INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
RICE UNIVERSITY

Essays on Parametric and Nonparametric Modeling and Estimation with Applications to Energy Economics

by

Weiyu Gao

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE

Bryan W. Brown, Chairman
Hargrove Professor of Economics

Robin C. Sickles, Co-Chairman
Professor of Economics

Peter R. Hartley
Professor of Economics

David W. Scott
Professor of Statistics

Houston, Texas
April, 1999
ABSTRACT

Essays in Parametric and Nonparametric Modeling and Estimation with Applications to Energy Economics

by

Weiyu Gao

Nonparametric and semiparametric modeling and estimation procedures are now widely applied in econometrics. Their popularity mainly comes from the reduction of the probability of misspecification compared with their parametric counterpart. My dissertation research is composed of two parts: a theoretical part on semiparametric efficient estimation of partially linear model and an applied part in energy economics under different dynamic settings. The essays are related in terms of their applications as well as the way in which models are constructed and estimated. In the first essay, estimation of the partially linear model is studied under different stochastic restrictions of the residual term. We work out the efficient score functions and efficiency bounds under four stochastic assumptions - independence, conditional symmetry, conditional zero mean, and partially conditional zero mean. A feasible efficient estimation method for the linear part of the model is also developed based on the efficient score function associated with each parametric submodel. A battery of specification test that allows for choosing between alternative assumptions is provided. A Monte Carlo simulation is also conducted to contrast and compare the finite sample properties of the efficient estimator with those of the Robinson's estimator.
The second essay presents a dynamic optimization model for a stylized oilfield resembling the largest developed light oil field in Saudi Arabia, Ghawar. We use data from different sources to estimate the oil production cost function and the revenue function that constitute part of our dynamic programming model. We pay particular attention to the dynamic aspect of the oil production by employing a petroleum engineering software to simulate the interaction between production control variables and reservoir state variables. A nonparametric smoothing technique (tensor spline) is employed to approximate the value function. Optimal solutions are studied under different scenarios to account for the possible changes in the exogenous variables and the uncertainty about the forecasts. The model is based on profit maximization hypothesis. While Saudi oil policy is likely to reflect many political and strategic motives, our analysis is nevertheless instructive in that it enables one to quantify the cost of pursuing these non-economic objectives.

The third essay examines the effect of oil price volatility on the level of innovation displayed by the U.S. economy. A measure of innovation is calculated by decomposing an output-based Malmquist productivity index. We also construct a nonparametric measure for oil price volatility. Technical-change and oil price volatility are then placed in a VAR framework with oil price and a variable indicative of monetary policy. The system is estimated and analyzed for significant relationships. We find that oil price volatility displays a significant negative effect on innovation. A key point of this analysis lies in the fact that we impose no functional forms for production technologies and the methods we employ keep technical assumptions to a minimum.
Acknowledgements

It is a pleasure to express my appreciation to my two advisors: Prof. Bryan W. Brown and Prof. Robin C. Sickles, for their patient and helpful guidance through my graduate studies at Rice. Without many discussions I had with them and their constructive and valuable advice, it would be almost impossible for me to have completed this research -- for this I feel grateful. I would also like to thank Prof. Peter R. Hartley for his insightful comments and suggestions and his very careful reading of one part of this dissertation. I also want to thank Prof. David W. Scott because he kindly agreed to sit on my graduate committee.

A special recognition goes to my classmates and officemates, in particular, Kenneth B. Medlock III. I certainly benefited from the discussions with Ken and appreciate the help he provided when I needed it.

I also wish to extend my gratitude to Prof. Yasar Barut, Prof. Yoosoon Chang, Prof. Mahmoud A. El-Gamal, Prof. Kevin Hasker, Prof. Monika Gehrig-Merz, and Prof. Yuka Ohno for their encouragement and help during my studies at Rice.

