\beta}_2$	$-.0560$	.0213	$-.0506$	.0147
$\hat{\sigma}$	32.4	52.92	34.78	52.39
R^2	.4174	.1714	.3837	.1393
t-statistics	Mean	% reject(.05)	Mean	% reject(.05)
$\tau^*_{\beta_0}=\frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	3.168	.7998	4.225	.9108
$\tau^*_{\beta_1}=\frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	5.775	.9785	7.574	.9993
$\tau^*_{\beta_2}=\frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}}$	-2.711	.7855	-3.522	.9679
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	.1606	.3967	.1382	.4548
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$	1.287	.3048	1.711	.3774
F-statistic	Mean	% reject(.05)	Mean	% reject(.05)
F at zero	36.20	.9997	57.95	1.000
ϕ_{β_1,β_2}	8.173	.5030	10.44	.5960
Default Misspecification Battery (% reject 0.05)				
Normality	.7548	.9877	.7062	1.000
D-W	1.999	.0851	1.997	.0762
Homosk.	15.86	.6081	25.53	.8131

Top row: true underlying model and estimated model.

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2-4B: True: LogNR / / Estimated: NLR				
	$n = 50$		$n = 100$	
Misspecification Test	Statistic	% reject(.05)	Statistic	% reject(.05)
D'AP Normality	38.95	.9666	83.57	1.000
SW Normality	.7548	.9877	.7062	1.000
Durbin Watson	1.999	.0851	1.997	.0762
AC Test (1):	.8319	.0324	.8693	.0426
AC Test (2):	1.011	.0569	.9896	.0584
White's Homosked.	15.86	.6081	25.53	.8131
BP Homosked.	29.02	.5691	85.31	.8057
Reset (2)	6.642	.3034	8.280	.3701
Joint Mean (A)	2.864	.2334	3.397	.2942
Trend in mean	.0223	.0546	.0004	.0489
Reset (2) linearity	.6133	.2992	.3850	.3653
$\widehat{u}_t$ in mean (1)	$-.2640$	.0350	$-.1810$	.0397
Joint Mean (B)	2.186	.2141	2.468	.2623
trend in mean	.0258	.0566	.0009	.0509
Reset (2) linearity	.5962	.2942	.3800	.3632
y_{t-1}, x_{t-1} in mean	1.057	.0616	1.011	.0611
Joint Variance	9.476	.3818	12.50	.5246
trend in variance	.0218	.0409	$-.0083$	.0357
Reset 2 Homosk.	2.977	.4678	3.836	.6305
ARCH(1)	$-.2552$	.0267	$-.1769$	.0285

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.4C: True: LogNR / / Estimated: NLR [Robust]				
	n=50		n=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	3.172	1.020	3.174	.8181
$\hat{\beta}_1$	.2942	.0684	.2820	.0750
$\hat{\beta}_2$	−.0560	.0208	−.0505	.0158
$\hat{\sigma}$	32.50	52.98	34.75	52.22
R^2	.4155	.1710	.3939	.1388
t-statistics	Mean	% reject(.05)	Mean	% reject(.05)
$\tau^*_{\beta_0}=\frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	3.522	.8612	4.359	.9101
$\tau^*_{\beta_1}=\frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	4.579	.9209	4.632	.9691
$\tau^*_{\beta_2}=\frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}}$	−2.903	.8582	−3.372	.9453
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	−.2705	.2514	−.2368	.2061
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$	1.663	.3289	1.759	.3606
F-statistic	Mean	% reject(.05)	Mean	% reject(.05)
F at zero	36.01	.9998	57.90	1.000
ϕ_{β_1,β_2}	8.225	.5050	10.49	.5965
Default Misspecification Battery (% reject 0.05)				
Normality	.7548	.9875	.7065	1.000
D-W	1.999	.0835	1.998	.0797
Homosk.	15.73	.6028	25.66	.8136

Top row: true underlying model and estimated model.

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.4D: True: LogNR / / Estimated: NLR [Robust]				
	$n = 50$		$n = 100$	
Misspecification Test	Statistic	% reject(.05)	Statistic	% reject(.05)
D'AP Normality	39.21	.9686	83.45	.9999
SW Normality	.7548	.9875	.7065	1.000
Durbin Watson	1.999	.0835	1.998	.0797
AC Test (1):	.8283	.0326	.8898	.0423
AC Test (2):	.9997	.0563	1.006	.0611
White's Homosked.	15.73	.6028	25.66	.8136
BP Homosked.	28.23	.5654	65.61	.8043
Reset (2)	6.531	.2963	8.039	.3698
Joint Mean (A)	2.830	.2283	3.325	.2958
Trend in mean	.0183	.0551	.0068	.0493
Reset (2) linearity	.5923	.2929	.3740	.3654
$\widehat{u}_t$ in mean (1)	$-.2400$	.0345	$-.1839$	.0410
Joint Mean (B)	2.164	.2090	2.433	.2666
trend in mean	.0229	.0562	.0079	.0513
Reset (2) linearity	.5773	.2898	.3680	.3625
y_{t-1}, x_{t-1} in mean	1.046	.0617	1.029	.0652
Joint Variance	9.737	.3770	12.79	.5298
trend in variance	.0226	.0418	$-.0039$	.0376
Reset 2 Homosk.	2.955	.4621	3.870	.6369
ARCH(1)	$-.2569$	.0262	$-.1653$	.0299

The percentage of rejections represents the actual error probabilites versus the nominal error probabilities (5%).

Table 2.5A: Five-Dimension M-S Test versus Simultaneous M-S Test

	NLR		NLR-trend		Dyn-NLR		LogNR	
	n=50	n=100	n=50	n=100	n=50	n=100	n=50	n=100
Normality	.0448	.0375	.0415	.0408	.1409	.1291	.9667	.9999
M_1	.0515	.0509	.6671	.9835	.1405	.1435	.2584	.5291
M_2	.0472	.0491	.0418	.0447	.2145	.2005	.3801	.5739
M_3	.0517	.0559	.4428	.8695	.2600	.2512	.1950	.3343
M_4	.0525	.0515	.0884	.0833	.2896	.2698	.2530	.3410
Simultaneous	.0588	.0586	.4330	.6725	.3255	.4390	.4560	.7010

Normality, Auxiliary Regression M-S Testing of the Conditional Moments (M),
and Simultaneous M-S Testing using the simple standardized residuals at (0.05)

Table 2.5B: Five-Dimension M-S Test$^{(*)}$ versus Simultaenous M-S Test$^{(*)}$

	NLR		NLR-trend		Dyn-NLR		LogNR	
	n=50	n=100	n=50	n=100	n=50	n=100	n=50	n=100
Normality	.0448	.0375	.0415	.0408	.1409	.1291	.9667	.9999
M_1^*	.0515	.0509	.6671	.9835	.1405	.1435	.2584	.5291
M_2^*	.0472	.0491	.0418	.0461	.0869	.1256	.3681	.5767
M_3^*	.0517	.0559	.0000	.0000	.0015	.0275	.0132	.1041
M_4^*	.0525	.0515	.0378	.0423	.0864	.1597	.1965	.2811
Simultaneous	.0588	.0586	.6110	.9730	.3770	.8390	.3830	.6750

Normality, Auxiliary Regression M-S Testing of the Conditional Moments (M),
and Simultaneous M-S Testing using the corrected standardized residuals at (0.05)

Table 2.6: Conditional Mean M-S Test						
	NLR-trend		Dyn-NLR		NLR	
	n=50	n=100	n=50	n=100	n=50	n=100
Orthogonality	.7906	.9988	.1035	.1134	.5407	.7996
Functional	.0532	.0485	.1158	.1264	.6045	.8542
Structural	.9597	.9991	.5336	.5438	.3344	.3343
Dependence	.0459	.0499	.1848	.2109	.0657	.0626

2.6 Appendix 2.B: Full-Fledged M-S Battery

For the simulations reported in this paper, several misspecification tests are applied.

Normality: D'Agostino-Pearson K2 omnibus test and Shapiro-Wilk which have shown to have good power over a wide range of alternative distributions.

Independence: The three autocorrelation tests include the Durbin-Watson and test of independence using the following auxiliary regressions:

$$\widehat{u}_t=\beta_0^*+\boldsymbol{\beta}_1'^*\mathbf{x}_t+\lambda_1 y_{t-1}+\boldsymbol{\lambda}_2'\mathbf{x}_{t-1}+v_t, \qquad \widehat{u}_t=\beta_0^*+\boldsymbol{\beta}_1'\mathbf{x}_t+\lambda_1\widehat{u}_{t-1}+v_t$$

Linearity: The functional form test here is a RESET(2) test. It is based on the following auxiliary regression:

$$\widehat{u}_t=\beta_0^*+\boldsymbol{\beta}_1'\mathbf{x}_t+\lambda_1\widehat{y}_t^2+v_t.$$

Homoskedasticity: The White's homoskedasticity test and the Breusch–Pagan homoskedasticity test are used.

Structural stability: The parameters' stability is assessed through the following joint tests:

$$\text{Joint Mean (A):} \quad \widehat{u}_t=\beta_0^*+\boldsymbol{\beta}_1'^*\mathbf{x}_t+\gamma_1\widehat{y}_t^2+\gamma_2 t+\gamma_3\widehat{u}_{t-1}+v_t$$

$$\text{Joint Mean (B):} \quad \widehat{u}_t=\beta_0^*+\boldsymbol{\beta}_1'^*\mathbf{x}_t+\gamma_1\widehat{y}_t^2+\gamma_2 t+\gamma_3 y_{t-1}+\boldsymbol{\gamma}_4'\mathbf{x}_{t-1}+v_t$$

Parameter stability, linearity, and independence are then assessed separately by testing the individual component of this joint test one by one.

Conditional Variance Test: The joint variance test simultaneously assesses the stability, homoskedasticity, and dynamic homoskedasticity:

$$\text{Joint Variance:} \quad \widehat{u}_t^2=\beta_0^*+\gamma_1\widehat{y}_t^2+\gamma_2 t+\gamma_3\widehat{u}_{t-1}^2+v_t$$

Parameter stability, static, and dynamic homoskedasticity are then assess separately by testing the individual components of this joint test one by one.

2.7 Appendix 2.C: Joint Conditional Moments Test

The system of auxiliary regressions:

$$\left.\begin{array}{lcl} E\left(\frac{u_t}{\sigma_t}\right)=0 & \Leftrightarrow & \left(\frac{\widehat{u}_t}{\widehat{\sigma}_t}\right)=\gamma_{10}+\boldsymbol{\gamma}_{11}'\mathbf{X}_t+\boldsymbol{\gamma}_{12}'\boldsymbol{\Upsilon}+\boldsymbol{\gamma}_{13}'\mathbf{X}_t^2+\boldsymbol{\gamma}_{14}'\mathbf{X}_{t-1}+\varepsilon_{1t}, \\ E\left(\frac{u_t^2}{\sigma_t^2}\right)=1 & \Leftrightarrow & \left(\frac{\widehat{u}_t}{\widehat{\sigma}_t}\right)^2=\gamma_{20}+\boldsymbol{\gamma}_{21}'\mathbf{X}_t+\boldsymbol{\gamma}_{22}'\boldsymbol{\Upsilon}+\boldsymbol{\gamma}_{23}'\mathbf{X}_t^2+\boldsymbol{\gamma}_{24}'\mathbf{X}_{t-1}+\varepsilon_{2t}, \\ E\left(\frac{u_t^3}{\sigma_t^3}\right)=0 & \Leftrightarrow & \left(\frac{\widehat{u}_t}{\widehat{\sigma}_t}\right)^3=\gamma_{30}+\boldsymbol{\gamma}_{31}'\mathbf{X}_t+\boldsymbol{\gamma}_{32}'\boldsymbol{\Upsilon}+\boldsymbol{\gamma}_{33}'\mathbf{X}_t^2+\boldsymbol{\gamma}_{34}'\mathbf{X}_{t-1}+\varepsilon_{3t}, \\ E\left(\frac{u_t^4}{\sigma_t^4}\right)=3 & \Leftrightarrow & \left(\frac{\widehat{u}_t}{\widehat{\sigma}_t}\right)^4=\gamma_{40}+\boldsymbol{\gamma}_{41}'\mathbf{X}_t+\boldsymbol{\gamma}_{42}'\boldsymbol{\Upsilon}+\boldsymbol{\gamma}_{43}'\mathbf{X}_t^2+\boldsymbol{\gamma}_{44}'\mathbf{X}_{t-1}+\varepsilon_{4t}, \end{array}\right\} t=1,2,...,n \tag{19}$$

can be accommodated in a Seemingly Unrelated Regression (SUR) framework, where the set of equations may be written as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_m \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_m \end{bmatrix} \tag{20}$$

where $\mathbf{X}_1 \neq \mathbf{X}_2 \neq \cdots \neq \mathbf{X}_m$. This condition allows the application of OLS to each equation separately to obtain the unrestricted estimators including the unrestricted variance-covariance matrix; that is, $\widehat{\mathbf{B}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and $\widehat{\boldsymbol{\Omega}}=\frac{1}{n}\widehat{\mathbf{U}}'\widehat{\mathbf{U}}$, where $\widehat{\mathbf{U}}=\mathbf{Y}-\mathbf{X}\widehat{\mathbf{B}}$. The imposition of the linear restrictions for the omnibus test takes the form of $\mathbf{D}_1\mathbf{B}+\mathbf{C}_1=\mathbf{0}$, where $\mathbf{D}_1 : p \times k$ $(p<k)$, $\text{rank}(\mathbf{D}_1)=p$ and $\mathbf{C}_1 : p \times m$ are matrices of known constants. For instance, to test whether a subset of coefficients in $\mathbf{B}$ is zero the restrictions become $\mathbf{D}_1 \equiv (\mathbf{0}, \mathbf{I}_{k_2})$, $\mathbf{B}=\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix}$, $\mathbf{C}_1=\mathbf{0}$, implying $\mathbf{B}_2=\mathbf{0}$. The constrained MLE of $\mathbf{B}$ and Ω are given by $\widetilde{\mathbf{B}}=\widehat{\mathbf{B}}-(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}_1'\left[\mathbf{D}_1(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}_1'\right]^{-1}\left(\mathbf{D}_1\widehat{\mathbf{B}}+\mathbf{C}_1\right)$, and $\widetilde{\boldsymbol{\Omega}}=\frac{1}{n}\widetilde{\mathbf{U}}'\widetilde{\mathbf{U}}=\widehat{\boldsymbol{\Omega}}+\frac{1}{n}\left(\widetilde{\mathbf{B}}-\widehat{\mathbf{B}}\right)'(\mathbf{X}'\mathbf{X})\left(\widetilde{\mathbf{B}}-\widehat{\mathbf{B}}\right)$, where $\widetilde{\mathbf{U}} \equiv \mathbf{Y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}$.

The null hypothesis is then that the restrictions hold, or that the distance $\left\|\mathbf{D}\widehat{\mathbf{B}}-\mathbf{C}\right\|$ is relatively close to zero. Defining $\widehat{G}=\left[\left(\widehat{\mathbf{U}}'\widehat{\mathbf{U}}\right)\left(\widehat{\mathbf{U}}'\widehat{\mathbf{U}}\right)^{-1}\right]$, a $m \times m$ random matrix, it is possible to use a Likelihood-Ratio type test by computing the determinant of $\widehat{G}$.

This is, $LR(\mathbf{y})$=det $\left[\left(\widehat{\mathbf{U}}'\widehat{\mathbf{U}}\right)\left(\widehat{\mathbf{U}}'\widehat{\mathbf{U}}\right)^{-1}\right]$. For large n,

$$-n^{*}\ln LR(\mathbf{y}) \underset{\alpha}{\overset{H_0}{\sim}} \chi^{2}(mp), \tag{21}$$

where $n^{*}=\left[n-k-\frac{1}{2}(m-p+1)\right]$ and α is a pre-specified significance level.

For the case when m=1, the restrictions among the parameters of $\boldsymbol{\beta}$ can be accommodated within the linear formulation $\mathbf{R}\boldsymbol{\beta}$=$r$, rank($\mathbf{R}$)=$m$, where $\mathbf{R}$ and $\mathbf{r}$ are $m\times k$ $(k>m)$ and $m\times 1$ known matrices. The null hypothesis is $H_0:\mathbf{R}\boldsymbol{\beta}$=$\mathbf{r}$ against the alternative $H_1:\mathbf{R}\boldsymbol{\beta}\neq\mathbf{r}$. To assess whether the distance $\left\|\mathbf{R}\widehat{\boldsymbol{\beta}}=\mathbf{r}\right\|$ is statistically close to zero or not, the F-type test-statistic takes the form:

$$\tau(\mathbf{y})=\frac{1}{m}\frac{\left(\mathbf{R}\widehat{\boldsymbol{\beta}}-\mathbf{r}\right)'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{R}\widehat{\boldsymbol{\beta}}-\mathbf{r}\right)}{s^{2}}\underset{\alpha}{\overset{H_0}{\sim}}F(m,n-k). \tag{22}$$

3 Beta Regression-like Models

3.1 Introduction

Of particular interest in economics and other social sciences is to model situations where the behavior of a response (dependent) random variable $\{Y_t,\ t{\in}\mathbb{N}\}$ can be modeled as a function of a set of explanatory variables $\{\mathbf{X}_t{=}X_{1,t},\ldots X_{K,t},\ t{\in}\mathbb{N}\}$. Beta regression-like models are particularly useful for dependent variables measuring *rates* or *proportions*, as well as any other random variables y whose support:

$$\mathbb{R}_Y{:=}\{y : f(y;\theta) > 0\} = [a,b], -\infty < a < b < \infty,$$

is bounded. In such cases the Beta distribution is more appropriate than the Normal in the sense that the former accounts for the measurement information pertaining to the bounded support. In addition, they provide the additional flexibility of allowing for the marginal distribution of Y_t to be asymmetric.

3.1.1 Some Properties of the Beta Distribution

For $Y_t{\sim}\mathsf{Beta}\,(a,b)$:

$$E(Y_t){=}\tfrac{a}{a+b}; Var(Y_t){=}\tfrac{ab}{(a+b)^2}\tfrac{1}{a+b+1}\ \text{for}a>0, b>0.$$

Y_t is unimodal and $Var(Y_t) < \frac{1}{12}$ for $a, b > 1$.

Y_t is U-shaped and $\frac{1}{12} < Var(Y_t) < \frac{1}{4}$ if $a, b < 1$.

If $a{=}b{=}1$, Y_t is identical to the uniform distribution.

If $a{=}b{=}\frac{1}{2}$, Y_t coincides with the arc-sine distribution.

Additionally, its skewness and kurtosis coefficients are:

$$\alpha_3{=}\tfrac{2(b-a)}{a+b+2}\sqrt{\tfrac{a+b+1}{ab}},\ \ \alpha_4{=}\tfrac{3(a+b+1)\left[2(a+b)^2+ab(a+b-6)\right]}{ab(a+b+2)(a+b+3)}.$$

Hence, a beta distribution will resemble a Normal if $a{=}b > 1$, since since $\alpha_3{=}0$ and for large values of $a{=}b$, say 24, $\lim\limits_{a=b\to\infty}\ \alpha_4| = 3$.

When one uses the conditional distribution of Y_t given $\mathbf{X}_t$ to model a regression-like model in the spirit of the Generalized Linear models (see McCullagh and Nelder, 1989), however, the modeler has to make two decisions about the functional form of the regression equation:

(a) what is the appropriate link function that contains the conditional mean in the $(0, 1)$ interval, and

(b) how the covariates in the regression equation should enter the link function. The discussion that follows contributes to addressing both questions. First, we show that by reinterpreting the values that Y_t can take as a Bernoulli distributed random variable, the logit link function is deemed the appropriate, and two, by studying the conditional distribution of Y_t given $\mathbf{X}_t$ it is possible to establish the right functional form for how the covariates enter the link function.

Section 3.2 introduces the simple beta model and evaluates its performance against the simple normal model. Section 3.3 introduces specification, estimation, inference, and misspecification of beta regression-like models using the probabilistic reduction approach. Section 3.4 introduces a discussion on the functional form of beta regression-like models, Section 3.5 presents simulation results of the performance between probabilistic beta regression-like models and simple-linear-predictor beta regression-like models, and section 3.6 summarizes the results.

3.2 Simple Beta Models

The family of Beta distributions is extremely versatile in modeling data with bounded support because one can easily extend the $(0, 1)$ interval into any other finite range using a recentering-rescaling transformation. When the support is $y \in (c, d)$, where c and d are known scalars and $c < d$, the beta distribution can be transformed into $y^* = (y - c) / (d - c)$. Notice that this transformation requires a priori knowledge of the boundaries and precludes y from taking either value. When the limits are unknown, a four-parameter beta distribution can be used instead. An additional transformation

can be applied to the data when the variable includes the boundaries, $y \in [c, d]$. The rescaling procedure involves a two step process[11]: 1) Compute y^* as indicated above and 2) Compute $y'' = [y'(n-1) + 0.5]/n$, where n is the sample size[12] (Smithson and Verkuilen, 2006).

The usefulness of the transformed Beta distribution is not limited to the range of the variable of interest but also to its ability to model non-symmetric data (very frequent in economic analysis); negatively skewed data as well as unimodal, strictly increasing, strictly decreasing, concave, convex, and uniform distributions. This flexibility encourages its empirical use in a wide range of applications (Johnson et al., 1995).

3.2.1 Specification

Before stating the properties of the simple beta model it is necessary to reparameterize the distribution. The probability distribution function of a continuous random variable y distributed beta with shape parameters a and b, that simultaneously control its skewness and dispersion, is given by

$$f(y; a, b) = \left[y^{a-1} (1-y)^{b-1} \right] / \mathsf{B}(a, b), \ 0<y<1, \ \theta := (a, b) \in \mathbb{R}_+ \times \mathbb{R}_+,$$

where $\mathsf{B}(.,.)$ denotes the Beta function. However, the probability distribution can be reparameterized in terms of its mean and dispersion parameter as,

$$f(y; \varphi) = \frac{\Gamma(\delta)}{\Gamma(\mu\delta)\Gamma((1-\mu)\delta)} y^{\mu\delta-1} (1-y)^{(1-\mu)\delta-1}, \ y \in (0, 1), \tag{23}$$

where $\mu = a/\delta$, and $\delta = a+b$. This reparameterization allows us to specify the distribution in terms of its mean and its variance via $E(Y_t) = \mu$ and $Var(Y_t) = \mu(1-\mu)/(\delta+1)$ (see Spanos, 2001). With this parameterization, which characterizes the distribution of y in a more familiar and interpretable fashion, it is possible to state the following

[11] Henceforth we will refer to these two steps as Algortihm 1.

[12] From this part on, we would use y to refer to y, y^*, and y'' indistinctively, since we will always be referring to the data (transformed if needed) in the $(0, 1)$ interval.

properties for a simple beta model for y, where conditions [1]-[5] imply that a realization from this model constitutes a random sample, where the observations are beta, independent, and identically distributed.

Table 3.1: The Simple Beta Model

SGM: $Y_t = \mu + u_t$, $t \in \mathbb{N}$	
[1] Beta	$Y_t \sim \mathsf{Beta}(.,.)$, $y \in (0,1)$
[2] Constant mean	$E(Y_t) = \mu$
[3] Constant variance	$Var(Y_t) = \frac{\mu(1-\mu)}{\delta+1} = \sigma_0^2$
[4] Independence	$\{Y_t, t \in \mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi} := (\mu, \delta)$ do not change with t
where $\boldsymbol{\varphi} := (\mu, \delta) \in (0,1) \times \mathbb{R}_+$.	

These conditions will also prove crucial for the assessment of the statistical adequacy of proposed simple beta models. The statistical Generating Mechanism (GM) is defined as an orthogonal decomposition of y into two orthogonal components, a systematic component and a non-systematic component $Y_t = E(Y_t) + u_t$, $t \in \mathbb{N}$; see Spanos (1999).

3.2.2 Estimation

For $Y_t \sim \mathsf{Beta}(\mu, \delta)$, the log-likelihood function is given by

$$l(\mu, \delta | \mathbf{y}) = \sum_{t=1}^{n} \ln f(y_t; \mu, \delta) = \ln \Gamma(\delta) - \ln \Gamma(\mu, \delta) - \ln \Gamma((1-\mu)\delta) + (\mu\delta - 1) \ln y_t + ((1-\mu)\delta - 1) \ln(1-y_t) \quad (24)$$

where $\Gamma(.)$ is the Gamma function. The maximum likelihood estimators for μ and δ can be derived as the solutions to the following score functions,

$$\frac{\partial l(\mu,\delta|\mathbf{y})}{\partial \mu} = \sum_{t=1}^{n} \frac{\partial \ln f(y_t;\mu,\delta)}{\partial \mu} = 0 \quad \text{and} \quad \frac{\partial l(\mu,\delta|\mathbf{y})}{\partial \delta} = \sum_{t=1}^{n} \frac{\partial \ln f(y_t;\mu,\delta)}{\partial \delta} = 0$$

which have to be obtained via a numerical algorithm since no closed-form solutions exist.

3.2.3 Small-Sample Properties

Little is known with respect to the sampling properties of the maximum likelihood estimators in the simple beta model besides the fact that they are biased in small samples (Gupta et al., 2004). Through simulation, Romero (2010) has determined that a good first approximation to the sampling distribution of the ML estimator of μ is given by

$$\widehat{\mu} \sim \mathsf{Beta}\left(\frac{a}{b}(a+b)\,n, (a+b)\,n\right),$$

which implies that $E(\widehat{\mu}) = a/(a+b) = \mu$ and $Var(\widehat{\mu}) = \mu(1-\mu)/[(\delta+1)(n)] = \sigma_0^2/(n)$. Although the ML estimators converge in probability to the normal; in small samples, inference conducted under the wrong probabilistic assumptions about their distribution will often lead to unreliable inferences.

Romero (2010) has also proposed a test statistic for hypotheses regarding μ of the form $H_0{:}\mu{=}\mu_0$ against $H_1{:}\mu{\neq}\mu_0$ (or $H_1{:}\mu{\lesseqgtr}\mu_0$) when δ is known. The framework uses a *natural distance* to test the discrepancy between μ and μ_0 using the score function and the fact that, if the null is true, $E[s(\mathbf{y};\mu)] \overset{\copyright}{=} 0$ and $Var[s(\mathbf{y};\mu)] = E[s^2(\mathbf{y};\mu)] = \mathcal{I}_n(\mu)$; where $s(\mathbf{y};\mu) = \frac{\partial l(\mu,\delta|\mathbf{y})}{\partial \mu}$ and $\overset{\copyright}{=}$ reads 'under the assumption of correct specification.'

For a given δ, the score function is:

$$\begin{aligned} s(\mathbf{y};\mu) &= \textstyle\sum_{t=1}^{n}\left[\delta\psi((1-\mu)\delta) - \delta\psi(\mu\delta) + \delta\ln(y_t) - \delta\ln(1-y_t)\right] = \\ &= \textstyle\sum_{t=1}^{n}\left[\delta(y_t^* - \mu^*)\right] \end{aligned}$$

where $y_t^* = \ln(y_t) - \ln(1-y_t)$, $\mu^* = \psi(\mu\delta) - \psi((1-\mu)\delta)$, and $\psi(.)$ is the digamma function. Since

$$\frac{\partial^2 \ln f(\mathbf{y};\mu)}{\partial \mu^2} = -\sum_{t=1}^{n}\left(\delta^2\left[\psi'((1-\mu)\delta) + \psi'(\mu\delta)\right]\right) = -n\delta^2\left[\psi'((1-\mu)\delta) + \psi'(\mu\delta)\right],$$

and

$$\mathcal{I}_n(\mu) = -E\left[\frac{\partial^2}{\partial \mu^2}\ln f(\mathbf{y};\mu)\middle|\,\mu\right] = n\delta^2\left[\psi'((1-\mu)\delta) + \psi'(\mu\delta)\right],$$

where $\psi'(.)$ is the trigamma function, the test-statistic is of the form:

$$\tau_B = \frac{\sum_{t=1}^{n}(y_t^* - \mu^*)}{\sqrt{n\upsilon^*}} \quad \overset{H_0}{\sim} \mathsf{N}(0,1) \tag{25}$$

where y_t^* and μ^* are defined as above, $\upsilon^* = \psi'((1-\mu)\delta) + \psi'(\mu\delta)$ and $\overset{H_0}{\sim}$ reads 'distributed under the null.'

3.2.4 Comparison with the Simple Normal Model

The fact that the simple beta model establishes the conditions to analyze beta distributed variables imposing constancy of the mean and the dispersion warrants the comparison of this model to the ubiquitous constant mean-dispersion model, the simple normal model (Table 3.2). While conditions (2)-(5) are equivalent, the distributional assumption and the unboundedness of the mean might lead the researcher to incorrect inferences even if the first four moments coincide (see Appendix 3.B).

Table 3.2: The Simple Normal Model

SGM: $Y_t = \mu + u_t$, $t \in \mathbb{N}$	
(1) Normal	$Y_t \sim \mathsf{N}(\mu, \sigma^2)$, $y \in (0,1)$
(2) Constant mean	$E(Y_t) = \mu$
(3) Constant variance	$Var(Y_t) = \sigma^2$
(4) Independence	$\{Y_t, t \in \mathbb{N}\}$ is an independent process
(5) t-homogeneity	$\boldsymbol{\varphi} := (\mu, \sigma^2)$ do not change with t
where $\boldsymbol{\varphi} := (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$.	

This is important because if a researcher ignores the naturally bounded range of the data (e.g. rates or proportions), and assumes Normality, inferences can be unreliable. To illustrate this point, we simulated six different Beta distributed random variables of different samples sizes $n = \{25, 50, 100, 500\}$. Using $10,000$ samples per random variable, we compared point maximum likelihood estimators of the mean and the variance of Y_t under both the Beta distribution and the Normal distribution. The

results are presented in Table 3.3[13].

Clearly, the ML estimators for the mean and the variance, under either distributional assumption, produce similar estimates (mean point value and empirical standard errors), if not for some rounding error. It is well-known, however, that point estimation is considered *inadequate* for the purposes of inference, because a 'good' point estimator, by itself, does not provide any measure of the reliability and precision associated with the estimate. What is important is whether the sampling distribution of this estimator, assuming Normality, is a good approximation of the true one under the Beta distribution. In particular, whether the relevant error probabilities are approximated well or not. Table 3.4 shows the discrepancy between the tail probabilities at a five percent significance level between the maximum likelihood estimators under Normality and the Beta distribution. The results indicate that the approximation is reasonable only when certain conditions of symmetry and large values of (a, b) hold.

[13]For ease of interpretation, all the results were multiplied by a factor of 100.

Table 3.3: The Simple Beta Model vs. The Simple Normal Model

	True Values μ=50, σ^2=12.5				True Values μ=50, σ^2=8.3				True Values μ=50, σ^2=5.0			
	Normal		Beta		Normal		Beta		Normal		Beta	
n	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$
25	50.0	12.5	50.0	11.8	49.9	8.3	49.9	7.84	49.9	5.0	49.9	4.7
	(7.1)	(1.9)	(6.5)	(1.5)	(5.8)	(1.5)	(5.6)	(1.3)	(4.4)	(1.1)	(4.4)	(0.9)
50	50.0	12.5	50.0	12.1	49.9	8.3	49.9	8.1	50.0	5.0	50.0	4.8
	(5.0)	(1.3)	(4.5)	(1.0)	(4.0)	(1.0)	(3.9)	(0.9)	(3.1)	(0.7)	(3.1)	(0.7)
100	49.9	12.0	49.9	12.3	50.0	8.3	50.0	8.2	50.0	5.0	50.0	4.9
	(3.5)	(0.9)	(3.1)	(0.7)	(2.8)	(0.7)	(2.7)	(0.6)	(2.2)	(0.5)	(2.1)	(0.5)
500	49.9	12.5	49.9	12.4	50.0	8.3	50.0	8.3	50.0	5.0	50.0	4.9
	(1.5)	(0.4)	(1.4)	(0.3)	(1.2)	(0.3)	(1.2)	(0.2)	(1.0)	(0.2)	(0.9)	(0.2)

	True Values μ=28.5, σ^2=2.5				True Values μ=83.3, σ^2=1.9				True Values μ=98.0, σ^2=4×10^{-02}			
	Normal		Beta		Normal		Beta		Normal		Beta	
n	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\mu}$	$\widehat{\sigma}^2$
25	28.6	2.5	28.5	2.4	83.3	2.0	83.3	1.9	98.0	4×10^{-02}	98.0	4×10^{-02}
	(3.2)	(0.7)	(3.2)	(0.6)	(2.8)	(0.7)	(2.8)	(0.6)	(0.4)	(2×10^{-02})	(0.4)	(2×10^{-02})
50	28.6	2.6	28.6	2.5	83.3	2.0	83.3	2.0	98.0	4×10^{-02}	98.0	4×10^{-02}
	(2.3)	(0.5)	(2.2)	(0.5)	(2.0)	(0.5)	(2.0)	(0.5)	(0.3)	(1×10^{-02})	(0.3)	(1×10^{-02})
100	28.6	2.6	28.6	2.5	83.3	2.0	83.4	2.0	98.0	4×10^{-02}	98.0	4×10^{-02}
	(1.6)	(0.3)	(1.6)	(0.3)	(1.4)	(0.4)	(1.4)	(0.3)	(0.2)	(1×10^{-02})	(0.2)	(8×10^{-03})
500	28.6	2.6	28.6	2.5	83.3	2.0	83.3	2.0	98.0	4×10^{-02}	98.0	4×10^{-02}
	(0.7)	(0.2)	(0.7)	(0.1)	(0.6)	(0.2)	(0.6)	(0.1)	(0.1)	(4×10^{-03})	(0.1)	(4×10^{-03})

Maximum Likelihood Estimators of the Mean and the Variance of Y assuming either a normal or a beta distribution.

Standard errors in parenthesis. All the results have been multiplied by a factor of 100.

Table 3-4 : Nominal vs. Actual Error Probabilities					
True Model: Beta // Estimated Model: Normal					
μ	σ^2	a	b	Nominal	Actual
0.500	0.125	0.5	0.5	.05	.16
0.500	0.083	1.0	1.0	.05	.10
0.500	0.050	2.0	2.0	.05	.07
0.285	0.025	2.0	5.0	.05	.08
0.833	0.019	5.0	1.0	.05	.13
0.980	0.000	50.0	1.0	.05	.16
Actual vs. Nominal Tail Probabilities at 5% under the assumption of normality.					

To illustrate the previous point, consider a Beta distributed variable with shape parameters a=0.5, b=0.5 (μ=0.5, σ^2=0.125) and with n=25 (Table 3.4). The approximation to the sampling properties of the ML estimator of μ yields $\widehat{\mu} \sim \mathsf{Beta}(25, 25)$, or $E(\widehat{\mu})$ =0.5 (or 50.0 using the factor 100) and $St.Dev.(\widehat{\mu})$ =6.804 1$\times 10^{-2}$ (or 6.8 using the factor 100), relatively close the the empirical mean and standard errors of the beta distributed ML estimators (Table 3.3). If the normal distribution was incorrectly assumed and used for inference purposes, even in the case where the estimated coefficients were identical to the true parameters, the actual error probabilities for a test like H_0: $\widehat{\mu}$=0.5 vs. H_1: $\widehat{\mu} \neq 0.5$ would be of 0.16 instead of the nominal error probability of 0.05 that a researcher would believe it is attained; the actual *type-I* error would be more than 3 times larger!

The situation would be no different with respect to the power of the test. Table 3.5 shows the results of evaluating H_0: $\widehat{\mu} \geq \mu_0$ for $n = \{10, 20, 30\}$ and $\mu_0 = \{0.5, 0.45, 0.4, 0.3, 0.2\}$ using both the Beta-warranted test-statistic proposed in (3) and an incorrect Normal-based z-statistic. Notice that, as expected, the power of the test is an increasing function of the sample size and of $\|\widehat{\mu} - \mu_0\|$. It is clear that assuming the wrong distributional assumptions (and thus the wrong statistic) will lead the

researcher to underestimate the true power of the test for practically every sample size.

Table 3.5: Power of the Test for $H_0 : \widehat{\mu} \geq \mu_0$ with (μ=0.5, δ=1)

H_1:	$\widehat{\mu} = 0.5$		$\widehat{\mu} = 0.45$		$\widehat{\mu} = 0.4$		$\widehat{\mu} = 0.3$		$\widehat{\mu} = 0.2$	
T	Beta	Normal	Beta	Normal	Beta	Normal	Beta	Normal	Beta	Normal
10	0.048	0.049	0.125	0.124	0.238	0.227	0.617	0.565	0.943	0.854
20	0.048	0.051	0.174	0.163	0.384	0.344	0.887	0.798	0.999	0.985
30	0.053	0.053	0.214	0.194	0.521	0.464	0.975	0.929	1.000	0.999

Beta indicates that the power of the test was computed using (3). Normal indicates that the power of the test was computed using the normal z-test.

3.3 Beta Regression-like Models

The basis of Regression-like models is the conditional distribution of $(Y_t|\mathbf{X}_t = \mathbf{x}_t)$, which is often reduced to the regression and skedastic functions. More frequently than not, knowledge of the joint or the conditional distribution of the relevant vector stochastic process is not available to the researcher. When this is the case, regression-like models, a variant of the generalized linear models methodology, can be implemented as long as certain conditions to assess the statistical adequacy of the proposed specifications are provided. The regression-like models framework proposed here incorporates the pseudo-generalized linear models' methodology and the probabilistic reduction approach (Spanos, 1986, 1999).

3.3.1 An Overview of the Probabilistic Reduction Approach

Let $\{Y_t,\ t\in\mathbb{N}\}$ be a stochastic process defined on a proper probability space and let $\{\mathbf{X}_t=X_{1,t},\dots X_{K,t},\ t\in\mathbb{N}\}$ be a vector stochastic process defined on the same probability space with joint density function $f_{\mathbf{x}}(\mathbf{X}_t;\boldsymbol{\phi}_2)$, where $\boldsymbol{\phi}_2$ is an appropriate set of parameters. Furthermore, assume that $E\left(\mathbf{X}_{k,t}^2\right)<\infty$ for $k\in K$, $t\in\mathbb{N}$.

The probabilistic structure of an observable vector stochastic process is fully de-

scribed by the joint distribution of $\mathbf{Z}_t$, that is, $\mathsf{D}(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n; \boldsymbol{\phi})$, for $t \in \mathbb{N}$, where $\boldsymbol{\phi}$ is a set of appropriate parameters. This distribution demarcates the relevant statistical information because it provides the most general description of the potential information contained in the data. Kolmogorov's theorem also warrants the existence of, not only the process itself, but also the 'few numerical values', the parameters $\boldsymbol{\phi}$, which will summarize or 'reduce' the statistical information contained in the process in a systematic manner. The size of the parameter set depends crucially on the invariant structure of the process.

Given that a complete description of the probabilistic structure of the vector stochastic process $\{\mathbf{Z}_t\}$ is provided by the joint distribution $\mathsf{D}(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n; \boldsymbol{\phi})$, it is possible to characterize all the statistical models in relation to $\mathbf{Z}_t$ by imposing t-invariant assumptions to its distribution from a set of testable probabilistic conditions; ergo Probabilistic Reduction (PR) approach. Different probabilistic conditions will lead to different statistical models. The different combinations yielded by the reduction assumptions can generate a wealth of statistical models that would have been impossible to construct otherwise. The reduction assumptions are obtained from three broad categories:

(D) Distribution (M) Dependence (H) Heterogeneity

Assuming that the joint vector stochastic process $\{\mathbf{Z}_t = (Y_t, \mathbf{X}_t),\ t \in \mathbb{N}\}$ is IID (conditions imposed from M and H), the joint distribution of $\mathbf{Z}_t$ can be reduced to,

$$\mathsf{D}(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n; \boldsymbol{\phi}) \overset{I}{=} \prod_{t=1}^{n} \mathsf{D}_t(\mathbf{Z}_t; \boldsymbol{\phi}_t) \overset{IID}{=} \prod_{t=1}^{n} \mathsf{D}(\mathbf{Z}_t; \boldsymbol{\phi}) \overset{IID}{=} \prod_{t=1}^{n} \mathsf{D}(Y_t, \mathbf{X}_t; \boldsymbol{\phi}) \qquad (26)$$

It is possible to decompose the resulting distribution into a conditional distribution and a marginal distribution, that is, $\mathsf{D}(Y_t, \mathbf{X}_t; \boldsymbol{\phi}) = \mathsf{D}(Y_t | \mathbf{X}_t; \boldsymbol{\phi}_1) \cdot \mathsf{D}(\mathbf{X}_t; \boldsymbol{\phi}_2)$, where $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are appropriate sets of parameters. It is important to note the role of each of the reduction assumptions and the reparameterization/restriction from the primary parameters $\boldsymbol{\phi}$ to the model parameters $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. Notice also that $\boldsymbol{\phi}_1$

and $\boldsymbol{\phi}_2$ might not necessarily be *variation free*. The last probabilistic condition of Distribution would then be imposed in this conditional/marginal decomposition. This distributional assumption will directly establish the probability distribution of both the conditional distribution and the marginal distribution. This task is not trivial. Additionally, note that if $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are *variation free*, then it is possible to impose weak exogeneity of $\mathbf{X}_t$ with respect to the parameters $\boldsymbol{\phi}_2$ one can ignore $\mathsf{D}(\mathbf{X}_t;\boldsymbol{\phi}_2)$ for inference purposes. With this imposition, the focus of the modeling endeavor relies exclusively on assessing the distributional properties of $\mathsf{D}(Y_t|\mathbf{X}_t;\boldsymbol{\phi}_1)$. For reasons that will become apparent in the sequel, weak exogeneity will be imposed in the distribution of a bivariate joint stochastic vector $\{(Y_t,\mathbf{X}_t),\, t\in\mathbb{N}\}$ by letting $\mathbb{P}(\mathbf{X}_t=\mathbf{x}_t)=1$.

3.3.2 Specification of Beta Regression Models

Distribution-based Specifications The question now is to impose a distributional structure to $\mathsf{D}(Y_t|\mathbf{X}_t;\boldsymbol{\phi}_1)$. For a beta regression model with one explanatory variable, a non-independent bivariate beta distribution suggests itself[14] as a first approximation to the joint distribution of the stochastic process $\{\mathbf{Z}_t\}$. This will allow the decomposition of the joint stochastic process into a systematic component and an unsystematic component. Incidentally, evaluating the expected value of the systematic component will lead us to the correct specification for the regression (and skedastic) function[15]. Unfortunately, there are several non-independent bivariate beta distributions in the literature that share the properties of beta distributed marginal distributions and beta distributed conditional distributions, each with different limitations in terms of the dependence between Y_t and $\mathbf{X}_t$, making the choice of distribution not a trivial problem.

The availability of several multivariate beta distributions creates an additional

[14] Since $\{\mathbf{X}_t\}$ can be rescaled to fit the K-dimensional unit interval $(0,1)^K$.

[15] 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.

level of complexity in the modeling of beta distributed data. The problem is exacerbated when deciding what beta distribution to use and the limitations of each probabilistic choice since very specific beta relationships have embedded specific ranges for their correlations, such as exclusively non-negative or exclusively non-positive correlations. This, of course, poses a monumental problem for the practitioner that might be attempting to model non-negatively correlated data with a bivariate beta distribution that only allows for non-positive correlations.