Last but not least, I want to thank my parents and my wife for their support. My parents helped and encouraged me during the most difficult period of my life. Without their love, encouragement, and support, I would never have had the necessary energy and courage to pursue my graduate study. My appreciation and love to my wife, Xiaoling, for her love, understanding, and continuous support.
Contents

1. Introduction 1

1.1. Parametric and Nonparametric Modeling and Estimation 1
1.2. Dissertation Structure 5

2. Semiparametric Efficient Estimation of Partially Linear Model 6

2.1. Introduction 6
2.2. Model and Previous Results 9
2.3. Semiparametric Efficiency Bounds 11
  2.3.1. Basic Concepts of Semiparametric Efficiency 12
  2.3.2. Semiparametric Efficiency Bound for Independence 15
  2.3.3. Semiparametric Efficiency Bound for Conditional Symmetry 16
  2.3.4. Semiparametric Efficiency Bound for Conditional Zero Mean 18
  2.3.5. Semiparametric Efficiency Bound for Partially Conditional Zero Mean 19
2.4. Feasible and Efficient Estimation 20
  2.4.1. Basic Approach for Feasible Estimation 21
  2.4.2. Feasible and Efficient Estimation under Independence 25
  2.4.3. Feasible and Efficient Estimation under Conditional Symmetry 26
  2.4.4. Feasible and Efficient Estimation under Conditional Zero Mean 28
2.5. Specification Tests 29
2.6. Sampling Results 31
2.7. Conclusion 39

3. Optimal Dynamic Production Policy: The Case of a Large Oil Field in Saudi Arabia 41

3.1. Introduction 41
3.2. Dynamic Modeling of Oil Production Decisions 44
3.3. Data and Estimation
   3.3.1. The Revenue Function 48
   3.3.2. The Production Cost Function 50
   3.3.3. The Dynamic Production Function 52
   3.3.4. Summary 59
3.4. Theoretical Issues 60
3.5. Simulation Results 63
3.6. Conclusion 73

4. The Effects of Oil Price Volatility on Technical Change 76
   4.1. Introduction 76
   4.2. Channels of Transmission and Oil Price Volatility 78
   4.3. A Measure of Volatility 82
   4.4. A Measure of Innovation (Technical Change) 85
   4.5. Vector Autoregression 92
   4.6. Conclusion 102

5. References 105

6. Appendix A 116

7. Appendix B 127
List of Tables

2.1 Simulation Results under Independence: Model (I) 33
2.2 Simulation Results under Independence: Model (I) 33
2.3 Simulation Results under Independence: Model (II) 34
2.4 Simulation Results under Independence: Model (II) 34
2.5 Simulation Results under Independence: Model (III) 36
2.6 Simulation Results under Independence: Model (III) 36
2.7 Simulation Results under Independence: Model (III) 37
2.8 Simulation Results under Independence: Model (III) 37
2.9 Simulation Results under Conditional Zero Mean: Model (I) 39

3.1 Oil Revenue and Expenditure of Saudi Arabia Government 42
3.2 Estimation of the Inverse Demand Equation 49
3.3 Estimation of the Water Injection Function 57
3.4 Estimation of the Short-term Capacity Function 58
3.5 Comparison of Simulated Oil Revenues with the True Ones 74

4.1 Productivity Decomposition under Constant Returns to Scale 91
4.2 Tests for Integration and Cointegration 94
4.3 Selected Variable Properties 95
4.4 VAR Estimation Results 96
4.5 Hypothesis Test Results 97
List of Figures

3.1 Approximated Value Function in the Time-Invariant State in Scenario I 65
3.2 Approximated Value Function in the Time-Invariant State in Scenario II 65
3.3 Approximated Value Function in the Time-Invariant State in Scenario III 66
3.4 Optimal Extraction Rates in Scenario I 67
3.5 Optimal Number of New Wells in Scenario I 67
3.6 Optimal Extraction Rates in Scenario II 68
3.7 Optimal Number of New Wells in Scenario II 68
3.8 Optimal Extraction Rates in Scenario III 69
3.9 Optimal Number of New Wells in Scenario III 69
3.10 Optimal Extraction Rates in Scenario IV 70
3.11 Optimal Number of New Wells in Scenario IV 70
3.12 Comparison of Simulated Optimal Productions with the Real Path 71
3.13 Comparison of Simulated Price Path and the Real Path 73