As a first approximation to the distributional choice problem, consider the bivariate beta distribution illustrated in Spanos (1999) based on Isserlis (1914) and Pearson (1923a). This distribution, while maintaining marginal and conditional probability densities in the beta family for both Y_t and X_t, has two very strict requirements: one, it can only model non-positive correlations, and two, it can model data only in the simplex index. The probability density function is given by:

$$f\left(y,x;\boldsymbol{\theta}\right)=\tfrac{\Gamma(v_1+v_2+v_3)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)}x^{v_1-1}y^{v_2-1}\left(1-x-y\right)^{v_3-1} \tag{27}$$

where $\boldsymbol{\theta}\mathbf{:=}(v_1,v_2,v_3)\in\mathbb{R}^3_+$, $x,y\geq 0$, and $x+y\leq 1$. The marginal and conditional distributions are also beta distributed (see Spanos, 1999) and $\rho_{X,Y}=-\left[\frac{v_1v_2}{(v_1+v_3)(v_2+v_3)}\right]^{1/2}\leq 0$.

The fact that the distribution can only admit non-positive correlations between Y_t and X_t translates directly not only into the regression function but also into the skedastic function,

$$E(Y|X=x)=\tfrac{v_2}{v_2+v_3}(1-x)\quad\text{and}\quad Var(Y|X=x)=\tfrac{v_2v_3}{(v_2+v_3)^2(v_2+v_3+1)}(1-x)^2\ , \tag{28}$$

since $(v_1,v_2,v_3)\in\mathbb{R}^3_+$.

Olkin and Liu (2003) derived a bivariate beta distribution with exactly the opposite problem, although relaxing the unit simplex condition. First, they noticed that the bivariate beta distribution generated from the Dirichlet distribution has the support on the simplex $0\leq x,y\leq 1$ with $0\leq x+y\leq 1$. By specifying x and y as unconditionally beta distributed variables and by linking them together via a gamma distributed third variable W_t, they were able to derive a bivariate beta distribution

for X_t and Y_t using the fact that each beta distributed variables is the result of a non-linear combination of gamma variates. The resulting bivariate beta distribution is given by,

$$f(y,x;\boldsymbol{\theta}) = \frac{\Gamma(v_1+v_2+v_3)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)} \frac{(1-x)^{v_2+v_3-1}(1-y)^{v_1+v_3-1}}{(1-xy)^{v_1+v_2+v_3}} x^{v_1-1}y^{v_2-1} \tag{29}$$

where $\boldsymbol{\theta} := (v_1, v_2, v_3) \in \mathbb{R}^3_+$, and $0 \leq x, y \leq 1$. Although the density preserves the three-parameter characteristic of the one illustrated in Spanos (1999), it allows the random variables X_t and Y_t to have only the restriction of beings simultaneously bounded between zero and one without having to belong in the unit simplex. Their marginal and conditional distributions belong to the standard beta family of distributions.

The caveat of this distribution is that both, the correlation coefficient and the regression function, involve the Generalized Hypergeometric Function and cannot be expressed in closed form. Additionally, the correlation coefficient of this density is non-negative and bounded at $[0, 1]$ (Olkin and Liu, 2003).

Arnold et al., (1999) eliminated the problems of unit index support and specific correlations by creating bivariate distributions not based on a priori marginals but rather based on the conditional distributions of Y_t given X_t and X_t given Y_t. Under this approach, the bivariate density yields beta conditional and marginal distributions. Additionally, both forms of correlation can be generated through their conditionally-beta distributions; a suitable asset for regression analysis. Unfortunately, the resulting non-standard bivariate distribution requires nine different parameters for its specification that have to simultaneously satisfy a very stringent set of conditions.

The bivariate density with beta conditionals is given by,

$$\begin{aligned} f(y,x;\boldsymbol{\theta}) &= [x(1-x)y(1-y)]^{-1} \exp[m_{11}\ln x \ln y + m_{12}\ln x \ln(1-y) + \\ &\quad + m_{21}\ln(1-x)\ln y + m_{22}\ln(1-x)\ln(1-y) + m_{10}\ln x + \\ &\quad + m_{20}\ln(1-x) + m_{01}\ln y + m_{02}\ln(1-y) + m_{00}] \end{aligned} \tag{30}$$

$$\text{for } 0 \leq x, y \leq 1; \tag{31}$$

$$
\begin{aligned}
&m_{10}+m_{11}\ln y+m_{12}\ln(1-y)>0; \quad m_{20}+m_{21}\ln y+m_{22}\ln(1-y)>0;\\
&m_{01}+m_{11}\ln x+m_{21}\ln(1-x)>0; \ \ m_{02}+m_{12}\ln x+m_{22}\ln(1-x)>0;\\
&m_{10}>0; m_{20}>0; m_{01}>0; m_{02}>0; \ \{m_{ij}\leq 0,\ i=1,2;\ j=1,2\}\\
&e^{m_{00}}=1/\iint_{D_1\times D_2} e^{-m_{00}} f\,(y,x;\boldsymbol{\theta})\,d\mu_1(x)d\mu_2(y)<\infty.
\end{aligned}
$$

The latter term (that must be computed numerically) along with the conditions on the parameters ensure integrability and properness of the distribution. The resulting regression function, after some manipulation, can be written as

$$
E(Y_t|X_t=x_t)=\frac{\beta_0+\beta_1\ln x_t+\beta_2\ln(1-x_t)}{(\alpha_0+\beta_0)+(\alpha_1+\beta_1)\ln x_t+(\alpha_2+\beta_2)\ln(1-x_t)} \tag{32}
$$

where $0<E(Y_t|X_t=x_t)<1$.

The methodology proposed by Arnold et al., (1999) stands as a monumental attempt to include the probability distribution of the variables involve in the conditioning set into the derivation of the regression function. Their treatment on bivariate relationships of probability distribution in the exponential family allowed us to derive the following regression specifications for a number of probability distributions,

$$
E\,(Y_t|X_t=x_t)=\begin{cases}
\frac{\beta_0+\beta_1x_t+\beta_2x_t^2}{(\alpha_0+\beta_0)+(\alpha_1+\beta_1)x_t+(\alpha_2+\beta_2)x_t^2} & \text{if } (X_t|Y_t=x_t) \text{ is normal}\\
\frac{\beta_0+\beta_1x_t+\beta_2\ln(x_t)}{(\alpha_0+\beta_0)+(\alpha_1+\beta_1)x_t+(\alpha_2+\beta_2)\ln(x_t)} & \text{if } (X_t|Y_t=x_t) \text{ is gamma}\\
\frac{\beta_0+\beta_1x_t+\beta_2}{(\alpha_0+\beta_0)+(\alpha_1+\beta_1)x_t} & \text{if } (X_t|Y_t=x_t) \text{ is exponential}\\
\frac{\beta_0+\beta_1x_t+\beta_2x_t^{-1}}{(\alpha_0+\beta_0)+(\alpha_1+\beta_1)x_t+(\alpha_2+\beta_2)x_t^{-1}} & \text{if } (X_t|Y_t=x_t) \text{ is inverse Gaussian}
\end{cases} \tag{33}
$$

In principle, it is possible to extend the previous analysis to multivariate conditioning sets as long as the variables belong to a prescribed family of densities. As it can be inferred, the resulting multivariate distribution would pose an even greater challenge for the modeler since not only the number of estimable parameters will increase, but also the number of restrictions in the estimable parameters will increase as well. Furthermore, the normalizing constant problem would have to be re-evaluated and readdressed at each stage.

If we impose an assumption of independence between the variables in the conditioning set, then the expansion of the regression functions would simply imply adding as many more explanatory variables with the right functional form as needed. For instance, letting Y_t be conditionally beta distributed on X_{1t} and X_{2t} and assuming $(X_{1t}|Y_t{=}y_t) \sim N(.,.)$ and $(X_{2t}|Y_t{=}y_t) \sim \Gamma(.,.)$. Then,

$$E\left(Y_t|X_t{=}x_t\right) = \frac{\beta_0{+}\beta_1 x_{1t}{+}\beta_2 x_{1t}^2{+}\beta_3 x_{2t}{+}\beta_4 \ln\left(x_{2t}\right)}{\left(\alpha_0{+}\beta_0\right) + \left(\alpha_1{+}\beta_1\right) x_{1t}{+}\left(\alpha_2{+}\beta_2\right) x_{1t}^2{+}\left(\alpha_3{+}\beta_3\right) x_{2t}{+}\left(\alpha_4{+}\beta_4\right) \ln\left(x_{2t}\right)} \tag{34}$$

Regression-like Specifications Ferrari and Crebari-Neto (2004) proposed a model where a response variable Y_t is measured continuously on the standard unit interval and is beta distributed with heterogeneous location and shape parameters. The source of heterogeneity in these parameters is modeled by imposing a regression-like structure in a mean-dispersion reparameterization of the heterogeneous distribution identical to Spanos, 2001.

The mean-dispersion reparameterization of a heterogeneous beta distributed variable is given by

$$f\left(y_t;\boldsymbol{\varphi}\right) = \frac{\Gamma\left(\delta\right)}{\Gamma\left(\delta\mu_t\right)\Gamma\left(\delta\left(1{-}\mu_t\right)\right)} y_t^{\delta\mu_t - 1}\left(1{-}y_t\right)^{\delta(1-\mu_t)-1},\ y_t \in (0,1) \tag{35}$$

where $E(Y_t){=}\mu_t$, and $Var\left(Y_t\right) = \mu_t\left(1{-}\mu_t\right)/\left(\delta+1\right)$. Notice that, aside from the fact that $E(Y_t){=}\mu_t$ is now heterogeneous, the distribution resembles that of (23). The corresponding model specification is given in Table 3.6.

Table 3.6: The Heterogeneous Beta Model	
SGM:	$Y_t = \mu_t + u_t,\ t \in \mathbb{N}$
[1] Beta	$Y_t \sim \mathsf{Beta}(.,.),\ y \in (0,1)$
[2] Heterogeneous mean	$E(Y_t) = \mu_t$
[3] Heterogeneous variance	$Var(Y_t) = \frac{\mu_t(1-\mu_t)}{\delta+1}$
[4] Independence	$\{Y_t,\ t \in \mathbb{N}\}$ is an independent process
where $(\mu_t, \delta) \in (0,1) \times \mathbb{R}_+$.	

As it stands, this model is not operational until the modeler establishes a functional form between the heterogeneous expected value of the dependent variable Y_t and a set of explanatory variables $\mathbf{X}_t$ $(k \times 1)$. In other words, our purpose is to establish some form of mean heterogeneity in y due to x. If it is possible to describe the heterogeneity in the mean as a function of the explanatory variables, i.e., $\mu_t = h(\mathbf{x}_t)$, then the previous model will become operational and can be stated as a regression-like model (Table 3.7),

Table 3.7: The Beta Regression-like Model	
SGM:	$Y_t = \mu_t + u_t,\ t \in \mathbb{N}$
[1] Beta	$Y_t \sim \mathsf{Beta}(.,.),\ y \in (0,1)$
[2] Heterogeneous mean	$E(Y_t) = \mu_t = h(\mathbf{x}_t)$
[3] Heterogeneous variance	$Var(Y_t) = \frac{h(\mathbf{x}_t)(1-h(\mathbf{x}_t))}{\delta+1}$
[4] Independence	$\{Y_t,\ t \in \mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi} := (\boldsymbol{\beta}, \delta)$ do not change with t
where $(\mu_t, \delta) \in (0,1) \times \mathbb{R}_+$.	

As it stands this model is non-operational until we are able to specify and estimate a set of parameteres that describe $\mu_t = h(\mathbf{x}_t; \boldsymbol{\beta})$ while ensuring $\mu_t \in (0,1)$. A potential solution would be to use multivariate CDF that would warrant $h(.) : \mathbb{R}^m \to [0,1]$. In practive, however, this has proven a monumental task and practitioners opted for way to overcome the multidimensionality of the CDF, first in binary and multnomial mod-

els and then to their continuous form extension represented by the beta regression-like models. This methodology gives birth to th class of generalized linear models except for the fact that the beta distribution assumption for the dependent variable does not belong to the exponential family of distributions. The methodology, than can be considered *pass the bucket* approach, consists in reducing the multidimensionality of $h(\mathbf{x}_k;\boldsymbol{\beta}),\ k\in\mathbb{N}$ to a univariate CDF $\mu_k = F(\mathbf{x}_k;\boldsymbol{\beta}),\ k\in\mathbb{N}$, where $F(.){:}\mathbb{R}_Z \to [0,1]$ by using linear combinations of the explanatory variables $z_k = \boldsymbol{\beta}^\top \mathbf{x}_k,\ \mathbf{x}_k \in \mathbb{R}^m$.

Nevertheless, $h(.)$ sheds no light as to what functional form should be imposed. Ferrari and Cribari-Neto (2004) follow the functional forms and implementation commonly used in generalized linear models analysis. These imply the use of the so-called 'link functions' (McCullagh and Nelder, 1989) in the heterogeneous mean to allow the covariates, or explanatory variables, to enter linearly in the model, that is, $h(\mathbf{x}_t) = g^{-1}(\eta_t)$, where $g^{-1}(.)$ is the link function; $\eta_t = \sum_{i=1}^{K}\beta_i x_{ti}$ is the linear predictor à la Nelder and Wedderburn, 1972; and K is the number of covariates. The choice of link function, left to the modeler, usually involves the probit link function: $h(\mathbf{x}_t) = \Phi^{-1}(\eta_t)$ where $\Phi^{-1}(.)$ is the cumulative distribution function of a standard normal random variable; the complementary log-log link function: $h(\mathbf{x}_t) = \ln[-\ln(1-\eta_t)]$; the log-log link function: $h(\mathbf{x}_t) = -\ln[-\ln(\eta_t)]$; amongst others. By far, the most commonly used link function is the logit, which establishes the sources of heterogeneity of the mean to the covariates as $h(\mathbf{x}_t) = \frac{e^{\eta_t}}{1+e^{\eta_t}}$. All these functional forms ensure that the conditional mean be bound in the unit interval. With respect to in what way the covariates have to enter the link function there is not much to be drawn from the model assumptions of the beta regression-like model. It is indeed a common practice that they enter linearly given that any other sources of nonlinearities are captured through the logit link function[16].

[16]The assumption of nonlinearity in the arguments of the link function has been relaxed to allow for additional non-linearities. Simas et al. (2008), allowed the regression structure to be nonlinear, similar to the way the exponential family nonlinear models extend the generalized linear models. For instance, they estimate a non-linear model of the form $\mathsf{logit}(\mu_t) = \beta_0 + \beta_1 x_{it} + \beta_2 x_{2t}^{\alpha_3}$. Additionally,

The dispersion parameters δ is generally assumed to enter as a constant, letting the entire source of variance heterogeneity to be completely captured by the mean. The assumption of constancy, perhaps misinterpreted as an imposition of homoskedasticity in the beta regression-like model, has been relaxed by Simas et al., 2008. They allow the dispersion parameter to have a pseudo-regression nonlinear functional form depending on either $\mathbf{X}_t$ or some other vector of explanatory variables $\mathbf{Z}_t$ that might or might not include $\mathbf{X}_t$[17].

3.3.3 Estimation

The mean-dispersion parameterization of the regression-like model warrants the following log-likelihood function,

$$\begin{aligned} l(\boldsymbol{\beta},\delta) &= \sum\nolimits_{t=1}^{n} l_t(\mu_t,\delta) = \ln\Gamma(\delta) - \ln\Gamma(\mu_t,\delta) - \ln\Gamma((1-\mu_t)\delta) + (\mu_t\delta - 1)\ln y_t + \\ &\quad + ((1-\mu_t)\delta - 1)\ln(1-y_t) \end{aligned} \tag{36}$$

where $\eta_t = \sum_{i=1}^{K} \beta_i x_{ti}$ and $\Gamma(.)$ is the Gamma function. From this log-likelihood function, the maximum likelihood estimators for β_i and δ can be derived as the solutions to the following score functions,

$$\frac{\partial l(\boldsymbol{\beta},\delta)}{\partial \beta_i} = \sum_{t=1}^{n} \frac{\partial l_t(\mu_t,\delta)}{\partial \mu_t}\frac{\partial \mu_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \beta_i} = 0 \quad \text{and} \quad \frac{\partial l(\boldsymbol{\beta},\delta)}{\partial \delta} = \sum_{t=1}^{n} \frac{\partial l_t(\mu_t,\delta)}{\partial \delta} = 0$$

Unfortunately, no closed-form solutions exists for $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\beta}}, \widehat{\delta}\right)$. To obtain the ML estimators, it is required to solve the previous system of equations simultaneously using a numerical optimization algorithm and setting $l\left(\widehat{\boldsymbol{\theta}}\right) = 0$, which is not a linear function of $\boldsymbol{\theta}$. There are several numerical algorithms which can be used to solve this problem, the one employed in this document is Matlab's Netwon-Raphson constrained optimization method.

---

they derived first and second order conditions for the estimation of this kind of non-linear models.

[17] Simas et al., (2008) entertained a logarithmic function with non-linear covariates of the form $\ln(\delta_t) = \gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{2t}^{\gamma_3}$ in their analysis.

3.3.4 Inference

Under the right specification, it is well known that the ML Estimators of $\boldsymbol{\theta}$ enjoy the following properties, amongst others,

[1] Consistency: $p\lim(\widehat{\boldsymbol{\theta}})=\boldsymbol{\theta}$

[2] Asymptotic Normality: $\widehat{\boldsymbol{\theta}}\overset{a}{\sim}\mathsf{N}\left(\boldsymbol{\theta},\mathbf{I}^{-1}\left(\boldsymbol{\theta}\right)\right)$, where $\mathbf{I}^{-1}\left(.\right)$ is the inverse of the Fisher Information Matrix.

The Observed Information Matrix evaluated at $\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}$ is given by,

$$I_n\left(\widehat{\boldsymbol{\theta}}\right)=\begin{bmatrix}\frac{\partial^2 l}{\partial\beta_1^2} & \frac{\partial^2 l}{\partial\beta_1\partial\beta_2} & \cdots & \frac{\partial^2 l}{\partial\beta_1\partial\beta_K} & \frac{\partial^2 l}{\partial\beta_1\partial\delta}\\ \frac{\partial^2 l}{\partial\beta_2\partial\beta_1} & \ddots & \cdots & \cdots & \vdots\\ \vdots & \vdots & \ddots & \cdots & \vdots\\ \frac{\partial^2 l}{\partial\beta_K\partial\beta_1} & \vdots & \vdots & \frac{\partial^2 l}{\partial\beta_K^2} & \vdots\\ \frac{\partial^2 l}{\partial\delta\partial\beta_1} & \cdots & \cdots & \cdots & \frac{\partial^2 l}{\partial\delta^2}\end{bmatrix}, \tag{37}$$

where

$$\frac{\partial^2 l}{\partial\delta\partial\beta_i}=\frac{\partial^2 l}{\partial\beta_i\partial\delta}=\sum_{t=1}^{n}\left[\left(y_t^*-\lambda_t\right)-\delta\left(\psi'\left(\mu_t\delta\right)\mu_t-\psi'\left[\left(1-\mu_t\right)\delta\right]\left(1-\mu_t\right)\right)\right]\frac{\partial\mu_t}{\partial\eta_t}\frac{\partial\eta_t}{\partial\beta_i},$$
$$\frac{\partial^2 l}{\partial\beta_i\partial\beta_j}=\sum_{t=1}^{n}\delta^2\left(\psi'\left(\mu_t\delta\right)+\psi'\left(\left(1-\mu_t\right)\delta\right)\right)\left(\frac{\partial\mu_t}{\partial\eta_t}\right)^2\frac{\partial\eta_t}{\partial\beta_i}\frac{\partial\eta_t}{\partial\beta_j},\ \lambda_t=\left(\psi\left(\mu_t\delta\right)-\psi\left(\left(1-\mu_t\right)\delta\right)\right)$$
$$\frac{\partial^2 l}{\partial\delta^2}=-\sum\left[\psi'\left(\mu_t\delta\right)\mu_t^2+\psi'\left[\left(1-\mu_t\right)\delta\right]\left(1-\mu_t\right)^2-\psi'\left(\delta\right)\right],\ y_t^*=\ln\frac{y_t}{1-y_t},$$

where $\psi(.)$ is the digamma function and $\psi'(.)$ is the trigamma function[18].

Several inference procedures can be conducted once the estimators and their asymptotic errors have been computed. For instance, to test $\beta_i=0$, a simple F-test of the form $F=\widehat{\beta}_i^2/\mathsf{Var}(\widehat{\beta}_i)\sim F\left(1,n-k\right)$ can be conducted. Similarly, confidence intervals can be established by using the same asymptotic errors, that is $\beta_i\pm t_{(\alpha,n-k)}\times\mathsf{s.e.}\left(\widehat{\beta}_i\right)$. For simultaneous hypothesis, it is possible to use, with large samples, a Likelihood Ratio test $\mathsf{LR}=\boldsymbol{\lambda}=L(\widetilde{\boldsymbol{\theta}})/L(\widehat{\boldsymbol{\theta}})$ where $L(\widetilde{\boldsymbol{\theta}})$ is the likelihood evaluated at the restricted estimates, $L(\widehat{\boldsymbol{\theta}})$ is the likelihood evaluated at the unrestricted maximum likelihood estimates, and $-2\ln\lambda\sim\chi^2\left(q\right)$, where q is the number of restrictions imposed.

[18] $\psi(x)=\frac{d}{dx}\ln\Gamma\left(x\right)=\frac{\Gamma'(x)}{\Gamma(x)}$ and $\psi'(x)=\frac{d^2}{dx^2}\ln\Gamma\left(x\right)$.

Additionally, a naive measure of the relative goodness of fit can be obtained by using the square of the Spearman correlation coefficient between the observed values of y_t and the estimated values $\widehat{y}_t$.

3.3.5 Misspecification Testing

We follow the misspecification testing procedure (M-S) described in (Romero, 2010), where the goal is to determine whether the assumptions of the model are valid vis-a-vis the data. The rationale behind M-S testing is to probe outside the boundaries of pre-specified models by testing H_0: $f_0(z) \in \mathcal{M}$ vs. $\overline{H}_0$: $f_0(z) \in (\mathcal{P}-\mathcal{M})$, where $\mathcal{P}$ denotes the set of all possible statistical models[19]. Detection of departures from the null in the direction of $\mathcal{P}_1 \subset (\mathcal{P}-\mathcal{M})$ can be considered sufficient to deduce that the null is false but not to deduce that $\mathcal{P}_1$ is true.

If indeed the proposed specification has been able to capture all the systematic information in the data through $h(\mathbf{X}_t)$, then any other function of the conditioning set $g(\mathbf{X}_t)$ will cause the following condition to hold[20]:

$$E([y_t - h(\mathbf{X}_t)]\, g(\mathbf{X}_t)) = 0,\ t \in \mathbb{N}. \tag{38}$$

The condition is referred to as the orthogonality expectation theorem. Using the standardized residuals, ancillary regressions of the form:

$$\widehat{v}_t = [y_t - g(\mathbf{X}_t)], \quad t \in \mathbb{N}, \tag{39}$$

where $\widehat{v}_t = \frac{\sqrt{n}(y_t - \widehat{y}_t)}{\widehat{\sigma}}$, $\widehat{\sigma} = \sqrt{Var(y_t | \mathbf{X}_t = \mathbf{x}_t)}$, can be estimated and used to assess deviations from the statistical model assumptions.

To assess the reliability of residual-based M-S testing and the precision of estimation under the TA and the PRA, a set of ancillary regressions will be used from $\widehat{y}_t = E(y_t | \mathbf{X}_t = \mathbf{x}_t)$, $\widehat{\sigma}^2 = Var(y_t | \mathbf{X}_t = \mathbf{x}_t)$, and $\widehat{u}_t = y_t - \widehat{y}_t$, to test for statistical model

[19] On how M-S testing differs from Neyman-Pearson testing see Spanos (1999), ch. 14.

[20] For the regression function, $g(\mathbf{X}_t) = E(y_t | \sigma(\mathbf{X}_t))$, where $\sigma(\mathbf{X}_t)$ denotes the σ-field generated by $\mathbf{X}_t$.

departures in the two conditional moments, as follows:

$$\left.\begin{array}{l}\left(\frac{\widehat{u}_t}{\widehat{\sigma}}\right)=\gamma_{10}+\boldsymbol{\gamma}'_{11}\mathbf{X}_t+\boldsymbol{\gamma}'_{12}\boldsymbol{\xi}_t+\boldsymbol{\gamma}'_{13}\mathbf{X}_t^2+\boldsymbol{\gamma}'_{14}\mathbf{X}_{t-1}+\varepsilon_{1t}, \\ \left(\frac{\widehat{u}_t}{\widehat{\sigma}}\right)^2=\gamma_{20}+\boldsymbol{\gamma}'_{21}\mathbf{X}_t+\boldsymbol{\gamma}'_{22}\boldsymbol{\xi}_t+\boldsymbol{\gamma}'_{23}\mathbf{X}_t^2+\boldsymbol{\gamma}'_{24}\mathbf{X}_{t-1}+\varepsilon_{2t},\end{array}\right\} t=1,2,...,n \qquad (40)$$

where $\mathbf{X}_t$ is the vector of regressors of the original specification, $\boldsymbol{\xi}_t$ is a vector of trends that capture structural change misspecification, $\mathbf{X}_t^2$ is a vector of monotonic transformations of $\mathbf{X}_t$ that allows the conditional standardized moment to have additional sources of nonlinearities, and $\mathbf{X}_{t-1}$ is a vector or lagged values of $\mathbf{X}_t$ and Y_t that allows for temporal or spatial dependence.

The previous system of equations is tested simultaneously for departures from the model assumptions, that is, $\boldsymbol{\gamma}_{(.0)}=\mathbf{0}$, using the standardized residuals. The joint test is conducted using a Multivariate Normal Linear Regression Framework.

3.4 A Digression on the Functional Form

It is clear that when the whole probabilistic structure of the data is taken into account, the functional form for the regression and the skedastic functions are given directly by evaluating the first and second moments of the conditional distribution of Y_t given $\mathbf{X}_t$. Unfortunately, too many restrictions would have to be imposed in the estimation routine to ensure the appropriate parameter space. In the previous equations(32-34), for instance, the value of the conditional mean is not necessarily bound between zero and one unless the restrictive conditions of the parameters hold. This problem is not shared by the regression-like models, where the value of the conditional mean is naturally bound between zero and one. The caveat of this approach, however, is that no additional information is provided on how the covariates should enter the regression function by the probabilistic structure of the data other than the linear covariates and/or arbitrary functional forms provided by the researcher's intuition or judgement.

Some light can be shed on what link function to use and how the covariates should enter the function if we re-interpret the support of Y_t as a set of the proportions of

success for a particular characteristic à la Bernoulli. Let Y_t represent the rate of success of a random variable that is binomially distributed. This necessarily implies that, for every t, the percentage of success of a particular characteristic is given by $Y_t = (Y_{it}=1)/n$, In principle,

$$\mu_t = h(\mathbf{x}_t) \approx E\left(\tfrac{Y_{it}}{n} | X_t = x_t\right) = \tfrac{1}{n} E(Y_{it} | X_t = x_t)$$

where Y_{it} is then a Bernoulli distributed random variable with $E(Y_i | X_t = x_t) = p_i$. It is possible to demonstrate that this reinterpretation of the dependent variable warant the used of the logit link function.

Let $\{Y_i, i=1, \ldots N\}$ be a stochastic process defined on a proper probability space where $Y_i \tilde{} \mathsf{bin}(0,1)$. Furthermore, let $\{\mathbf{X}_i = (X_{1,i}, \ldots X_{K,i}), i=1, \ldots N\}$ be a vector stochastic process defined on the same probability space with joint density function $f(\mathbf{X}_i; \psi_2)$ where ψ_2 is an appropriate set of parameters.

The joint density function of the vector stochastic process $\{(Y_i, \mathbf{X}_i), i=1, \ldots N\}$ takes the form $f(Y_1, \ldots Y_N, \mathbf{X}_1, \ldots \mathbf{X}_N; \phi)$ where ϕ is an appropriate set of parameters.

Assuming that the joint vector stochastic process is independent (I) and identically distributed (ID), the joint distribution can be reduced to

$$f(Y_1, \ldots Y_N, \mathbf{X}_1, \ldots \mathbf{X}_N; \phi) \overset{IID}{=} \prod_{i=1}^{N} f(Y_i, \mathbf{X}_i; \varphi) \overset{IID}{=} \prod_{i=1}^{N} f(Y_i | \mathbf{X}_i; \psi_1) f(\mathbf{X}_i; \psi_2)$$

where ψ_1 and ψ_2 are appropriate sets of parameters.

The existence of $f(Y_i, \mathbf{X}_i)$ is dependent upon the compatibility of the conditional density functions, $f(Y_i | \mathbf{X}_i; \psi_1)$ and $f(\mathbf{X}_i | Y_i; \varkappa_1)$, where $\varkappa_1$ is an appropriate set of parameters (Arnold and Castillo, 1999). Since

$$f(Y_i | \mathbf{X}_i; \psi_1) f(\mathbf{X}_i; \psi_2) = f(\mathbf{X}_i | Y_i; \varkappa_1) f(Y_i; p) = f(Y_i, \mathbf{X}_i; \varphi)$$

where $f(Y_i; p) = p^{Y_i}(1-p)^{1-Y_i}$, the following relationship can be derived,

$$\frac{f(\mathbf{X}_i | Y_i=1; \varkappa_1)}{f(\mathbf{X}_i | Y_i=0; \varkappa_1)} \cdot \frac{f(Y_i=1; p)}{f(Y_i=0; p)} = \frac{f(Y_i=1 | \mathbf{X}_i; \psi_1)}{f(Y_i=0 | \mathbf{X}_i; \psi_1)} \cdot \frac{f(\mathbf{X}_i; \psi_2)}{f(\mathbf{X}_i; \psi_2)} \tag{41}$$

Furthermore, assume that $f\left(Y_i|\mathbf{X}_i;\psi_1\right)$ is a conditional Bernoulli density function with the following functional form:

$$f\left(Y_i|\mathbf{X}_i;\psi_1\right)=h\left(\mathbf{X}_i;\psi_1\right)^{Y_i}\left[1-h\left(\mathbf{X}_i;\psi_1\right)\right]^{1-Y_i} \tag{42}$$

where $h\left(\mathbf{X}_i;\psi_1\right):\mathbb{R}^K\times\Theta_1\rightarrow(0,1)$.

Substituting (42) into (41) and letting $\pi_j=p^j\left(1-p\right)^{1-j}$ for $j=0,1$ gives

$$h\left(\mathbf{X}_i;\psi_1\right)=\frac{\pi_1\cdot f\left(\mathbf{X}_i|Y_i=1;\varkappa_1\right)}{\pi_0\cdot f\left(\mathbf{X}_i|Y_i=0;\varkappa_1\right)+\pi_1\cdot f\left(\mathbf{X}_i|Y_i=1;\varkappa_1\right)}$$

Using the transformation $x=\exp\left\{\ln\left(x\right)\right\}$ and rearranging (see Kay and Little, 1987), $h\left(\mathbf{X}_i;\psi_1\right)$ becomes,

$$h\left(\mathbf{X}_i;\psi_1\right)=\frac{\exp\left\{\eta\left(\mathbf{X}_i;\varkappa_1\right)\right\}}{1+\exp\left\{\eta\left(\mathbf{X}_i;\varkappa_1\right)\right\}}$$

where $\eta\left(\mathbf{X}_i;\varkappa_1\right)=\ln\frac{f(\mathbf{X}_i|Y_i=1;\varkappa_1)}{f(\mathbf{X}_i|Y_i=0;\varkappa_1)}+\kappa$ and $\kappa=\ln\left(\pi_1\right)-\ln\left(\pi_0\right)$.

Notice that the composite function represents the logistic cumulative density function. This allows us to rewrite these probabilities of success as the logarithm of the odds of success ratio. That is:

$$\ln\left[\frac{h(\mathbf{x}_t)}{1-h(\mathbf{x}_t)}\right]=\eta_t.$$

Thus, μ_t can be rewritten as:

$$h\left(\mathbf{x}_t\right)=\frac{e^{\eta_t}}{1+e^{\eta_t}}.$$

Then clearly, a naturally derived link function is the logit link function, where while modeling of the logarithm of the odds of success ratio we guarantee that $h\left(\mathbf{x}_t\right)$ maintains the right support.

This re-interpretation also opens a plethora of opportunities for the modeling of the $h(.)$ function and help establish how the covariates should enter the function. Bergtold et al., (2005), in their study on the probability based functional forms for logit models have established the conditions to determine the right specification for

the argument of the link function. They have argued that $h(.)$ can be modeled as the logarithm of the conditional distribution of X_t given that Y_{it}=1 was a success over the conditional density of X_t given that Y_{it}=0 was a failure,

$$\eta\left(X;\eta\right)=\ln\left(\frac{f_{X|Y=\text{Success};\varkappa_1}}{f_{X|Y=\text{Failure};\varkappa_1}}\right)+\kappa$$

where $\kappa=\ln\left(\pi_s/\pi_f\right)$ is the logarithm of the odds of success over failure.

Of course, in practice, we do not have information of the success or of the failures but rather on the proportion of both; yet, the study of the conditional distribution of the conditioning set given Y_t reflects similarly the underlying probability distribution of the covariates. This is clear when noticing that the functional forms provided by this approach produce the same functional transformation for the covariates than the one provided via regression models using the appropriate conditional distributions. As an illustration, consider the case where (Y_t, X_t) is jointly beta distributed with both conditional and marginal beta distributions. Then, from Bergtold et al. (2005), the argument of the logit link function is[21]

$$\eta_t=\beta_0+\beta_1\ln\left(x_t\right)+\beta_2\ln\left(1-x_t\right).$$

Notice that, compared to (32), this functional form is more parsimonious and estimates a smaller number of parameters while ensuring $\mu_t\in(0,1)$.

3.5 Simulation Analysis

To assess whether knowledge of the probability distribution of the explanatory variables helps the researcher produce an orthogonal decomposition between the systematic and the unsystematic components, we designed four different two-variable experiments. In each instance, the dependent variable Y_t is beta distributed and X_t is either beta distributed, gamma distributed, or *beta-induced* distributed from

[21] For a more detailed list of conditional distributions for $\mathbf{X}_t|Y_t=y_t$ and their respective functiona forms see Bergtold (2004) pp. 20-21.

a gamma distribution (see below). We compared the performance of two different specifications: a 'naive specification' using the logit-link where the covariates enter linearly and one 'probabilistic specification,' also using the logit-link function, where the covariates enter with the appropriate transformation granted by the probability distribution of X_t, following Bergtold et al. (2005). The appropriateness of each specification will be evaluated via the M-S testing framework described above.

3.5.1 Data Generation Process

To avoid any bias in the data generation process, we simulated R=10,000 samples of n= $\{25, 50, 100, 500\}$ from a Gaussian copula with two different correlation coefficients ρ= $\left\{-0.77, -0.8\dot{6}\right\}$, and corresponding values for the $\widetilde{R}^2$= $\{0.60, 0.75\}$ From the raw-data, selected inverse functions from the CDF of both Y_t and X_t were applied. The used probability distributions and selected $\widetilde{R}^2$'s are shown in Table 3.8. All the simulations and computations were conducted using Matlab R2009b.

Table 3.8: Experimental Design

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Y_t	$\sim$Beta $(5, 2)$	$\sim$Beta $(1, 3)$	$\sim$Beta $(1.5, 5)$	$\sim$Beta $(1.5, 5)$
X_t	$\sim$Beta $(1, 5)$	$\sim$Beta $(3, 1)$	$\sim\Gamma(2, 2)$	$\sim$Beta $(\Gamma(2, 2))$
$\widetilde{R}^2$	0.60	0.75	0.60	0.60

It is important to note that with respect to Experiment 4, $X_t\sim$Beta $(\Gamma(2, 2))$ indicates that X_t is the result of inducing a beta distribution transformation on gamma distributed data by applying the algorithm shown in Section 3.2.

3.5.2 Empirical Results

Experiment 1 For illustration purposes, Panel 3.A shows a realization of Experiment 1 with n=5,000. Skewness and non-normality of the marginal distributions of y_t and x_t is evident through the histograms. The scatter-plot underneath illustrates the the non-linear and non-normal relationship between y_t and x_t.

Specification of the naive and the probabilistic model is given by

$$\boxed{\text{Naive: } \eta_t=\beta_0+\beta_1 x_t} \quad \boxed{\text{Probabilistic: } \eta_t=\beta_0+\beta_1 \ln(x_t)+\beta_2 \ln(1-x_t)}, \tag{43}$$

where knowledge of the marginal probability distribution of X_t allows us to incorporate additional sources of nonlinearities in the regression-like function. The full model assumptions are presented in Table 3.9.

The results are presented in Table 3.12. The table shows the mean value and empirical standard error of the ML estimators of both specifications as well as the mean and empirical standard error of the coefficient of determination. It also shows the mean F-statistic of the significance of each estimator (including the coefficient of dispersion) and the power of the test at each sample size. Notice that, although the estimates from the regression-like function are not directly comparable, the magnitude of the coefficient of dispersion δ is consistently larger under the probabilistic specification than under the naive specification at each sample size, becoming more statistically significant as the sample size increases.

With respect to the coefficient of determination, it is clear that both methodologies produce almost identical results (if not for some rounding error at n=25). This coincidence might lead a researcher to the conclusion that both specifications would produce 'similar' results and that in practice selecting one over the other would be a matter of choice. Unfortunately, the previous justification is flawed for at least two reasons. First of all, as Table 3.11 shows, the average marginal response of Y_t given a change in X_t, that is, $\frac{\partial}{\partial x}E(Y_t|X_t=x_t^\circ)$ when $x_t^\circ=E(x_t)$ are different even at a relatively large sample size (n=500). Thus, selecting one functional form over the other will consistently lead the researcher to different conclusions regarding the effect of X_t on Y_t.

Secondly, and more important, Table 3.12 shows the results of conducting the two-equation misspecification testing on the estimated residuals from both methodologies, in particular the assessment of the existence of additional sources of nonlinearities in both the regression and the skedastic functions. Clearly, the naive specification is

statistically misspecified and the degree of misspecification (revealing the existence of additional sources of nonlinearities) increases with the sample size. As such, the decomposition of systematic and unsystematic information warranted by the naive specification is not orthogonal and any statistical inference, even under unwarranted asymptotic claims of robustness, would be invalid. In fact, the larger the sample size, the more unreliable the inference. In contrast, the decomposition warranted by the probabilistic specification appears to be orthogonal and the researcher may conduct reliable statistical inference in that case.

Experiment 2 Similar to Experiment 1, Experiment 2 sports two non-independent beta distributed random variables (Panel 3.B). While Y_t presents a positively skewed distribution, the negatively skewed distribution of X_t is a mirror image of the former. The scatter plot also reveals the non-normal non-linear dependence between y_t and x_t. Also, similar to Experiment 1, the specification of the naive and the probabilistic models is given by (43) and the full model assumptions are presented also in Table 3.9.

The results, presented in Table 3.13, are analogous to those of Experiment 1. Notice that, also in this case, the magnitude of the coefficient of dispersion $\widehat{\delta}$ is consistently larger under the probabilistic specification than under the naive specification at all sample sizes. Similarly, both methodologies produce identical results with respect to the proposed coefficient of determination.

Marginal changes are also different, as verified by Table 3.11. The average marginal response of Y_t given a change in X_t, is consistently over-estimated by the naive specification as compared to the values produced by the probabilistic model even at the relatively large sample size.

Lastly, and akin to Experiment 1, the naive specification is statistically misspecified and the severity of the misspecification exacerbates with the sample size, rendering any statistical inference conducted on the naive model unreliable. On the other

hand, the probabilistic model warrants reliable inferences.

Experiment 3 For experiment 3, we let the distribution of X_t to continue being positively skewed and allowed its range to be $(0, +\infty)$. Panel 3.C shows the resulting histograms and scatter-plots of this change in distribution. The corresponding specifications under the naive and the probabilistic model are given by

$$\boxed{\text{Naive} \quad \eta_t = \beta_0 + \beta_1 x_t} \qquad \boxed{\text{Probabilistic} \quad \eta_t = \beta_0 + \beta_1 x_t + \beta_2 \ln(x_t)} \tag{44}$$

with the full model assumptions presented in Table 3.10. Not surprisingly, analogous results to the previous two experiments are found; in particular, the similarity between the coefficients of determination, the discrepancy between the coefficients of dispersion and the average marginal responses that do not decrease with the sample size, and the lack of statistical adequacy in the naive specification compared to the probabilistic specification (Table 3.14).

Experiment 4 Experiment 4 uses the same data from Experiment 3 but transforms X_t to fit in the open unit interval, using the algorithm shown in Section 3.2. The histograms and the scatter-plots (Panel 3.D) resemble those of Experiment 3 except for the rescaled axis for X_t that is restricted now to $(0, 1)$. The results are analogous to the previous three cases taking into consideration the fact that the relevant specification under the probabilistic model is that of (32), with equivalent model assumptions (Table 3.9)

These results suggests that the beta distribution transformation has the potential to become a valuable resource when the researcher is unable to assess the underlying marginal and/or conditional distributions of the explanatory variables (Table 3.15).

3.6 Conclusion

The flexibility of the beta distribution makes it ideal for modeling rates and proportions. Maximum likelihood estimators of its mean-variance representation (and their respective standard errors) are strikingly similar to those obtained assuming normality. Nevertheless, inference conducted under the wrong distributional assumption will lead the researcher to astray conclusions.

With respect to modeling the response of a beta distributed random variable as a function of a set of explanatory variables, it is clear that by reinterpreting the values that Y_t can take as a Bernoulli distributed random variable, the logit link function results in the appropriate link function to select. Pairing this result with the study of the conditional distribution of the conditioning set given Y_t, it is possible to establish the right functional form for how the covariates enter the link function following the specifications provided by Bergstold et al., 2009).

When this probabilistic information about X_t is ignored, the average marginal response of Y_t given a change in X_t, obtained under naive specifications will lead the researcher to biased estimates and to statistically misspecified models. If this is the case, no reliable statistical inference can be conducted even under false claims of asymptotic robustness.

We also provide a simple framework, under the probabilistic reduction approach, where the statistical adequacy of proposed beta regression-like models can be put to the test. We also provide a set of conditions that would have to be satisfied by beta regression like models in order to ensure statistical adequacy.

3.7 Appendix 3.A: Tables and Figures

Table 3.9: The Beta-Beta Regression-like Model

$Y_t = \mu_t + u_t,\ t \in \mathbb{N}$	
[1] Beta	$Y_t \sim \mathsf{Beta}(.,.),\ y \in (0,1)$
[2] Non-linearity	$E(Y_t) = \mu_t = \frac{e^{\eta_t}}{1+e^{\eta_t}}$
[3] Heteroskedasticity	$Var(Y_t) = \frac{\mu_t(1-\mu_t)}{\delta+1}$
[4] Independence	$\{Y_t,\ t \in \mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi} := (\boldsymbol{\beta}, \delta)$ do not change with t

where $\eta_t = \beta_0 + \beta_1 \ln(x_t) + \beta_2 \ln(1-x_t)$; $(\boldsymbol{\beta}, \delta) \in \mathbb{R}^k \times \mathbb{R}_+$.

Table 3.10: The Beta-Gamma Regression-like Model

$Y_t = \mu_t + u_t,\ t \in \mathbb{N}$	
[1] Beta	$Y_t \sim \mathsf{Beta}(.,.),\ y \in (0,1)$
[2] Non-linearity	$E(Y_t) = \mu_t = \frac{e^{\eta_t}}{1+e^{\eta_t}}$
[3] Heteroskedasticity	$Var(Y_t) = \frac{\mu_t(1-\mu_t)}{\delta+1}$
[4] Independence	$\{Y_t,\ t \in \mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi} := (\boldsymbol{\beta}, \delta)$ do not change with t

where $\eta_t = \beta_0 + \beta_1 x_t + \beta_2 \ln(x_t)$; $(\boldsymbol{\beta}, \delta) \in \mathbb{R}^k \times \mathbb{R}_+$.