4.1 Oil Price and Oil Price Volatility 84
4.2 A Comparison of Volatility Measures 85
4.3 The Malmquist-based Index and Decomposition by Distance Function 87
4.4 Impulse Response Analysis 100

A1 Water Saturation - Initial State 127
A2 Gas Saturation - without Water Injection 128
A3 Water Saturation (Fraction) – with Water Injection 129
A4 Geometry of a Reservoir Wedge 130
A5 Geometry of the Stitched Wedges for Entire Reservoir 131
A6 Typical Well Production Schedule (Rates at 5%) 132
A7 Typical Well Production Schedule (Rates at 6%) 133
A8 Typical Well Production Schedule (Rates at 7%) 134
A9 Typical Well Production Schedule (Rates at 8%) 135
A10 Typical Well Production Schedule (Rates at 5%) 136
Chapter 1

Introduction

1.1 Parametric and Nonparametric Modeling and Estimation

In this section, I would like to address, in general, the motivation for parametric and nonparametric (and semiparametric) modeling and estimation and issues regarding the applications. The specific motivation and background for each individual research project is left to the introduction of each of the respective chapters.

The basic incentive for nonparametric models comes from the price that we might have to pay for using pure parametric models, i.e., the possible misspecification that could lead to highly biased results in estimation and prediction in terms of both finite sample and asymptotic properties. The nonparametric method, hence, provides an alternative estimation procedure in the absence of strong a priori restriction.

In addition, a nonparametric estimation approach may also have the following potential advantages, as summarized by Härdle (1990).

- a versatile method of exploring a general relationship between variables;
- predictions of observations to be made without reference to a fixed parametric model;
- a tool for finding spurious observations by studying the influence of isolated points;
- a flexible method of substituting for missing values or interpolating between adjacent data points.
The motivations behind using nonparametric method in this paper, apart from reducing the possibility of specification errors, lie mainly in the first two advantages listed above. For example, the Malmquist index, constructed in chapter 4 using a linear programming method, is a nonparametric productivity index, which requires no specification of a production function in constructing the productivity and innovation indices. In chapter 3, we use tensor spline to explore the relationship between the optimal value and the state variables involved in Saudi’s optimal production model, which also requires no specification of functional form for the value function. These nonparametric techniques not only lead to reduction of the approximation bias from possible misspecifications, but also furnish us with more adaptive tools to study the relationships between economic variables.

While we enjoy the advantages of nonparametric methods, the main cost associated with them, compared with a correctly specified parametric model, is the loss of efficiency, i.e., reduction in the precision of estimations and predictions. The idea to achieve balance between the bias caused by possible misspecification in the parametric technique and the poor precision generated from the nonparametric method gives rise to our interest in semiparametric modeling and estimation, which can be viewed as a hybrid of the pure parametric and nonparametric methods. The advantage of the semiparametric approach is that we can specify the part of the model that we have confidence in and focus on the estimation of “parameter of interest” and leave the part uninterested or unknown – the so called “nuisance parameter” – to be unspecified. As a combination of parametric and nonparametric approaches, semiparametric method does share the advantage and disadvantage of the two in a more balanced way, which is the reason for the popularity of semiparametric modeling.
In chapter 2, we study an important semiparametric model known as the partially linear model (PLM). The PLM is a typical example of semiparametric model, which has one part of the regressors formatted in linear form and the other part in unknown functional form. Its popularity can be attributed to its flexible functional form. We construct a consistent estimator that follows a *Normal* distribution and attains the efficiency bound in the limit. Monte Carlo simulation results also show that the estimator in general performs better in finite sample than Robinson's estimator, the most widely cited estimator of the PLM in the previous literature. Although the PLM is the only semiparametric model we study in this dissertation, it does not mean that our method is only limited to the PLM. As a matter of fact, it is possible that the general idea of deriving the semiparametric efficient estimator from efficient score can be extended to other cases.