Table 3.11: Marginal Response of $E(Y\|X=x)$ evaluated at $E(X)$						
	Experiment 1		Experiment 2		Experiment 3	
n	Naive	Probabilistic	Naive	Probabilistic	Naive	Probabilistic
25	-3.086	-2.630	-1.038	-1.021	-0.0569	-0.0459
50	-3.011	-2.573	-1.021	-0.999	-0.0561	-0.0454
100	-3.002	-2.570	-1.013	-1.002	-0.0564	-0.0456
500	-2.969	-2.540	-1.013	-0.988	-0.0560	-0.0452

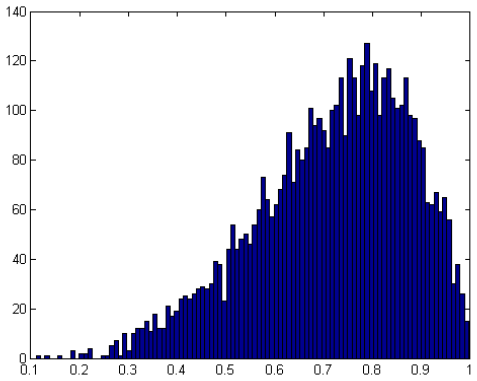
Table 3.12: Experiment 1: Naive Misspecification vs. Probabilistic Specification								
	Naive		**Probabilistic**		**Naive**		**Probabilistic**	
	$n=25$				$n=50$			
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
$\widehat{\beta}_0$	1.678	0.189	0.314	0.641	1.665	0.129	0.354	0.392
$\widehat{\beta}_1$	−4.257	0.923	−0.401	0.198	−4.148	0.613	−0.384	0.121
$\widehat{\beta}_2$	.	.	1.142	1.191	.	.	1.155	0.693
$\widehat{\delta}$	18.15	5.810	22.02	7.303	16.69	3.783	19.66	4.456
$\widehat{R}^2$	0.558	0.135	0.560	0.134	0.562	0.098	0.562	0.098
$\phi^*_{\beta_0}=\frac{\hat{\beta}_0^2}{\hat{\sigma}^2_{\hat{\beta}_0}}$	8.861	0.999	0.490	0.371	12.90	.999	0.901	0.622
$\phi^*_{\beta_1}=\frac{\hat{\beta}_1^2}{\hat{\sigma}^2_{\hat{\beta}_1}}$	4.610	0.999	2.020	0.944	6.762	.999	3.139	0.995
$\phi^*_{\beta_2}=\frac{\hat{\beta}_2^2}{\hat{\sigma}^2_{\hat{\beta}_2}}$	.	.	0.958	0.651	.	.	1.667	0.890
$\phi^*_{\delta}=\frac{\widehat{\delta}^2}{\hat{\sigma}^2_{\hat{\delta}}}$	3.124	0.995	1.958	0.937	4.411	0.999	4.412	0.999
M-S Joint	1.933	0.148	0.149	0.001	2.444	0.281	0.189	0.001
Randomness	3.0645	.067	3.0691	.067	3.0321	.052	2.9638	.052
	$n=100$				$n=500$			
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
$\widehat{\beta}_0$	1.662	0.094	0.344	0.266	1.655	0.043	0.343	0.114
$\widehat{\beta}_1$	−4.137	0.434	−0.386	0.083	−4.090	0.190	−0.384	0.035
$\widehat{\beta}_2$	.	.	1.144	0.460	.	.	1.120	0.177
$\widehat{\delta}$	15.83	2.417	18.54	2.811	15.49	1.082	18.01	1.325
$\widehat{R}^2$	0.572	0.068	0.572	0.068	0.578	0.028	0.578	0.028
$\phi^*_{\beta_0}=\frac{\hat{\beta}_0^2}{\hat{\sigma}^2_{\hat{\beta}_0}}$	17.59	0.999	1.289	0.789	37.84	0.999	3.008	0.993
$\phi^*_{\beta_1}=\frac{\hat{\beta}_1^2}{\hat{\sigma}^2_{\hat{\beta}_1}}$	9.517	0.999	4.636	0.999	21.47	0.999	10.85	0.999
$\phi^*_{\beta_2}=\frac{\hat{\beta}_2^2}{\hat{\sigma}^2_{\hat{\beta}_2}}$	.	.	2.487	0.979	.	.	6.302	0.999
$\phi^*_{\delta}=\frac{\widehat{\delta}^2}{\hat{\sigma}^2_{\hat{\delta}}}$	6.548	0.999	6.597	0.999	14.30	0.999	13.59	0.999
M-S Joint	4.493	0.594	0.267	0.999	19.27	0.999	0.798	0.001
Randomness	2.9429	.09	2.9361	.05	2.7295	.07	2.8556	.03
$Y\sim$Beta $(5,2)$, $X\sim$Beta $(1,5)$, $\widetilde{R}^2$=0.60								

Table 3.13: Experiment 2: Naive Misspecification vs. Probabilistic Specification								
	Naive		**Probabilistic**		**Naive**		**Probabilistic**	
	$n=25$				$n=50$			
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
$\widehat{\beta}_0$	2.285	0.528	−0.566	0.648	2.217	0.357	−0.573	0.375
$\widehat{\beta}_1$	−4.742	0.701	−1.178	0.792	−4.645	0.484	−1.146	0.439
$\widehat{\beta}_2$	.	.	0.631	0.245	.	.	0.613	0.141
$\widehat{\delta}$	16.41	5.149	20.66	7.003	14.87	3.238	18.20	4.319
$\widehat{R}^2$	0.704	0.109	0.704	0.109	0.713	0.074	0.713	0.074
$\phi^*_{\beta_0}=\frac{\hat{\beta}_0^2}{\hat{\sigma}^2_{\hat{\beta}_0}}$	4.323	0.999	0.872	0.607	6.195	0.999	1.526	0.858
$\phi^*_{\beta_1}=\frac{\hat{\beta}_1^2}{\hat{\sigma}^2_{\hat{\beta}_1}}$	6.762	0.999	1.487	0.848	9.590	0.999	2.608	0.984
$\phi^*_{\beta_2}=\frac{\hat{\beta}_2^2}{\hat{\sigma}^2_{\hat{\beta}_2}}$	.	.	2.571	0.982	.	.	4.324	0.999
$\phi^*_{\delta}=\frac{\widehat{\delta}^2}{\hat{\sigma}^2_{\hat{\delta}}}$	3.187	0.995	2.950	0.992	4.593	0.999	4.215	0.999
M-S Joint		0.154		0.001		0.280		0.000
Randomness	3.0612	.077	2.965	.067	3.016	.051	2.9386	0.051
	$n=100$				$n=500$			
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
$\widehat{\beta}_0$	2.193	0.248	−0.536	0.283	2.192	0.119	−0.576	0.115
$\widehat{\beta}_1$	−4.613	0.337	−1.102	0.317	−4.610	0.160	−1.130	0.125
$\widehat{\beta}_2$	.	.	0.627	0.105	.	.	0.604	0.041
$\widehat{\delta}$	14.55	2.209	17.72	3.155	13.83	0.978	16.70	1.277
$\widehat{R}^2$	0.727	0.053	0.727	0.053	0.731	0.021	0.731	0.021
$\phi^*_{\beta_0}=\frac{\hat{\beta}_0^2}{\hat{\sigma}^2_{\hat{\beta}_0}}$	8.823	0.999	1.893	0.928	18.41	0.999	4.996	0.999
$\phi^*_{\beta_1}=\frac{\hat{\beta}_1^2}{\hat{\sigma}^2_{\hat{\beta}_1}}$	13.66	0.999	3.476	0.997	28.70	0.999	8.982	0.999
$\phi^*_{\beta_2}=\frac{\hat{\beta}_2^2}{\hat{\sigma}^2_{\hat{\beta}_2}}$	.	.	5.927	0.999	.	.	14.70	0.999
$\phi^*_{\delta}=\frac{\widehat{\delta}^2}{\hat{\sigma}^2_{\hat{\delta}}}$	6.586	0.999	5.616	0.999	14.13	0.999	13.07	0.999
M-S Joint		0.590		0.010		0.999		0.120
Randomness	2.9773	.02	2.7198	.02				
$Y\sim$Beta $(1,3)$, $X\sim$Beta $(3,1)$, $\widetilde{R}^2$=0.75								

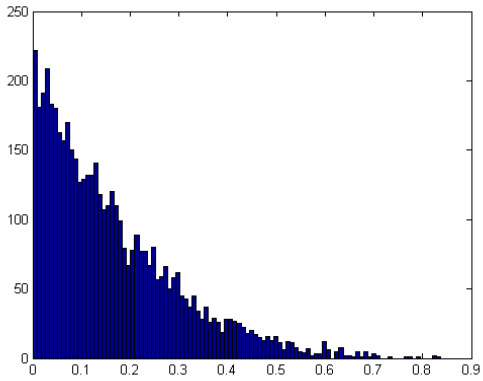
Table 3.14: Experiment 3: Naive Misspecification vs. Probabilistic Specification								
	Naive		**Probabilistic**		**Naive**		**Probabilistic**	
	$n = 25$				$n = 50$			
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
$\widehat{\beta}_0$	-0.244	0.241	-0.285	0.217	-0.258	0.163	-0.303	0.144
$\widehat{\beta}_1$	-0.278	0.062	-0.115	0.133	-0.271	0.041	-0.115	0.077
$\widehat{\beta}_2$	.	.	-0.531	0.432	.	.	-0.504	0.251
$\widehat{\delta}$	18.37	6.083	20.63	6.812	16.80	3.992	18.41	4.280
$\widehat{R}^2$	0.588	0.135	0.559	0.134	0.562	0.098	0.562	0.098
$\phi^*_{\beta_0}=\frac{\hat{\beta}^2_0}{\hat{\sigma}^2_{\hat{\beta}_0}}$	1.013	0.678	1.310	0.796	1.579	0.871	2.108	0.953
$\phi^*_{\beta_1}=\frac{\hat{\beta}^2_1}{\hat{\sigma}^2_{\hat{\beta}_1}}$	4.441	0.999	0.863	0.602	6.479	0.999	1.495	0.851
$\phi^*_{\beta_2}=\frac{\hat{\beta}^2_2}{\hat{\sigma}^2_{\hat{\beta}_2}}$	.	.	1.228	0.767	.	.	2.008	0.943
$\phi^*_{\delta}=\frac{\widehat{\delta}^2}{\hat{\sigma}^2_{\hat{\delta}}}$	3.020	0.993	3.028	0.993	4.209	0.999	4.301	0.999
M-S Joint	0.792	0.021	0.104	0.001	1.082	0.039	0.116	0.001
	$n = 100$				$n = 500$			
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
$\widehat{\beta}_0$	-0.252	0.119	-0.297	0.102	-0.260	0.055	-0.306	0.049
$\widehat{\beta}_1$	-0.271	0.030	-0.115	0.053	-0.268	0.013	-0.113	0.018
$\widehat{\beta}_2$	.	.	-0.506	0.173	.	.	-0.504	0.064
$\widehat{\delta}$	15.97	2.535	17.37	2.652	15.63	1.128	16.89	1.248
$\widehat{R}^2$	0.572	0.068	0.572	0.068	0.578	0.028	0.578	0.028
$\phi^*_{\beta_0}=\frac{\hat{\beta}^2_0}{\hat{\sigma}^2_{\hat{\beta}_0}}$	2.110	0.953	2.899	0.991	4.658	0.999	6.232	0.999
$\phi^*_{\beta_1}=\frac{\hat{\beta}^2_1}{\hat{\sigma}^2_{\hat{\beta}_1}}$	9.018	0.999	2.168	0.958	20.33	0.999	5.994	0.999
$\phi^*_{\beta_2}=\frac{\hat{\beta}^2_2}{\hat{\sigma}^2_{\hat{\beta}_2}}$	.	.	2.924	0.992	.	.	7.835	0.999
$\phi^*_{\delta}=\frac{\widehat{\delta}^2}{\hat{\sigma}^2_{\hat{\delta}}}$	6.299	0.999	6.550	0.999	13.85	0.999	13.54	0.999
M-S Joint	1.844	0.133	0.144	0.001	7.109	0.990	0.381	0.001
$Y \sim \mathsf{Beta}\,(1.5, 5)$, $X \sim \mathsf{\Gamma}\,(2, 2)$, $\widetilde{R}^2$=0.60								

Table 3.15: Experiment 4: Naive Misspecification vs. Probabilistic Specification								
	Naive		**Probabilistic**		**Naive**		**Probabilistic**	
	$n = 25$				$n = 50$			
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
$\widehat{\beta}_0$	-0.335	0.264	-1.988	0.455	-0.328	0.181	-2.011	0.346
$\widehat{\beta}_1$	-3.074	0.927	-0.562	0.201	-3.417	0.806	-0.554	0.147
$\widehat{\beta}_2$	.	.	0.266	0.232	.	.	0.331	0.206
$\widehat{\delta}$	18.37	6.083	19.70	6.366	16.80	3.992	17.43	17.43
$\widehat{R}^2$	0.558	0.135	0.560	0.133	0.562	0.098	0.562	0.098
$\phi^*_{\beta_0}=\frac{\hat{\beta}_0^2}{\hat{\sigma}^2_{\hat{\beta}_0}}$	1.270	0.783	4.367	0.999	1.805	0.915	5.809	0.999
$\phi^*_{\beta_1}=\frac{\hat{\beta}_1^2}{\hat{\sigma}^2_{\hat{\beta}_1}}$	3.312	0.996	2.789	0.989	4.235	0.999	3.760	0.998
$\phi^*_{\beta_2}=\frac{\hat{\beta}_2^2}{\hat{\sigma}^2_{\hat{\beta}_2}}$	.	.	1.143	0.735	.	.	1.609	0.878
$\phi^*_{\delta}=\frac{\widehat{\delta}^2}{\hat{\sigma}^2_{\hat{\delta}}}$	3.020	0.993	3.094	0.994	4.209	0.999	4.338	0.999
M-S Joint	0.792	0.021	0.134	0.000	1.082	0.039	0.243	0.001
	$n = 100$				$n = 500$			
	mean	s.e.	mean	s.e.	mean	s.e.	mean	s.e.
$\widehat{\beta}_0$	-0.304	0.127	-2.065	0.303	-0.287	0.061	-2.108	0.205
$\widehat{\beta}_1$	-3.870	0.765	-0.566	0.115	-4.815	0.732	-0.569	0.075
$\widehat{\beta}_2$	.	.	0.435	0.221	.	.	0.860	0.269
$\widehat{\delta}$	15.97	2.535	16.42	2.486	15.63	1.128	16.34	1.232
$\widehat{R}^2$	0.572	0.068	0.572	0.068	0.578	0.028	0.578	0.028
$\phi^*_{\beta_0}=\frac{\hat{\beta}_0^2}{\hat{\sigma}^2_{\hat{\beta}_0}}$	2.383	0.973	6.806	0.999	4.706	0.999	10.27	0.999
$\phi^*_{\beta_1}=\frac{\hat{\beta}_1^2}{\hat{\sigma}^2_{\hat{\beta}_1}}$	5.059	0.999	4.906	0.999	6.572	0.999	7.560	0.999
$\phi^*_{\beta_2}=\frac{\hat{\beta}_2^2}{\hat{\sigma}^2_{\hat{\beta}_2}}$	.	.	1.962	0.937	.	.	3.189	0.995
$\phi^*_{\delta}=\frac{\widehat{\delta}^2}{\hat{\sigma}^2_{\hat{\delta}}}$	6.299	0.999	6.605	0.999	13.85	0.999	13.26	0.999
M-S Joint	1.844	0.133	0.449	0.001	7.109	0.990	1.318	0.001
$Y \sim \mathsf{Beta}\,(1.5, 5)$, $X \sim \mathsf{Beta}\,(\Gamma\,(2, 2))$, $\widetilde{R}^2$=0.60								

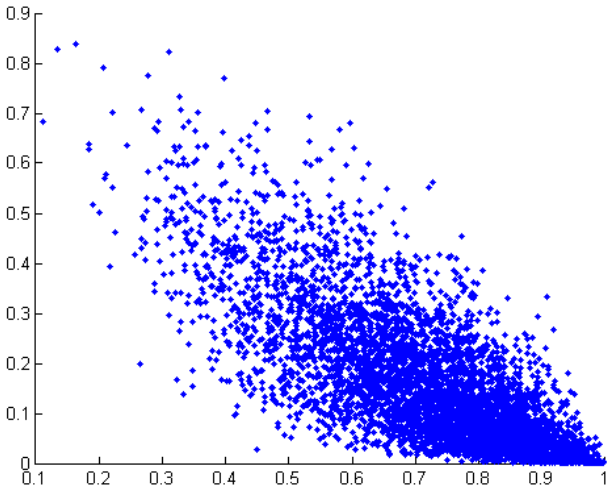
Panel 3.A: Histograms and scatter-plots of (y_t, x_t) from Experiment 1



Histogram of $Y_t \sim \mathsf{Beta}\,(5, 2)$

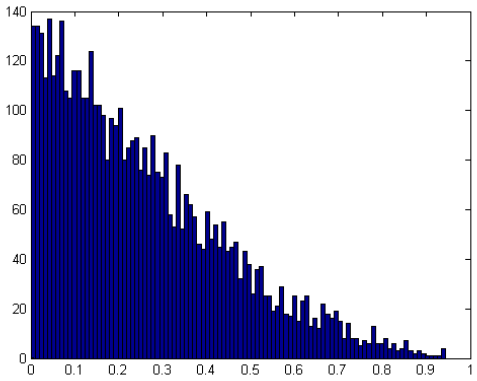


Histogram of $X_t \sim \mathsf{Beta}\,(1, 5)$

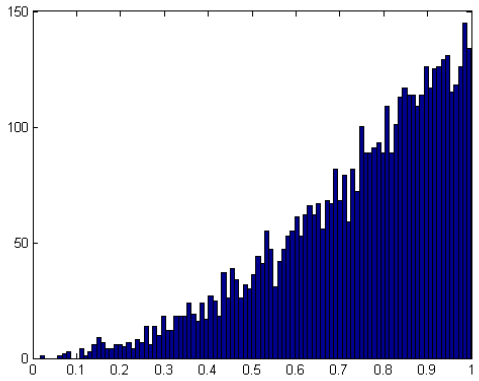


Scatter-plot of (y_t, x_t)

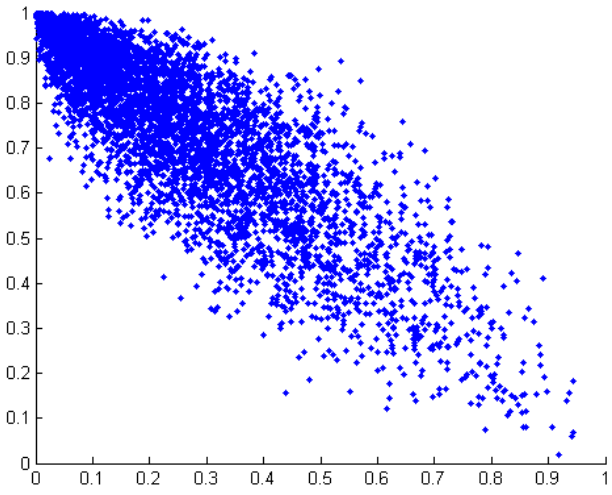
Panel 3.B: Histograms and scatter-plots of (y_t, x_t) from Experiment 2



Histogram of $Y_t \sim \mathsf{Beta}\,(1, 3)$

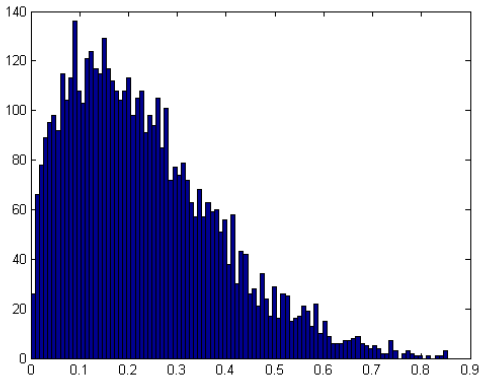


Histogram of $X_t \sim \mathsf{Beta}\,(3, 1)$

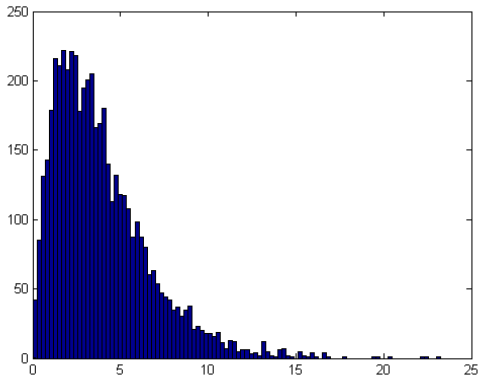


Scatter-plot of (y_t, x_t)

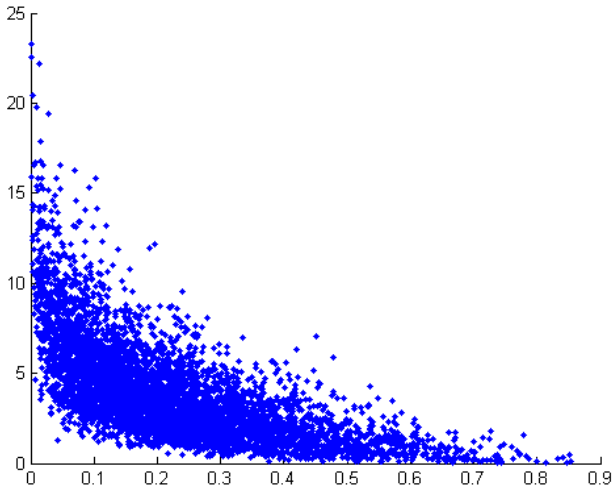
Panel 3.C: Histograms and scatter-plots of (y_t, x_t) from Experiment 3



Histogram of $Y_t \sim \mathsf{Beta}\,(1.5, 5)$

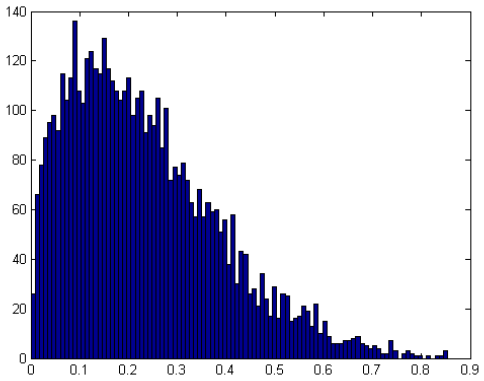


Histogram of $X_t \sim \Gamma\,(2, 2)$

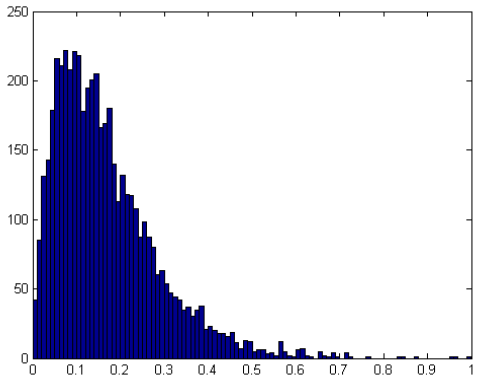


Scatter-plot of (y_t, x_t)

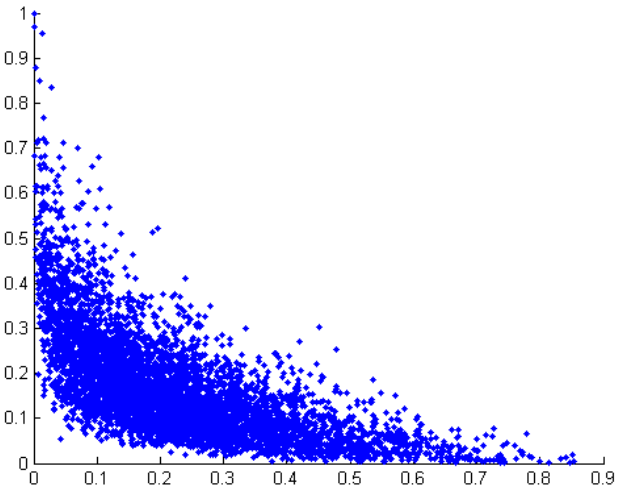
Panel 3.D: Histograms and scatter-plots of (y_t, x_t) from Experiment 4



Histogram of $Y_t \sim \mathsf{Beta}\,(1.5, 5)$



Histogram of $X_t \sim \mathsf{Beta}\,(\Gamma\,(2, 2))$



Re-scaled Scatter-plot of (y_t, x_t)

4 LogNormal Regression Models

4.1 Introduction

Consistent estimation of price and income elasticities of the demand is a perennial need in economic analysis, particularly in gasoline markets. The several drastic changes in the oil markets that have occurred since the 1970's, and their impact on gasoline prices 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 1974-1986, period that encompasses two major oil supply disruptions: the Arab Oil Embargo and the Iranian Revolution/Iran-Iraq Conflict. This period will allow us to describe the time path of the price elasticity to find drastic changes in its value. It seems that price elasticity became relatively less elastic after the outset of the Iranian Revolution.

Several studies corroborate the existence of changes in the value of the elasticity of demand. Hughes et al. (2006, 2008) provide evidence that the short-run price elasticity of gasoline demand is statistically different from the 1975-1980 lustrum to the 2001-2006 one. Short-run price elasticities differ considerably between periods, ranging from $(-0.21, -0.75)$ to $(-0.034, -0.077)$, just twenty years later.

To have an idea of the time consistency of the estimator of price elasticity, we conducted two thought experiments. In the first one, we ran several regressions with 60 observations each, following a double-log specification (see Equation 1), using a moving window estimator over the period 1973-2006, obtaining a set of estimators for the price elasticity. These 'naive' window estimators will help evaluate their time stationarity. In the second experiment, we ran cumulative regressions, starting with the first 60 observations and adding one observation at the time, also following a double-log specification (45). This 'naive' procedure will also help assess the existence of nonstationarity in the estimators.

The common specification for either estimation is given by

$$\ln G_{jt} = \beta_0 + \beta_1 \ln P_{jt} + \beta_2 \ln Y_{jt} + e_j + e_{jt} \tag{45}$$

where G_t is per capita gasoline consumption in gallons in month j and year t, P_t is the real retail price of gasoline in month j and year t, Y_t 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[22].

Each estimate, with its empirical 95 percent confidence interval, corresponds to the estimation of the price-elasticity coefficient under either the 60 month moving windows estimation (left panel of Figure 4.1) or the cumulative estimation (right panel of Figure 4.1). As the right panel of Figure 4.1 suggests, the shift in the value of the price elasticity occurred early in the sample. Note that, after the shift, the apparent time consistency of the estimator does not improve, and hovers around -0.05. Its time consistency is questionable, as it can be seen from the right panel of the same figure.

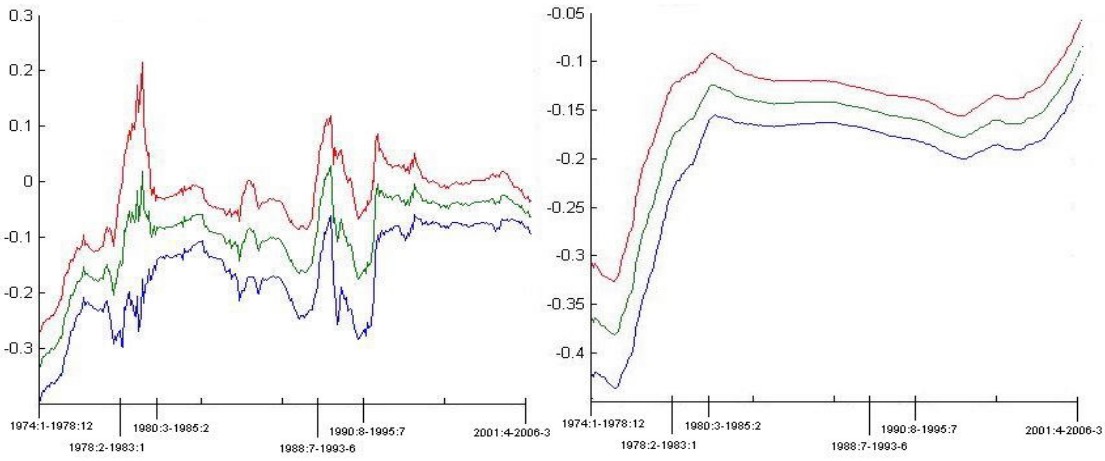


Figure 4.1: Moving Windows Estimators and Cumulative Estimators of the Price Elasticity of Gasoline Demand for the period 1973-2006.

[22]The specification is analogous to Hughes et al. (2006, 2008) that find it 'robust' to the inclusion of additional variables and functional form as well as 'exogeneous' to supply side factors.

Based on these preliminary results, we narrowed down the sample size from Jan:1974 to Sep:1986. This will allow us to identify the timing of the structural break in the price elasticity of gasoline. A formal search for the structural break reveals that the price elasticity of gasoline demand changed at around Jan:1979, roughly coinciding with the outbreak of the Iranian Revolution. The rest of this paper is divided as follows, Section 4.2 derives the specification of lognormally distributed regression models, Section 4.3 explaining the testing framework for identifying structural breaks, Section 4.4 presents the results of establishing the existence of a structural break in the price elasticity of gasoline demand, and finally Section 4.5 summarizes this work.

4.2 Modeling Nonnegative Data

Certain random variables have a nonnegative support, including several economic variables. Nominal interest rates or nominal gross domestic product, for instance, can only take values in the $(0, \infty)$ interval. The quintessential example of nonnegative values are prices and quantities, widely used in supply and demand analysis. The fact that the random variables involved cannot take negative values, however, may cause a problem in estimation if this information is not taken into account by the modeler. Even after assuming exogeneity in the estimated equations, classical regression analysis would fail to provide consistent estimators if the underlying probability distribution of the random variables is not only not normal but also its skewness coefficient is statistically different from zero.

Nonnegative data can arise from at least two different reasons. One, the underlying distribution of a random variable Y_t has indeed $\mathbb{R}$ as a domain but has been truncated and/or censored at $Y_t > 0$. Two, the random variable has indeed $(0, \infty)$ as its support. With respect to the first case, there exist already in the literature several ways of dealing with these problems, whether the use of truncated regression model or the use of censored regression models, for instance. With respect to the second case, several attempts to model nonnegative data have been made in the statistics lit-

erature but not in the econometrics literature. The former discussion does not mean, however, that logarithmic transformations have not been employed to model this kind of data but rather that the implementation of these transformation is incomplete.

Lognormally distributed variables are particularly important in economics where the rate of growth of the variables is unknown but of exponential base. Nominal prices are economic entities that are not only naturally nonnegative but their magnitudes may increase proportionally with market sizes and/or other economic conditions. When the base of their growth is e, it is possible to obtain a measure of their proportional change by using natural logarithms. In most economic data driven models, it is a common practice to take the natural logarithm of the nonnegative dependent variable before conducting econometric analysis. This has empirically proven to have two effects: one, it allows the modeler to obtain a measure of the growth rate of the data in question and two, it allows the modeler to estimate a constant growth rate of the variance of the dependent variable (the *rule of thumb* is that, by taking logarithms, 'data anomalies,' like heteroskedasticity, are 'fixed' most of the time).

One crucial step, however, has been missing when reverting the model from the logarithmic units to the original units. In practice, the modeler simply obtains the antilogarithm of the regression function, adjusting the resulting estimated residuals from additive to multiplicative while changing their distribution from normal to lognormal. This procedure is incomplete and may lead to severely biased representations of the true statistical relationship of the lognormally distributed variable and its conditioning set.

4.2.1 Simple Lognormal Models

Let Y_t be a lognormally distributed variable with support in $(0, \infty)$. In this univariate setting, the lognormal probability distribution function is given by,

$$f(y_t; \mu, \sigma^2) = \frac{1}{y_t\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln(y_t) - \mu)}{2\sigma^2}\right], \; y_t \in (0, +\infty).$$

where $E(Y_t)=e^{\mu+\frac{1}{2}\sigma^2}$, $Var(Y_t)=\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$, and $(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+$. Note that the kernel of the distribution function resembles that of a normal distribution function for $\ln(y_t)$ but the shape component includes and additional $\{1/y_t\}$. The lognormal distribution is completely characterized (similarly to the normal distribution), by two parameters: a central tendency parameter $\mu\in\mathbb{R}$ and a central dispersion parameter, $\sigma^2\in\mathbb{R}_+$. To avoid confusion in this and subsequent sections, we will define $Y_t^*=\ln(Y_t)$ where Y_t^* is a normally distributed variable with mean μ and variance σ^2 (the same parameter set that characterizes the lognormal distribution). These two variables are related in the following way,

$$\text{If } Y_t^*\sim\mathsf{N}\left(\mu,\sigma^2\right) \text{ then } Y_t:=\exp\left[Y_t^*\right]\sim\mathsf{LN}\left(\mu,\sigma^2\right)$$

Whereas these two random variables share the same parameter space, their first and second central moments are different and one cannot be obtained from the mere application of the logarithm (antilogarithm) of the other, as is sometimes suggested in introductory econometrics textbooks (see Dougherty, 2002). This is clearly seen in Table 4.1.

Table 4.1: Moment correspondence between $Y_t^*\sim N(\mu,\sigma^2)$ and $Y_t\sim LN(\mu,\sigma^2)$

Normal	Antilogarithmic Transformation	LogNormal
$E(Y_t^*)=\mu$	$e^{E(Y_t^*)}=e^{\mu}$	$E(Y_t)=e^{\mu+\frac{1}{2}\sigma^2}$
$Var(Y_t^*)=\sigma^2$	$e^{Var(Y_t^*)}=e^{\sigma^2}$	$Var(Y_t)=\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}$

Notice that the antilogarithmic transformation of the expected value of the normal counterpart is not equal to the expected value of the lognormally distributed variable. It includes not only the antilogarithm of the expected value of Y_t^* but also the antilogarithm of half the magnitude of the variance of Y_t^*. Of course, the 'damage' of ignoring the second central moment of Y_t^* will depend on its magnitude.

With this information, it is possible to state the following properties for a simple lognormal model for Y_t, where conditions [1]-[5] imply that a realization from this

model constitutes a random sample, where the observations are lognormal, independent, and identically distributed (Table 4.2).

Table 4.2: The Simple LogNormal Model	
SGM: $Y_t = \mu + u_t$, $t \in \mathbb{T}$	
[1] LogNormal	$Y_t \sim \mathsf{LN}(.,.)$, $y_t \in (0, +\infty)$
[2] Constant mean	$E(Y_t) = e^{\mu + \frac{1}{2}\sigma^2} = \eta_0$
[3] Constant variance	$Var(Y_t) = \left(e^{\sigma^2} - 1\right) e^{2\mu + \sigma^2} = \omega_0^2$
[4] Independence	$\{Y_t,\ t \in \mathbb{N}\}$ is an independent process
[5] t-homogeneity	(μ, σ^2) do not change with t
where $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ and $(\eta_0, \omega_0^2) \in \mathbb{R}_+ \times \mathbb{R}_+$	

4.2.2 Lognormal Regression Models

The relationship between a univariate normally distributed random variable and a derived lognormally distributed random variable is preserved in $n-$dimensions, as long as each of the variables in $n \in \mathbb{N}$ is also lognormally distributed. This allows extending the previous discussion from the univariate simple lognormal model to a multivariate framework.

Let $\mathbf{Z}^* = \{Z_1^*, Z_2^*, \ldots, Z_N^*\}$ be a random vector having a multivariate normal distribution[23] with mean $\boldsymbol{\mu}$ and variance-covariance matrix given by $\boldsymbol{\Sigma} = [\sigma_{ij}]$, where $\sigma_{ij} = Var(Z_i^*)$ if $i = j$ and $\sigma_{ij} = Cov\left(Z_i^*, Z_j^*\right)$ otherwise. Now we use the transform $Z_i = \exp\{Z_i^*\}$, $i \in \mathbb{N}$, and define $\mathbf{Z} = \{Z_1, Z_2, \ldots, Z_N\}$, also a random vector.

The density of $\mathbf{Z}$ is multivariate lognormal distribution given by

$$f(\mathbf{z}) = \frac{1}{\mathbf{z}(2\pi)^{N/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{[\ln(\mathbf{z}) - \boldsymbol{\mu}]^{\intercal} \boldsymbol{\Sigma}^{-1} [\ln(\mathbf{z}) - \boldsymbol{\mu}]}{2}\right\},\ 0 < \mathbf{z} < \infty \qquad (46)$$

where $\ln(\mathbf{z})$ is a $n-$component column vector and $Z_i = \exp\{Z_i^*\}$. Note that, also in the multivariate case, the kernel resembles that of the multivariate normal distribu-

[23] The subindex $t \in \mathbb{T}$ has been omitted for clarity.

tion[24] but the shape component now includes and additional $\{1/\mathbf{z}\}$ factor.

In a natural extension to the univariate case, $E(\mathbf{Z})=[\eta_1, \eta_2, \ldots, \eta_n]$, where $\eta_i = \exp\left(\mu_i + \frac{1}{2}\sigma_{ii}\right)$, and σ_{ii} is the ith diagonal element of $\mathbf{\Sigma}$. The variance-covariance matrix of $\mathbf{Z}$ is given by $\mathbf{\Omega}=E\left[(\mathbf{Z}-\boldsymbol{\eta})(\mathbf{Z}-\boldsymbol{\eta})^{\intercal}\right]=\omega_{ij}$, where

$$\omega_{ij}=\left\{\left[\exp\left(\left(\mu_i+\mu_j\right)+\frac{(\sigma_{ii}+\sigma_{jj})}{2}\right)\right]\left[\exp\left(\sigma_{ii}\right)-1\right]\right\},$$

σ_{ii} is the ith diagonal element of $\mathbf{\Sigma}$, and σ_{ij} is the ith-jth element of $\mathbf{\Sigma}$ (see Kotz et al., 2000).

It is relatively straightforward to obtain the conditional probability density functions of the multivariate lognormal distribution. Let $\mathbf{Z}_t := [Y_t, \mathbf{X}_t]^{\intercal}$ be multivariate lognormal distributed where $\{Y_t,\ t\in\mathbb{T}\}$ is the dependent or conditioned random variable and $\mathbf{X}_t = \{X_{1,t}, \ldots, X_{K,t},\ t\in\mathbb{T}\}$ is a is a set of conditioning variables. Additionally, let $\mathbf{Z}_t^* := \begin{pmatrix} Y_t^* \\ \mathbf{X}_t^* \end{pmatrix}$ be distributed multivariate normal, where $E\left(\mathbf{Z}_t^*\right) = \begin{pmatrix} \mu_{y^*} \\ \boldsymbol{\mu}_{\mathbf{x}^*} \end{pmatrix}$ and $Var\left(\mathbf{Z}_t^*\right) = \begin{pmatrix} \sigma_{y^*}^2 & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$.

The conditional distribution of Y_t given $\mathbf{X}_t = \mathbf{x}_t$ can then be derived as follows (see Yue, 2000),

$$f_{Y|\mathbf{X}=\mathbf{x}}(y)=\frac{f_{Y,\mathbf{X}}(y,\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}=\frac{1}{y\sigma_{y^*|\mathbf{x}^*}\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\ln(y)-\mu_{y^*|\mathbf{x}^*}}{\sigma_{y^*|\mathbf{x}^*}}\right)^{1/2}\right\}$$

where $\mu_{y^*|\mathbf{x}^*}=\mu_{y^*}-\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{22}^{-1}\left(\ln\left(\mathbf{x}\right)-\boldsymbol{\mu}_{\mathbf{x}^*}\right)$, and $\sigma^2_{y^*|\mathbf{x}^*}=\sigma^2_{y^*}-\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}=\sigma_0^2$. Notice that $\mu_{y^*|\mathbf{x}^*}$ and $\sigma^2_{y^*|\mathbf{x}^*}$ correspond to the regression function and the skedastic function, respectively, for the vector $\mathbf{Z}_t^* := [Y_t^*, \mathbf{X}_t^*]^{\intercal}$. Notice additionally that the conditional distribution of $(Y_t|\mathbf{X}_t = \mathbf{x}_t)$ is univariate lognormally distributed. This information allows us to define the first and second conditional moments as

$$E\left(Y|\mathbf{X}_t=\mathbf{x}_t\right)=e^{\mu_{y^*|\mathbf{x}^*}+\frac{1}{2}\sigma^2_{y^*|\mathbf{x}^*}} \text{ and } Var\left(Y|\mathbf{X}_t=\mathbf{x}_t\right)=\left(e^{\sigma^2_{y^*|\mathbf{x}^*}}-1\right)e^{2\mu_{y^*|\mathbf{x}^*}+\sigma^2_{y^*|\mathbf{x}^*}}$$

[24] In this case, $f(\mathbf{z}^*)=\frac{1}{(2\pi)^{N/2}|\mathbf{\Sigma}|^{1/2}}\exp\left\{-\frac{[\mathbf{z}^*-\boldsymbol{\mu}]^{\intercal}\mathbf{\Sigma}^{-1}[\mathbf{z}^*-\boldsymbol{\mu}]}{2}\right\}, -\infty<\mathbf{z}^*<\infty.$

Thus, the lognormal regression model can be stated as follows (Table 4.3),

Table 4.3: The Lognormal Regression Model

	$y_t = \alpha_0 \prod_{k=1}^{K} X_{k,t}^{\beta_k} + u_t$
[1] Lognormality	$(y_t \mid \mathbf{X}_t = \mathbf{x}_t; \boldsymbol{\varphi}_1) \sim \mathsf{LN}(.,.)$
[2] Exponential Growth	$E(y_t \mid \mathbf{X}_t = \mathbf{x}_t; \boldsymbol{\varphi}_1) = \alpha_0 \prod_{k=1}^{K} X_{k,t}^{\beta_k}$
[3] Heteroskedasticity	$Var(y_t \mid \mathbf{X}_t = \mathbf{x}_t; \boldsymbol{\varphi}_1) = \delta_0 \prod_{k=1}^{K} X_{k,t}^{2\boldsymbol{\beta}_k}$
[4] Independence	$\{(y_t \mid \mathbf{X}_t = \mathbf{x}_t), t \in \mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi}_1 := (\alpha_0, \delta_0, \beta_k, k=0,1,\ldots,K, \sigma_0^2)$ do not change with t
$\alpha_0 = \exp\left\{\beta_0 + \frac{\sigma_0^2}{2}\right\}$, $\delta_0 = \alpha_0^2\left(e^{\sigma_0^2} - 1\right)$, $\beta_0 = \mu_{y^*} - \boldsymbol{\beta}^{\intercal}\boldsymbol{\mu}_{\mathbf{x}^*}$,	
$\boldsymbol{\beta} := [\beta_1, \beta_2, \ldots, \beta_K]^{\intercal} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, $\sigma_0^2 = \sigma_{y^*}^2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}$	

Bivariate Case As an illustration, let us consider the bivariate case. Let $(Y_t, X_t) \sim \mathsf{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} := \binom{\mu_{y^*}}{\mu_{x^*}}$ and $\boldsymbol{\Sigma} := \begin{pmatrix} \sigma_{y^*}^2 & \sigma_{y^*x^*} \\ \sigma_{y^*x^*} & \sigma_{x^*}^2 \end{pmatrix}$, then, the bivariate lognormal regression model is defined as (Table 4.4),