An interesting question is how we can possibly use the PLM and the derived efficient estimation method in applications. Two famous examples are Engle, et al. (1986) and Hausman and Newey (1995). Engle et al. employ the PLM to model the relationship between electricity demand and weather. Hausman and Newey estimate the demand for gasoline in the PLM format, and then use the estimated equation to calculate the consumer surplus and deadweight loss. In the latter, only the covariates like time and region dummies are formatted in the linear part and the relationship between the dependent variable and the main regressors, such as income and gasoline price, are modeled as an unknown function. Their main interest lies in estimating the unknown function, which is a little different from our focus. We are interested in estimating the parameter in the linear part. An interesting and possible application of our method is the estimation of the price (or income) elasticity of gasoline demand, as demand may depend
on a set of variables, like price, income, the number of automobiles, etc. We can format the gasoline consumption (in log form) as the dependent variable and the gasoline price (in log form) as the regressor in the linear part leaving the other variables in the unknown function, such as $g(.)$ in equation (2.1). We can thus concentrate on the estimation of the price elasticity with the effects of the other variables relevant to gasoline consumption taken into account, thereby reducing the likelihood of specification error of the other variables in the demand equation. Although we have not pursued this study in this dissertation, the above example may be a good application of our method for future research. Furthermore, it is related to another area that this dissertation covers: Energy Economics.

We recognize the benefits of nonparametric and semiparametric methods, but it certainly does not mean that standard parametric approaches, like least square and maximum likelihood estimators, are not useful for econometric applications. On the contrary, they could be even more accurate if the specification is correct. They may also appear to be simple and convenient in many cases, especially when a closed functional form is required to study the relationship between one dependent variable and a set of regressors. For example, in chapter 3, to solve for the dynamic programming model built up for a typical light oil field in Saudi Arabia, we need closed form of its revenue, cost, and dynamic production functions in order to place them into numerical simulation for value functions and optimal policy paths. Nonparametric and semiparametric methods will only generate estimations either for conditional means or part of the closed form of these relationships. Although these estimations could still be applied in simulation, they will certainly cause inconvenience in approximation and increase the computation time due to their nonparametric features. On the other hand, experimenting with different
parametric specifications using standard least squares might be a better alternative. It generally yields good approximations to the relationship in which we are interested with explicit functional form required in later simulations. In addition, it is simpler, and saves a lot of computation time in simulation compared with some nonparametrical estimations like spline smoothed functions.

1.2 Dissertation Structure

Chapter 2 is on the semiparametric efficient estimation of partially linear model, and the rest of this dissertation is devoted to two topics of energy economics with applied econometric and dynamic programming techniques. In chapter 3, we build a dynamic programming model to study questions like: what are the optimal production rates for the major oil exporter in the world oil market - Saudi Arabia - and what is the possible cost associated with the other political and strategical objectives in its oil policy? Chapter 4 addresses the possible negative effect of oil price volatility on innovation and, hence, economics growth in a VAR system.

All the mathematical proofs for theorems in each chapter are contained in Appendix A. Some picture presentations of Black Oil simulation results (see chapter 3 for details) are included in Appendix B, which may aid in understanding the simulated reservoir engineering conditions and the estimated dynamic oil production function for a typical light oil field in Saudi Arabia.
Chapter 2

Semiparametric Efficient Estimation of Partially Linear Model

2.1 Introduction

In a semiparametric econometric model, part of the model is specified parametrically and part is specified nonparametrically. The attraction of semiparametric models in econometric research is the ability to focus on the parametric part where we have confidence in the specification. The gain from such approaches is to reduce the possibility of bias and inconsistency resulting from imposing inappropriate assumptions by allowing parts of the model to be nonparametric. The cost is, of course, the loss in efficiency from not imposing parametric restrictions on all components of the model in the estimation procedure. The art of econometrics is introduced through the intelligent choice of which parts of the model to represent in a parametric fashion and which to capture with a nonparametric specification.

In regression models, the component of the model that is typically handled in a nonparametric fashion is the distribution of the disturbance terms. Indeed, the classical results for linear least squares do not require a complete distributional specification. An extension to this model, which has received considerable attention, is the partially linear model. In this model, the dependent variable is specified as additive in a linear function of the regressors of primary interest and an unknown (possibly nonlinear) function of a second set of regressors. Thus the model is parametric (linear) in the first set of
regressors and nonparametric with respect to the second set of regressors and the distribution of the disturbances. This model is known as the partially linear model.

Early work on versions of this model includes: Cosslett (1984), Stock (1985), and Engle et. al. (1986), among others. Robinson (1988), first developed a $\sqrt{n}$ - consistent estimator of the coefficients on the linear component when the second set of regressors is stochastic and arbitrary dimension. Chamberlain (1992) has established the semiparametric efficiency bound for estimation of these coefficients when the disturbances are assumed to only satisfy conditional zero mean. Chamberlain (1992) also developed bounds for the model when the nonparametric regression function is restricted to be additively separable into two or more nonparametric functions of different sets of regressors. Li (1998) has proposed estimators that attain the latter bound when the disturbances turn out to be homoskedastic.

The purpose of this chapter is a unified treatment of semiparametric efficiency bounds and feasible estimators that attain those bounds for a number of variants of the partially linear model. We assume alternatively, that the disturbances are either stochastically independent of the regressors, conditionally symmetric given the regressors, conditionally zero mean, and uncorrelated with respect to the primary regressors but conditional zero mean given the secondary regressors. We focus on the case of a single additive nonparametric regression function, but discuss the extension to multiple additive nonparametric functions. The lower bounds for each case are established using the tangent set approach discussed by Begun, et. al. (1983), Bickel, et. al. (1992), and Newey (1990a). The feasible estimators that asymptotically attain these bounds utilize nonparametric estimation techniques to estimate certain unknown components of the efficient influence function.
This chapter is organized as follows. The next section formally presents the model and alternative assumptions along with a quick review of previous results. The efficient scores and corresponding semiparametric efficiency bounds for the model under the alternative assumptions are established in the third section. The efficient scores are utilized, in the fourth section, to develop feasible estimators, which attain the bounds asymptotically. A battery of specification tests that allow for choosing between the alternative assumptions regarding the disturbances are introduced in the fifth section. The results of a Monte Carlo study are presented in the sixth section to gain some idea of the relative finite sample behavior of the efficient estimators. In the final section, conclusions and extensions are presented.

Among the principle contributions of this study are the following. A unified treatment of the semiparametric efficiency bounds is given for all the alternative assumptions. For the conditional zero mean case this provides an alternative, simpler, approach than the multinomial approximation approach taken by Chamberlain. The Robinson estimator is shown to obtain the semiparametric efficiency bound under the assumption of disturbances uncorrelated with the primary regressors but conditional zero mean given the secondary regressors. Efficient estimation of the partially linear model under the assumption of conditional symmetry is studied for the first time. A generalized score estimator is proposed which asymptotically attains the semiparametric efficiency bound for both the conditional symmetry and independence cases. Feasible tests are proposed which enable choosing between the alternative disturbance assumptions.
2.2 Model and Previous Results

Formally, following Robinson (1988), we consider the model

\[ y = x'_1 \beta + g(x_2) + \varepsilon, \]

(2.1)

where \( z' = (y, x'_1, x'_2) \) is a \( R \times R^p \times R^q \)-valued observable random variable, \( \beta \) is a \( R^p \)-valued unknown parameter, and \( g \) is a real function with unknown functional form. This model can be restricted to the case where the unknown function \( g(x_2) \) can be written as additively separable in functions of subsets of \( x_2 \) and take the form

\[ y = x'_1 \beta + g_1(x_2) + g_2(x_3) + \varepsilon. \]