Table 4.4: The Bivariate Lognormal Regression Model

	$y_t = \alpha_0 X^{\beta_1} + u_t$
[1] Lognormality	$(y_t \mid X_t = x_t; \boldsymbol{\varphi}_1) \sim \mathsf{LN}(.,.)$
[2] Exponential Growth	$E(y_t \mid X_t = x_t; \boldsymbol{\varphi}_1) = \alpha_0 X^{\beta_1}$
[3] Heteroskedasticity	$Var(y_t \mid X_t = x_t; \boldsymbol{\varphi}_1) = \delta_0 X^{2\beta_1}$
[4] Independence	$\{(y_t \mid X_t = x_t), t \in \mathbb{N}\}$ is an independent process
[5] t-homogeneity	$\boldsymbol{\varphi}_1 := (\alpha_0, \delta_0, \beta_1, \sigma_0^2)$ do not change with t
$\alpha_0 = \exp\left\{\beta_0 + \frac{\sigma_0^2}{2}\right\}$, $\delta_0 = \alpha_0^2\left(e^{\sigma_0^2} - 1\right)$, $\beta_0 = \mu_{y^*} - \beta_1\mu_{\mathbf{x}^*}$,	
$\beta_1 = \sigma_{y^*x^*}/\sigma_{x^*}^2$, $\sigma_0^2 = \sigma_{y^*}^2 - (\sigma_{y^*x^*})^2/\sigma_{x^*}^2$	

4.2.3 Estimation

It is clear that direct estimation of $\boldsymbol{\eta}$ and $\boldsymbol{\Omega}$ via maximum likelihood can be avoided by exploiting the connection between the lognormal and the normal distributions

and estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ instead. To see this, let $\mathbf{Z} \sim \mathsf{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{Z}^* \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $f_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $f_{\mathbf{Z}^*}(\mathbf{z}^*; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ represent their density functions, respectively. Then, it is clear, from (2), that $f_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{z} f_{\ln(\mathbf{Z})}(\ln(\mathbf{z}); \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The log-likelihood function of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{z})$ is then,

$$l_{\mathbf{Z}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{z}_t) = -\sum_{i=1}^{T} \ln(\mathbf{z}_t) + l_{\ln(\mathbf{Z})}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\ln(\mathbf{z}_t)). \tag{47}$$

We can deduce, from the results of the multivariate normal distribution (see Spanos, 1986; ch. 15), that $\widehat{\mu}_i = \frac{1}{T}\sum \ln(z_{it})$, and $\widehat{\sigma}_{ij} = \sum (\ln(z_{it}) - \widehat{\mu}_i)\left(\ln(z_{jt}) - \widehat{\mu}_j\right)$; hence,

$$\begin{aligned} \widehat{\eta}_i &= \exp\left\{\widehat{\mu}_i + \frac{1}{2}\widehat{\sigma}_{ij}\right\} \\ \widehat{\omega}_{ij} &= \left[\exp\left\{\widehat{\mu}_i + \widehat{\mu}_j\right\} + \frac{1}{2}\left(\widehat{\sigma}_{ii} + \widehat{\sigma}_{jj}\right)\right]\left[\exp\left(\widehat{\sigma}_{ij}\right) - 1\right] \end{aligned}$$

4.3 Testing for Structural Breaks

The issue of parameters constancy and its importance in statistical inference has taken particular importance since the seminal work of Chow (1960), who attempted to establish a methodology for the detection of structural breaks in stationary time series data. Since then, several complementary and alternative methodologies have been proposed. Koutris et al. (2008) surveyed and tested several methodologies on single break test-statistics based on Chow (Quandt, 1960; Gardner, 1969; Nyblom, 1989; Hansen, 1991; Andrews, 1993; Andrews and Ploberger, 1994), on Recursive Residuals (Brown, Durbin, and Evans, 1975; Ploberger and Cramer, 1992; Ploberger and Cramer, 1996), as well as extensions to multiple breaks (Bai and Perron, 1998; Hansen, 2000), revealing that there is no dominant statistic for the detection of structural breaks.

In particular, Koutris et al. (2008) evaluated, using Monte Carlo experiments, the power of structural breaks tests based on Andrews (1993), Andrews and Ploberger (1994), and Hansen (2000); three of the most popular tests for structural changes.

Through the experiments, which under a Normal Linear Regression model framework included one time shifts in the mean; time trends in the mean; one time shifts in the covariance matrix; and time trends in the covariance matrix (with some variation that accounted for smooth mean trends and smooth variance trends); they found no evidence of any test being superior. Furthermore, for small and medium sample sizes (100 to 400 observations), the actual error probabilities turned out to be almost three times larger than the nominal error probabilities (generating an actual Type I error probability of 15 percent versus the expected 5 percent). Additionally, the tests demonstrated to have very low power against continuos parameter changes.

They proposed the combination of Rolling Overlapping Window Estimators (ROWE) and Maximum Entropy (ME) Bootstraps[25] to improve the power of structural break tests. The resampling procedure, based on Vinod (2004), provides the researcher with replicas of the original non-stationary and temporally dependent realization of the process that contain the same amount of statistical information. This procedure boosts the informational base of a single realization of an observed series.

4.3.1 Rolling Overlapping Window Estimators

The ROWE is defined in the following fashion. Let $\{R_t\}_{t=1,\ldots,n}$ be a random process, θ be the unknown parameter to be estimated, and $\widehat{\theta}=g(\mathbf{R})$ be an estimator based on the process. Additionally, let $P_R=\{P_{R_i}\}_{i\in I}$ be a partition of the process, such that,

$$P_{R_{t_i}}=\{R_t{:}t\in[t_i,t_i-1+l]\},\quad t_i=1,2,\ldots,n-(l-1),$$

[25]The ME bootstrap is similar to Efron's traditional bootstrap but avoids the following three restrictions over a time series x_t in the range $t=1,\ldots,T$. Restriction 1: The traditional bootstrap sample repeats some x_t values, requiring that none of the resample values can differ from the observed ones. Restriction 2: It also requires the bootstrap resamples to lie in the interval $[\min(x_t),\max(x_t)]$. Restriction 3: The traditional bootstrap resample shuffles x_t in such a way that all dependence and heterogeneity information in the time series sequence is lost. The ME bootstrap simultaneously avoids all three problems.

where l is a fixed window size. The ROWE $\widehat{\theta}_{r_{t_i}}$ of the unknown parameter θ is defined as,

$$\widehat{\theta}_{r_{t_i}}=g\left(P_{R_{t_i}}\right) \text{ for } t_i=1,2,\ldots,n-(l-1)$$

Thus, the estimators are based on a variant subsample of fixed length l that moves sequentially through the sample,generating a series of estimates for θ.

4.3.2 Testing Framework

With the ROWE, the testing procedure consists of the sequential application of F-tests. It can be described in the following 7 steps:

1. Select a variable whose time invariance is to be assessed. Determine the appropriate window size l. For $n\leq 150$, the proposed rule of thumb is $l=\left[\frac{n}{10}\right]-2$, where n is the total sample size.

2. For each window of size l, generate an additional number of ME bootstrap samples.

3. Estimate the sample mean and variance for each window. This generates a sequence of $T=n-(l-1)$ sample means, $\widehat{\mu}(t_i)$, and variance estimates, $\widehat{\sigma}^2(t_i)$.

4. Test for time invariance in the mean, assuming first order Markov dependence, with H_0:$\mu(t_i)=\mu$ for $t_i=1,\ldots,n-(l-1)$ being the null hypothesis and

   $$\widehat{\mu}(t_i)=a_0+a_1\widehat{\mu}(t_i-1)+u_{r\mu}(t_i)$$

   being the restricted formulation whose parameters have to be estimated. Keep the Restricted Sum of Squared Residuals ($RSSR$).

5. Test for time trends in the mean using Bernstein orthogonal polynomials of sufficient high degree, assuming first order Markov dependence, with H_1:$\mu(t_i)\neq\mu$ being the alternative hypothesis and

   $$\widehat{\mu}(t_i)=a_0'+a_1'\widehat{\mu}(t_i-1)+B_{k,t_i}+u_{u\mu}(t_i)$$

being the unrestricted formulation for the mean whose parameters have to be estimated and where B_{k,t_i} is the k^{th} degree Bernstein Orthogonal polynomial[26] at time t. Keep the Unrestricted Sum of Squared Residuals ($USSR$).

6. Calculate the $F-$statistic based on the $RSSR_\mu$ and the $USSR_\mu$ adjusting for the appropriate degrees of freedom, $(T-(k+2),k)$. .

7. Repeat the same procedure for all the relevant variables in the model.

4.4 The Elasticity of Gasoline Demand

4.4.1 Data Analysis

The data set consists of monthly observations of per capita gasoline consumption in gallons, G_t, the retail price of gasoline, P_t, and the per capita disposable income, Y_t, from January 1974 to March 2006. The time period used in this document spans from January 1974 to September 1986 for reasons that will become apparent in the sequel. All variables are in measured in 2000 dollars. Gasoline consumption is approximated as monthly product supplied, calculated as domestic production plus imports less exports and changes to stock. Real gasoline prices are U.S. city average prices for unleaded regular fuel. The data was kindly provided to us by Hughes et al. (2008), who collected it from several sources, including the U.S. Energy Information Administration, the U.S. Bureau of Labor Statistics and the U.S. Bureau of Economic Analysis.

Given that the time heterogeneity of the estimators can be modeled only through the underlying parameters of the joint distribution of the data, we studied the individual time consistency properties of the three time-series variables involved in the model to detect structural breaks and structural changes. To reach this goal, we conducted Non-Overlapping Window Estimator Tests as described in Koutris et al. (2008) over the entire sample size. This approach utilizes the principle of maximum entropy to

[26] $B_{k,t_i}=\sum_{j=0}^{k}\beta_j\binom{k}{j}t_i^j\left(1-t_i\right)^{k-j}$, where $\left\{\beta_j\right\}_{j=1,2,\ldots,k}$ are unknown constant model parameters.

construct a bootstrap sample of the observations in each window and provide the researcher with more precise estimates of the underlying parameters of the distribution of the process over each individual window. These estimates can then be tested for time heterogeneity. We selected a windows size of six observations to generate the maximum entropy bootstrap samples. Our method to detect the breaks consisted on finding the largest Chow statistic for structural break using all the windows and then within each windows. Using this approach we discovered that the first two structural breaks in the gasoline consumption series and the first two structural breaks in the price series variables overlapped, with no apparent overlapping breaks in the income variable. The first break occurred on January, 1979. The second break occurred on August 1985, creating a sub-sample that spawned until September of 1986. Thus, the sample size for this empirical work is from January 1974 to September 1986. The period of consideration encompasses several major events that potentially affected the world oil market: (1) The end of the Arab-Israeli War and the Arab Oil Embargo (10/73-3/74); (2) The civil war in Lebanon with the disruption to the Iraqi exports (4/76-5/76), (3), The damage to Saudi oil fields (5/77); (4) The Iranian Revolution (11/78-4/79); and (5) The outbreak of the Iran-Iraq war (10/80-12/80).

4.4.2 Empirical Results

We estimated the following model for the logarithm of gasoline consumption on the logarithm of price and the logarithm of income allowing the estimate of the price elasticity to adjust structurally in the three periods of consideration. We tested the redundancy of the periods and decided to account for a single structural change in the sample, encompassing the outset of the Iranian Revolution and the Iran-Iraq conflict. Thus, we have two different estimates for the price elasticity, one for the period Jan:1974 - Jan:1979 and one from the period Feb:1979 - Sep:1986. No statistically significant breaks were found for the income elasticity. After allowing the structural breaks in the first conditional moment equation, we tested for additional

sources of heterogeneity, additional sources of nonlinearities and additional sources of autocorrelation. The resulting statistically adequate model also includes time trends and a third order lagged variable for gasoline consumption as well as the monthly heterogeneity variables that capture the heterogeneity through the i's. The results are summarized in table,

As expected, the price elasticities and the income elasticities are statistically significant at a 5% significance level and have the correct signs. The income elasticity is positive whereas the estimated demand appears to be relatively price-inelastic becoming more inelastic after the first period under investigation. Additionally, notice how gasoline demand is expected to be higher during the summer months than during the winter months, as we would expect. The following figure shows the evolution of the gasoline estimates during the sample using the base specification (Equation 4-1) and the statistically adequate specification (left panel and right panel, respectively). Additionally, notice the relative homogeneity of the standardized residuals in our specification.

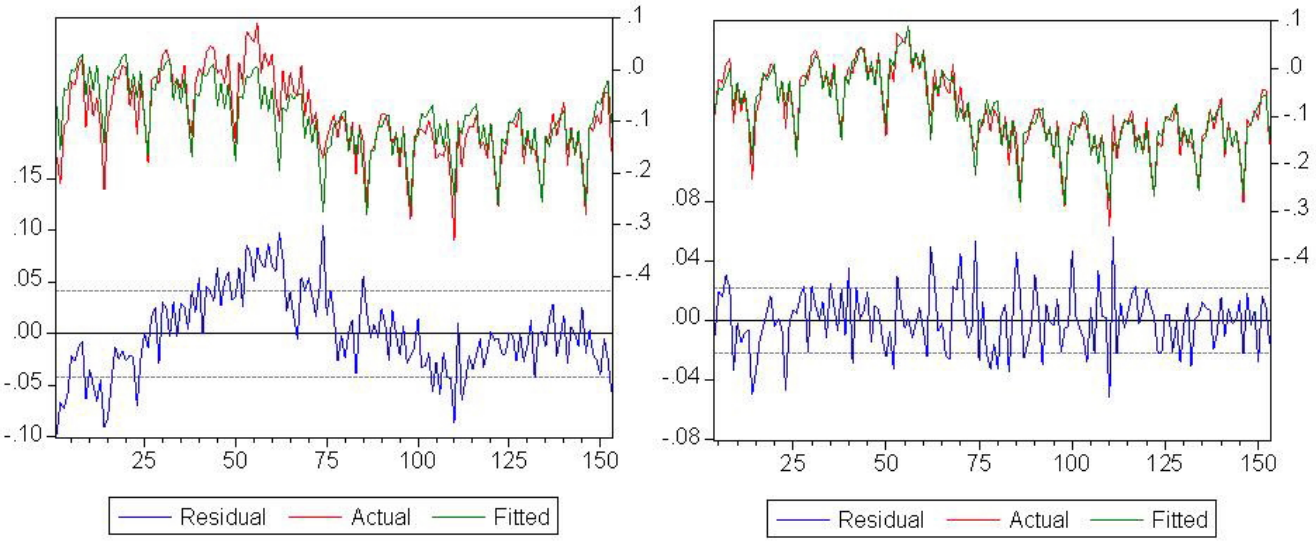


Figure 4.2: Simple Double-log Specification vs. Statistically Adequate Model

The equation for the Gasoline Consumption per Capital in levels is then given by the following function,

$$G_{it} = \begin{cases} 52.85 Y_{it}^{0.4398} P_{it}^{-0.1592} G_{it-3}^{0.3754} \epsilon_i \varepsilon_{it} & \text{Jan:1974-Jan:1979} \\ 52.85 Y_{it}^{0.4398} P_{it}^{-0.0621} G_{it-3}^{0.3754} e^{-.0014t} \epsilon_i \varepsilon_{it} & \text{Feb:1979-Sep:1986} \end{cases}$$

4.4.3 Misspecification Testing

To further assess the statistical adequacy of our model, we conducted several misspecification tests. The model withstood several misspecification tests, including autocorrelation (up to six lags), several forms of heteroskedasticity, Normality, and a Non-parametric test for independence and identical distributions (see Appendix 4.B).

4.5 Discussion and Conclusion

Our estimates of elasticities are at odds with previous estimates. However, this should not be surprising. Drollas (1984) reviewed published gasoline elasticity of demand literature for the time period 1950-1972. He concluded that the "price elasticities of demand obtained from the time series models suggest that although gasoline demand is price inelastic, it is not far from possessing unitary elasticity." Our estimates reveal relatively more inelastic values than those reported by Drollas. However, our two periods under consideration coincide with two major events in the world oil market that alters the probabilistic structure of the price-consumption relationship: (1) The Arab Oil Embargo and (2) The Iranian Revolution paired with the Iran-Iraq War (Figure 4.3). It is possible to conclude that the reason for the change in the nature of the relationship between price and consumption is mainly due to shocks in price. At the outset of these events, the market was characterized by a severe negative shock to the oil supply due to the sharp reduction in supply and then an additional shock to supply starting at the outset of the Iranian Revolution. Notice how the end of our

sample period also coincides with the end of the Iran-Iraq conflict. Of course, the evolution of the elasticity over the two periods will be different due to the way the government and the public reacted to both shocks.

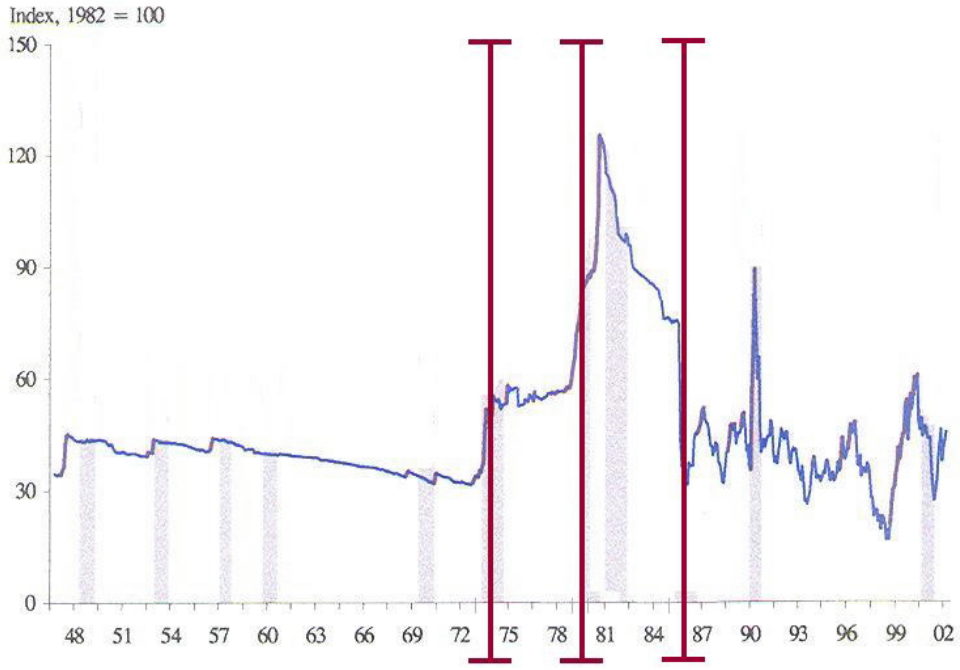


Figure 4.3: Evolution of the Price of Oil and Structural Breaks.

4.5.1 Period I: Jan 1974 - Jan 1979

Immediately after the end of Bretton-Woods and the consequent devaluation of the dollar, there was a call for changing the currency to price oil. By 1973, OPEC oil producers possessed most of the world's surplus crude oil production capacity. World crude oil prices were denominated in US dollars, and since the USA pulled out of the Bretton Woods system on August 15, 1971, the US dollar depreciated relative to other world currencies. Western countries were increasing crude oil consumption at a rate of 5% per year, crude oil prices were falling, and prices of imported goods to OPEC countries were continuously rising. These economic pressures, plus political pressures from the Yom Kippur War caused Arab members of OPEC to embargo

oil imports to the USA between October and December 1973. The embargo, which ended in January 1974, coincides with the outset of this first period. The shock from the embargo sent crude oil prices souring from $3 per barrel to $12 per barrel (in nominal terms), an increase of almost 300%. Nevertheless, regular gasoline prices rose from 38.5 cents in May 1973 to 55.1 cents in June of 1974 (in nominal terms), an increase of almost 43%, only a fraction of the increase in oil prices. The relatively small gasoline price increase was due to price controls. Additionally, psychological fears of oil shortages far exceeded the actual cutback in oil imports because of US inventory buildup earlier in 1973. It is argued that after this even OPEC succeeded in establishing a market sharing and price fixing cartel among oil producing countries.

Price controls on both crude oil and gasoline in the USA were imposed during the 1973-1981 time period as a response to the economic and psychological aftermath of the oil embargo. Price controls were harsher and more strict during the first half of our period I to be relatively more relaxed at the outset of our period II and completely abolished by the half of the same period. By January 1981 crude oil and gasoline price controls were completely abolished in the USA. Figure 4.4 shows the impact of price controls on domestic crude oil prices.

Price controls in the USA gasoline market, coupled with allocated gasoline supplies to different geographic regions, impeded the market's ability to let prices equate supply and demand, resulting in long lines due to gasoline shortages. Black markets for gasoline developed. By way of contrast, Germany and Japan did not experience long gasoline lines in the 1970's even though they were completely dependent on crude oil imports. Price controls and regulation made the US public think and feel crude oil and gasoline shortages were here to stay. Arguably, the price system was not allowed to function (Friedman, Milton and Rose, 1982; and Rockoff and Hugh, 2008).

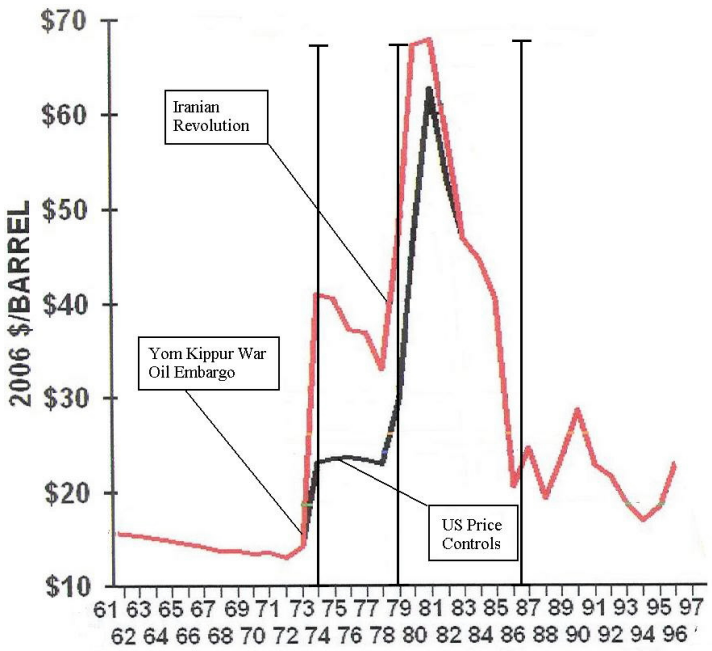


Figure 4.4: International vs. Domestic Price of Oil and Structural Breaks

4.5.2 Period II: Feb 1979 - Sep 1986

The second period was characterized by three different events that simultaneously contributed to another structural break in the price elasticity of the gasoline consumption, being the first event an additional shock to the world oil market. During 1979-1980, the world experienced a second and more severe oil shock in the supply of oil. In 1979, the Iranian Revolution caused crude oil production to fall by 3.5 millions of barrel of oil, an increase of almost 35% in the shortage created by the oil embargo. The supply of oil continued being interrupted throughout the revolution and then during the Iran-Iraq Conflict. Even after OPEC partially offset the decline, net crude oil production still decreased by about 4%. In 1980, the Iraq/Iran war haltered Iranian crude oil production and severely reduced Iraq's output. This caused crude oil prices to increase from $14 per barrel in 1978 to $35 per barrel in 1981

(in nominal terms), an increase of 150%. Over the same period USA gasoline prices increased 160%. This larger increase in the domestic price of gasoline was due to the second event, the softening of the price controls during the period that eventually led to their abolition.