(2.2)

For purposes of expositional clarity, we will primarily focus on (2.1) in this chapter. We also restrict our attention to the case where the observable variables \( z \) and hence the generating variables \( (x', \varepsilon') \) are jointly i.i.d. The distributions of the generating variables are unrestricted, except for regularity conditions and stochastic restrictions on the dependence of \( \varepsilon \) on \( x \). The alternative stochastic restrictions considered are: (i) \( \varepsilon \) independent of \( x \); (ii) \( \varepsilon \) conditionally symmetric given \( x \); (iii) \( \varepsilon \) conditional zero mean given \( x \); and (iv) \( \varepsilon \) uncorrelated with \( x_1 \) but conditional zero mean given \( x_2 \). The model is semiparametric since the form of additive function \( g(x_2) \) and the distribution of \( \varepsilon \) are nonparametric.

Robinson (1988) provided an estimator for (2.1) that is \( \sqrt{n} \)-consistent under conditional zero mean specification (iii) and other regularity conditions. This result also requires that the dimension of the regressors \( (x_2) \) in the unspecified function is no greater than 3. Define \( S_{A-\hat{A},B-\hat{B}} = (A_1, \ldots, A_n) \times \{ \text{diag}(1, \ldots, 1) \text{D'} \text{diag}(1, \ldots, 1) \} \times (B_1, \ldots, B_n) \)' and \( S_{A-\hat{A}} = S_{A-\hat{A},A-\hat{A}} \), where \( 1_i = I(\hat{f}(x_{2i}) > b) \), \( \hat{f}(x_{2i}) = (nh^q)^{-1} \sum_{j=1}^n K_{ij} \) is the Nadaraya-
Watson density estimator of $x_2$ evaluated at $x_{2i}$ with kernel weights $K_{ij}$, $h$ is a window width parameter, $b$ is a data trimming parameter, $D = I_N - C$, and $C_{ij} = \left(K_{ij} / \hat{f}_j\right)$. Then Robinson's estimator can be written as

$$\tilde{\beta} = S^{-1}_{x-x}S_{x-y-y},$$

(2.3)

This estimator can be seen to result from least squares regression of $y - \hat{E}(y \mid x_2)$ on $x_i - \hat{E}(x_i \mid x_2)$ for the observations satisfying the trimming condition $1(\hat{f}(x_{2i}) > b)$, where $\hat{E}(\cdot \mid x_2)$ is the Nadaraya-Watson kernel estimator of the conditional moment. Given $\tilde{\beta}$, $g(x_2)$ can then be estimated by $\tilde{g}(x_{2i}) = y_i - x_i \tilde{\beta}$. This $\sqrt{n}$ - consistent estimator is easy to calculate.

Chamberlain (1987) obtains the semiparametric efficiency bound for both (2.1) and (2.2) under the assumption of conditional zero mean (iii). He establishes his result when the underlying distribution is multinomial with arbitrary finite support. Since any distribution can be approximated arbitrarily closely with a multinomial, the result may be applied to the case of continuous support. The bounds for the conditional zero mean case, as well as the conditional symmetry and independence cases, will be established directly below using the tangent set/projection approach and thus will not be given now. It should be noted that the Robinson estimator does not attain this bound for (2.1), in general. Only if it turns out that the residual $\varepsilon$ is homoskedastic with zero conditional mean or $\varepsilon$ is Normally distributed with independence given will it attain the bound.

Cuzick (1992) examined efficient estimation of (2.1) under the assumption of independence (i). He considers the efficient influence function when both the distribution of the disturbances and the conditional expectations of $y$ and $x_i$ given $x_2$. 

He then utilizes preliminary series (spline) estimators of the unknown components of the influence function and shows that the asymptotic distribution of the resulting estimator is unaffected by the use of the preliminary estimators. The approach taken below results in an estimator that is asymptotically equivalent to Cuzick’s but uses kernel techniques. Moreover, the approach yields an explicit lower bound and generalizes easily to the model given by (2.2).

Li (1996) showed that Robinson’s estimator based on the standard second-order kernel applied to the case $q \leq 5$. Recently, Li (1998) suggested a series estimation approach for the model in (2.2) under the conditional zero mean assumption (iii). He establishes the consistency of the approach and shows that his estimator attains Chamberlain’s semiparametric efficiency bound if the residual term turns out to be homoskedastic. More generally, Li’s estimator will not attain the bound.

2.3 Semiparametric Efficiency Bounds

In this section, we develop semiparametric efficiency bounds for the partially linear model under each of the assumptions introduced above. The approach taken is based on first finding the efficient score using tangent sets and projections and draws heavily on Newey (1990a). The lower bound for covariance matrices of well behaved estimators is then taken as the inverse covariance matrix of the efficient score. As an example, details on how to derive the efficient scores and semiparametric efficiency bound are provided for the independence case. The method is also applied to the other three cases to find the corresponding efficient scores and lower bounds.
2.3.1 Basic concepts of semiparametric efficiency

We first formalize the notion that the semiparametric specification is a very general parameterization so its lower bound must be at least as large as that of any restricted version of the model.

Definition 1: A parametric submodel is any member of the set of parametric submodels consistent with the parametric and semiparametric assumptions. A parametric model is regular if it is smooth with a nonsingular information matrix.

Definition 2: The Asymptotic semiparametric efficiency bound for estimation of the parameter $\beta$ is defined as the supremum of the Cramer-Rao bounds for all regular parametric submodels, and denoted as $V_\beta^*$. The asymptotic variance of any semiparametric estimator should be no smaller than $V_\beta^*$. Of course, the lower bound does not apply to superconsistent estimators, so we restrict our attention to $\sqrt{n}$-consistent estimators with some smoothness. In the following, let $\eta$ denote the nuisance parameters and $\theta' = (\beta', \eta')$ the complete parameter vector for each parametric submodel.

Definition 3: An estimator is said to be regular if for each possible sequence $\theta_n$ of each parametric submodel, such that $n^{\frac{1}{2}}(\theta_n - \theta)$ is bounded and $(z_1, \ldots, z_n)$ distributed corresponding to $\theta_n$, then $n^{\frac{1}{2}}(\hat{\beta} - \beta(\theta_n))$ has a limiting distribution that does not depend on $\{\theta_n\}$ for the parametric model.
Definition 4: An estimator is said to be efficient if it is regular and its limiting distribution is \( N(0, \mathcal{V}_\beta^*) \).

Newey (1990a) provides a way to determine whether an asymptotically linear estimator is regular or not. For a particular submodel, \( S_\beta(z) \), \( S_\eta(z) \), and \( S_\theta(z) \) denote the scores with respect to the parameters \( \beta, \eta, \) and \( \theta \).

Definition 5: An estimator, \( \hat{\beta} \), is said to be asymptotic linear if there is a function \( \psi_\beta(z) \) of the observations, denoted the influence function, such that,

\[
\frac{n}{n}(\hat{\beta} - \beta_0) = \frac{n}{n} \sum_{i=1}^{n} \psi_\beta(z_i) + o_p(1).
\]

Lemma 2.1 Suppose that (i) \( \hat{\beta} \) is asymptotically linear with influence function \( \psi_\beta(z) \), and (ii) for all regular parametric submodels \( \beta(\theta) \) is differentiable and \( E_\theta[||\psi_\beta(z)||^2] \) exists and is continuous on a neighborhood of \( \theta_0 \). Then \( \hat{\beta} \) is regular if and only if, for all regular semiparametric submodels, \( \partial \beta(\theta_0)/\partial \theta' = E[\psi_\beta(z) \cdot S_\theta(z)] \).