The third event was the change in consumer behavioral patterns. The time period 1979-1986 had a far greater percentage of the US population living in suburbs surrounding cities than at any other earlier period. In the suburbs realm, there are few close transportation substitutes due to the lack of public transportation, compared to larger cities. Additionally, the fact that more families in the suburbs tend to have two income earners causes people to be less responsive to gasoline price increases. A larger percentage of the population drove. All these longer-term reasons cause people to need to drive no matter the cost of gasoline. Gasoline became less price elastic as suburbs grew.

4.6 Appendix 4.A: Tables

Table 4.A1: OLS Regression Results			
β_0	3.948	0.715	0.000
$\ln Y_t$	−0.159	0.065	0.016
$\ln P_t$	0.440	0.081	0.000
Jan	−0.083	0.007	0.000
Feb	−0.139	0.010	0.000
Mar	−0.030	0.009	0.002
Apr	−0.007	0.008	0.400
May	0.059	0.010	0.000
Jun	0.011	0.007	0.151
Jul	0.032	0.006	0.000
Aug	0.023	0.008	0.005
Sep	−0.057	0.008	0.000
Oct	−0.022	0.008	0.009
Nov	−0.075	0.010	0.000
$\ln G_{t-3}$	0.375	0.063	0.000
$\ln P_t \times (D_1 + D_2)$	0.097	0.047	0.039
$D_1 \times t$	−0.001	0.000	0.000
$D_2 \times t$	−0.001	0.000	0.000
R^2	0.934	F-statistic	110.82
$\overline{R}^2$	0.926	Prob(F)	0.000
D-W	2.142	BIC	−4.355

Table 4.A2: Battery of Misspecification Testing			
Serial Correlation LM Test			
F-statistic	0.979	Prob.	0.441
nR^2	6.687	Prob.	0.350
ARCH Test			
F-statistic	0.815	Prob.	0.559
nR^2	4.966	Prob.	0.548
White Heteroskedasticity Test			
F-statistic	1.653	Prob.	0.042
nR^2	34.78	Prob.	0.054
Normality			
Jarque-Bera	0.576	Prob.	0.749
Levene IID			
Statistic	6.168	Prob.	0.186
Max Run	4.00	Actual	Expected
R_1		64	62.5
R_2		36	27.2
R_3		3	7.7
R_4'		1	2.0

4.7 Appendix 4.B: Non-parametric Randomness Test

To assess randomness in the estimated residuals (a term that encompasses the statistical notions of independence and identical distributions), we derived a non-parametric test, following Spanos 1999, based on Levene, 1944.

Let $\mathbf{Y} := (Y_1, Y_2, \ldots Y_n)$ be a sample whose randomness we are trying to evaluate. Define a series of values representing the sign of the difference from one observation in the sample to the previous one, $d_{n-1} = y_n - y_{n-1}$ and concentrate exclusively in the number of subsequences of one type (pluses only or minuses only) immediately preceded and succeeded by an element of the other type. These subsequences will be defined as runs.

By treating the sequence of pluses and minuses as Bernoulli trials, Levene (1942) used combinatorial arguments to evaluate the number of runs expected if the observations were indeed independent. Let us define the following random variables based on the number of runs,

$$\begin{aligned} R: &\quad \text{number of runs of any size} \\ R_k: &\quad \text{number of runs of size } k \\ R'_k: &\quad \text{number of runs of size } k \text{ or greater,} \end{aligned}$$

Levene (1942) showed that

$$\begin{aligned} E(R) &= \tfrac{2n-1}{3} \\ E(R_k) &= 2\left[n\left(\tfrac{k^2+3k+1}{(k+3)!}\right) - \left(\tfrac{k^3+3k^2-k-4}{(k+3)!}\right)\right],\ k \leq (n-2) \\ E(R'_k) &= 2\left[n\left(\tfrac{k+1}{(k+2)!}\right) - \left(\tfrac{k^2+k-1}{(k+2)!}\right)\right],\ k \leq (n-2) \end{aligned}$$

For large enough sample size, the standardized form of these random variables can be shown to be approximately standard normal distributed. The test based on R takes the form $\mathbb{P}(|Z_R| > z_R;\ H_0 \text{ is valid}) = p$, where

$$Z_R = \frac{R - E(R)}{\sqrt{Var(R)}} = \frac{R - \frac{2n-1}{3}}{\sqrt{\frac{16n-29}{90}}} \underset{\alpha}{\overset{H_0}{\sim}} \mathsf{N}(0,1)$$

It is important to note that these test will be sensitive to departures from both the independence and the identical distribution assumptions.

References

[1] Abramowitz, M. and Stegun, I. A. (Eds.). "Psi (Digamma) Function." §6.3 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 258-259, 1972.

[2] Alston, Julian M., and James A. Chalfant. (1991) "Can We Take the Corn Out of Meat Demand Studies?" West. J. Agr. Econ., Vol. 16, No. 1, pp. 36-48.

[3] Andrews, W.K., (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point", Econometrica, 61, 821-856. (Corrigendum, 71, 395-397)

[4] Andrews, W.K., Ploberger W., (1994), "Optimal Tests when a Nuisance Parameter is Present only Under the Alternative", Econometrica, 62, 1383-1414.

[5] Arnold, B.C., E. Castillo and J.M. Sarabia. Conditional Specification of Statistical Models. New York: Springer Verlag, 1999.

[6] Bai, J., Perron P., (1998), "Estimating and Testing Linear Models with Multiple Structural Changes", Econometrica, 66, 47-78.

[7] Bergtold, Jason; Aris Spanos; Ebere Onukwugha (2005), "Bernoulli regression models: Revisiting the specification of statistical models with binary dependent variables," Annual meeting, American Agricultural Economic Association.

[8] Bernanke, Ben S. (2006). Speech before the Economic Club of Chicago: Energy and the Economy.

[9] Blaug, Mark (1980), The methodology of economics, or, How economists explain, Cambridge University Press.

[10] Bonett, Douglas G. and Edith Seier (2003), "Confidence intervals for mean absolute deviations," The American Statistician, 57, 4, 223.

[11] Box, G.E.P. and G.M. Jenkins (1970/1976), Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco.

[12] Brown, R., L., Durbin J., Evans, M., (1975), "Techniques for testing the Constancy of Regression Relationships over Time", Journal of the Royal Statistical Society, B 37, 149-192.

[13] Brown, Stephen; Yucel, Mine; Thompson, John (2003) "Business Cycles: The Role of Energy Prices," Research Department Working Paper 0304, Federal Reserve Bank of Dallas.

[14] Chalmers, A.F. (1999), What Is This Thing Called Science? Hackett Pub Co; 3 ed.

[15] Chan, Siew-Pang (2006), "Beta regression analysis of Singapore general election results," Proceeding of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications, Universiti Sains Malaysia, Penang.

[16] Chow, G.C., (1960), "Tests of Equality Between Sets of Coefficients In Two Linear Regressions", Econometrica, 28, 591-605.

[17] Colander, D. (2009). "Economists, incentives, judgment, and the European CVAR approach to macroeconometrics," Economics: the open access, open-assessment E-journal, Vol. 3.

[18] Cox, D. R. (1990). "Role of models in statistical analysis," Statistical Science, Vol. 5, pp. 169-174.

[19] Cribari-Neto, Francisco and Klauss L.P. Vasconcellos (2002), "Nearly unbiased maximum likelihood estimation for the beta distribution," Journal of Statistical Computation and Simulation, Volume 72, Number 2, 107-118.

[20] D'Agostino, Ralph B.; Albert Belanger, and Ralph B. D'Agostino, Jr. (1990), "A Suggestion for Using Powerful and Informative Tests of Normality," The American Statistician, November 1990, Vol. 44, No. 4

[21] Dougherty, Christopher (2002), Introduction to Econometrics, Oxford University Press, USA; 2 edition.

[22] Drollas, Leonidas P. (1984). "The Demand For Gasoline, Further Evidence." Energy Economics.

[23] Eichner, Alfred S. (1983), (ed.) Why economics is not yet a science, M.E. Sharpe.

[24] Energy Information Administration (2006). Annual Oil Market Chronology: 1970 - 2006 .

[25] Espinheira, Patricia; Silvia L.P. Ferrari; Francisco Cribari-Neto (2008) "On beta regression residuals," Journal of Applied Statistics, Vol. 35, No. 4, 407-419.

[26] Ferrari , Silvia L.P. and Francisco Cribari-Neto (2004), "Beta regression for modeling rates and proportions," Journal of Applied Statistics, Vol. 31, No. 7, 799-815.

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

[28] Friedman, M. (1953), "The Methodology of Positive Economics," in Essays in Positive Economics, Chicago University Press, Chicago.

[29] Friedman, Milton and Anna J. Schwartz, (1982), Monetary Trends in the United States and the United Kingdom, University of Chicago Press.

[30] Friedman, Milton and Rose (1979). Free To Chose. PBS.

[31] Gardner L. A. (1969) "On Detecting Changes in the Mean of Normal Variates", The Annals of Mathematical Statistics, vol 40, 116-126

[32] Greene, W. H. (2000), Econometric Analysis, 4th ed., Prentice Hall, NJ.

[33] Gupta, Arjun K. and Saralees Nadarajah (Editors) (2004), Handbook of Beta Distribution and Its Applications, Statistics: A Series of Textbooks and Monographs, CRC Press, 1 edition.

[34] Hansen E. Bruce (1991) "A Comparison of Tests for Parameter Instability: An Examination of Asymptotic Local Power"

[35] Hansen E. Bruce (2000) "Testing for Structural Change in Conditional Models", Journal of Econometrics, vol 97, 93-115

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

[37] Hendry, D.F. (2003), "J. Denis Sargan and the Origins of LSE Econometric Methodology," Econometric Theory, 19, 457-480.

[38] Hoover, K., Johansen, S., and Juselious, K. (2008). "Allowing the data to speak freely: The macroeconomics of the cointegrated vector autoregression," American Economic Review, 98, pp. 251-255.

[39] 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.

[40] Hughes, Jonathan; Christopher R. Knittel; Daniel Sperling (2008). "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand," The Energy Journal.

[41] Ireland, P. (2004), "Technology Shocks in the New Keynesian Model," NBER Working Paper, No. 10309, National Bureau of Economic Research, Cambridge.

[42] Johansen, S. (2007). "Confronting the economic model with the data," in D. Colander, Post Walrasian Macroeconomics: Beyond the DSGE Model. Cambridge University Press.

[43] Johnson , Norman L.; Samuel Kotz; N. Balakrishnan (1995), Continuous Univariate Distributions, Vol. 2, Wiley Series in Probability and Statistics, Wiley-Interscience; 2 edition.

[44] Johnston, J and John Dinardo (1996), Econometric Methods, McGraw-Hill/Irwin; 4 edition

[45] Johnston, J. (1963), Econometric Methods, McGraw-Hill, New York.

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

[47] Jones, M.C. (2002), "The complementary beta distribution," Journal of Statistical Planning and Inference, 104, 329-337.

[48] Juselious, K. and Franchi, M. (2007). "Taking a DSGE model to the data meaningfully," Economics: the open access, open-assessment E-journal.

[49] Kay, R. and S. Little. "Transformations of the Explanatory Variables in the Logistic Regression Model for Binary Data." Biometrika. 74(September, 1987): 495 – 501.

[50] Keeley, Sam R. (1976), "A Note on Estimating the Temporal Stability of Reaction Functions of the Monetary Authorities." Journal of Monetary Economics 2.

[51] Kennedy, P. (2008), A Guide to Econometrics, 6th edition, MIT Press, MA.

[52] Kotz, Samuel and Johan Rene van Dorp (2004), Beyond beta: Other continuous families of distributions with bounded support and applications, World Scientific Publishing Company.

[53] Kotz, Samuel; N. Balakrishnan; Norman L. Johnson (2000), Continuous Multivariate Distributions, Volume 1, Models and Applications, Wiley-Interscience; 2nd edition.

[54] Koutris, Andreas; Maria S. Heracleous; Aris Spanos (2008), "Testing for Nonstationarity Using Maximum Entropy Resampling: A Misspecification Testing Perspective ," Econometric Reviews, Volume 27, Issue 4 - 6, pages 363 - 384 .

[55] Kydland, F. and P. Prescott (1991), "The Econometrics of the General Equilibrium Approach to Business Cycles," Scandinavian Journal of Economics, 93.

[56] Lazzari, Salvatore (1990). The Windfall Profits Tax on Crude Oil: Overview of the Issues. CRS Report for Congress, September 12.

[57] Leamer, Edward E. (1978), Specification Searches: Ad Hoc Inference with Nonexperimental Data, New York: John Wiley & Sons.

[58] Lehmann, E. L. (1990). "Model specification: the views of Fisher and Neyman, and later developments," Statistical Science, Vol. 5, pp. 160-168.

[59] Leontief W.W. (1971), "Theoretical Assumptions and Nonobserved Facts," American Economic Review, 1-7.

[60] Lester, C. Thurow (1983), Dangerous Currents: The State of Economics, Random House, Inc., New York.

[61] Levene, H. and J. Wolfowitz (1944), "The Covariance Matrix of Runs Up and Down," Ann. Math. Statist. Volume 15, Number 1: 58-69.

[62] Lucas, R.E. and T.J. Sargent (1981), Rational Expectations and Econometric Practice, Allen & Unwin, London.

[63] Luenberger, G.M. (1969). Optimization by Vector Space Methods. Wiley, New York.

[64] Mayo, D. G. (1996) Error and the Growth of Experimental Knowledge, The University of Chicago Press, Chicago.

[65] McGarvey, Ronald G.; Enrique del Castillo; Tom M. Cavalier; E. Amine Lehtihet (2002), "Four-parameter beta distribution estimation and skewness test," Quality and Reliability Engineering International, 18: 395-402.

[66] McGuirk, A. and Spanos, A. (2009). "Revisiting error-autocorrelation correction: common factor restrictions and Granger non-causality," Oxford Bulletin of Economics and Statistics, Vol .71, pp. 273-294.

[67] McGuirk, A., Driscoll, P., and Alwang, J. (1993). "Misspecification testing: a comprehensive approach," Amer. J. Agr. Econ, Vol. 75, pp. 1044-1055.

[68] Mill, J.S. (1844), Essays on Some Unsettled Questions of Political Economy, 2nd ed., Longmans, London.

[69] Minhajuddin, Abu T.; Ian R. Harris; William R. Schucany (2004), "Simulating multivariate distributions with specific correlations," Journal of Statistical Computation and Simulation, Volume 74, Issue 8, 599 – 607.

[70] Nyblom, J., (1989), "Testing for the Constancy of Parameters Over Time", Journal of the American Statistical Association, 84, 223-230.

[71] Olkin, Ingram and Ruixue Liu (2003), "A bivariate beta distribution," Statistics & Probability Letters, 62, 407-412.

[72] Ospina, Raydonal; Francisco Cribari-Neto; Klaus L.P. Vasconcellos (2006), "Improved point and interval estimation for a beta regression model," Computational Statistics & Data Analysis, 51, 960-981.

[73] Paolino, Philip (2001), "Maximum likelihood estimation of models with beta-distributed dependent variables," Political Analysis, 9:4.

[74] Ploberger Werner, Kramer Walter (1992) "The Cusum Test with Ols Residuals", Econometrica, vol 60, 271-285

[75] Ploberger Werner, Kramer Walter (1996) "A Trent-Resistant Test for Structural Change Based on Ols Residuals", Journal of Econometrics, vol 70, 175-185

[76] Poole, William (2007). Energy Prices and the U.S. Business Cycle. Global Interdependence Center Abroad in Chile Conference. March 2.

[77] Quandt, R.E., (1960), "Tests of the hypothesis that a linear regression system obeys two separate regimes", Journal of the American Statistical Association, 55, 324-330.

[78] Rockoff, Hugh. (2008). Price Controls. The Concise Encyclopedia of Economics.

[79] Romero, A.A. (2009a), "Are we teaching students to do the wrong things? Misspecification testing on an array of econometrics textbook examples," Working Paper, Virginia Tech and The College of William and Mary.

[80] Romero, A.A. (2009b), "Does the conditioning set matter for Monte Carlo?" Working Paper, Virginia Tech and The College of William and Mary.

[81] Romero, A.A. (2009c), "Elasticities and lognormal regression models." Working Paper, Virginia Tech and The College of William and Mary.

[82] Romero, A.A. (2010), "A simulation analysis of the small sample properties of the beta distribution," Working Paper, Virginia Polytechnic Institute and State University.

[83] Roubini, Nooriel and Setser, Brad (2004), "The Effects of the Recent Oil Price Shock on the US And Global Economy," Unpublished Draft.

[84] Sargan, J.D. (1964) "Wages and prices in the U.K.: a study in econometric methodology," in P.E. Hart et al. (eds.) Econometric Analysis for National Planning, pp.25-59. London: Butterworth.

[85] Simas, Alexandre B.; Wagner Barreto-Souza; Andrea V. Rocha (2010), "Improved estimators for a general class of beta regression models," Computational Statistics & Data Analysis, Volume 54, Issue 2, 348-366.

[86] Sims, C.A. (1980), "Macroeconomics and Reality," Econometrica, 48, 1-48.

[87] Sims, C.A., (1982), "Policy Analysis with Econometric Models," Brookings Paper on Economic Activity, 107-164.

[88] Smithson, Michael and Jay Verkuilen (2005) "Beta regression: Practical issues in estimation," Working Paper, The Australian National University and University of Illinois at Urbana-Champaign.

[89] Smithson, Michael and Jay Verkuilen (2006), "A better lemon squeezer? Maximum likelihood regression with beta-distributed dependent variables," Psychological Methods, Vol. 11, No. 1, 54-71.

[90] Solow, R.M. (2007), "Comment on Colander's Survey," in D. Colander (ed.), The Making of an Economist Redux, Princeton University Press, Princeton, N.J.

[91] Spanos, A. (1995). "On theory testing in econometrics: modeling with nonexperimental data," Journal of Econometrics, Vol. 67, pp. 189-226.

[92] Spanos, A. (1999), Probability Theory and Statistical Inference: Econometric Modeling with Observational Data, Cambridge University Press, Cambridge.

[93] Spanos, A. (2000), "Revisiting Data Mining: 'hunting' with or without a license," The Journal of Economic Methodology, 7: 231-264.

[94] Spanos, A. (2006a). "Where do statistical model come from? Revisiting the problem of specification," IMS lecture notes-monograph series, Vol. 49, pp. 98-119.

[95] Spanos, A. (2006b), "Econometrics in Retrospect and Prospect," pp. 3-58 in Mills, T.C. and K. Patterson, New Palgrave Handbook of Econometrics, vol. 1, MacMillan, London.

[96] Spanos, A. (2007), "Using the Residuals to Test the Validity of a Statistical Model: Revisiting Sufficiency and Ancillarity," Working paper, Virginia Tech.

[97] Spanos, A. (2009). "The pre-eminence of theory versus the European CVAR perspective in macroeconometric modeling," Economics: the open access, open-assessment E-journal, Vol. 3.

[98] Spanos, A. and A. McGuirk (2001), "The Model Specification Problem from a Probabilistic Reduction Perspective," Journal of the American Agricultural Association, 83: 1168-1176.

[99] Spanos, A., (1986) Statistical Foundations of Econometric Modeling, Cambridge University Press, Cambridge.

[100] Spanos, Aris (2001), "The beta regression-like model," Working Paper, Virginia Polytechnic Institute and State University.

[101] Taylor, Jerry and Van Doren, Peter (2003) "Time to Lay the 1973 Oil Embargo to Rest," CATO Institute.

[102] Thomopoulos, Nick T. and Arvid C. Johnson (2004), "Some measures on the standard bivariate lognormal distribution," Decision Sciences Institute Proceedings, Boston.

[103] U.S. Department of Labor (2008). Consumer Price Index.

[104] Vinod, H.D., (2004), "Ranking Mutual Funds Using Unconventional Utility Theory and Stochastic Dominance,"Journal of Empirical Finance, Vol. 11(3) 2004, pp. 353-377.

[105] Williams, James L. (2007) "Oil Price History and Analysis." WTRG Economics.

[106] Wold, Herman(1938). A Study in the Analysis of Stationary Time Series. Almquist and Wicksell, Uppsala.

[107] Wolfowitz, J. (1944), "Asymptotic Distribution of Runs Up and Down," The Annals of Mathematical Statistics, Vol. 15, No. 2: 163-172 .

[108] Wooldridge, J. M. (2006), Introductory Econometrics: a modern approach, Thomson, South-Western.

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

[110] Yue, Sheng (2000), "The bivariate lognormal distribution to model a multivariate flood episode," Hydrological Processes, 14, 2575-2588.

[111] Yue, Sheng (2002), "The bivariate lognormal distribution for describing joint statistical properties of a multivariate storm event," Environmetrics, 13:811-819.