Proof: see Theorem 2.2 in Newey (1990).

The basis for an efficient semiparametric estimator is obtained by orthogonalizing \( S_\beta(z) \) with respect to the nuisance score, \( S_\eta(z) \), for all parametric submodels.

Definition 6: The nonparametric tangent set, \( \mathcal{T} \), is the mean square closure of all possible \( k \)-dimensional linear combination of \( S_\eta(z) \) for all regular parametric submodels satisfying the semiparametric assumptions; i.e.,
$$S = \left\{ t \in \mathbb{R}^q : E[t't] < \infty, \exists B_j, S_{\eta_j}(z), s.t. E\left[\|t - B_j S_{\eta_j}(z)\|^2\right] = o(1) \right\}.$$ 

**Definition 7:** The residual of the projection of score $S_\beta(z)$, for any parametric submodel, which includes the truth, on the parametric tangent set,

$$S(z) = S_\beta(z) - \text{proj}(S_\beta(z) | \mathcal{S})$$

is known as the efficient score for $\beta$.

Loosely speaking, the nonparametric tangent set, $\mathcal{S}$, is the linear space spanned by the score functions with respect to the nuisance parameters, $S_\eta(z)$, for all parametric submodels.

There is more than one way to find the semiparametric efficient estimator for the linear part of (2.1). The following theorem, due to Newey (1990), provides one of the feasible methods.

**Theorem 2.1** Suppose that (i) $f(z|\beta)$ is smooth with score $S_\beta(z)$; (ii) $\mathcal{S}$ is linear, and the residual $S$ of the projection of $S_\beta(z)$ on $\mathcal{S}$ satisfies $E[S(z) \cdot S(z)^\top]$ nonsingular. Then $\beta$ is differentiable with $\psi_\beta^*(z) = V_\beta^* \cdot S(z)$ and $V_\beta^* = E[S(z) \cdot S(z)^\top]^{-1}$.

Proof: See Theorem 2.2 in Newey (1990).

Combining Lemma 2.1 and Theorem 2.1, it is not difficult to see that asymptotically linear $\hat{\beta}$ with influence function $\psi_\beta^*(z)$ is an efficient estimator.
2.3.2 Semiparametric efficiency bound for independence

We now apply the approach implicit in Lemma 2.1 and Theorem 2.1 to find the efficient score and lower bound for each of the stochastic assumptions. First, we treat the assumption of stochastic independence of $\varepsilon$ and $x$. Formally, we assume the following regarding (2.1):

Assumption A1: $(x, \varepsilon)$ are jointly i.i.d.;

Assumption A2: $\varepsilon$ independent of $x$; and

Assumption A3: $E[\varepsilon] = 0$ and $E[x_i] = 0$.

Note that (A3) is not restrictive, since we always can include a nonzero mean in the nonparametric function $g(x_2)$.

Suppose $\varepsilon$ has density $f_\varepsilon(\varepsilon; \eta_2)$ and $x$ has density $f_x(x; \eta_3)$, where $\eta_2$ and $\eta_3$ are nuisance parameters. Then the joint likelihood function for this semiparametric model can be expressed as,

$$ L(y, x; \beta) = \sum_{i=1}^{n} \left[ \log f_\varepsilon(y - g(x_2; \eta_1) - x_i' \beta; \eta_2) + \log f_x(x; \eta_3) \right] $$

(2.4)

where $\eta_1$ is also a nuisance parameter. We can use the above likelihood function and the given semiparametric assumptions to construct the tangent set and then calculate the efficient score and the lower bound. The idea is formalized in the next theorem.

Theorem 2.2 Under (A1)-(A3), the efficient score function associated with the model (2.1) is

$$ S(\varepsilon) = -S_\varepsilon(\varepsilon|x_i - E(x_i | x_2)) $$

(2.5)

where $\varepsilon = (y - E(y | x_2)) - (x_i - E(x_i | x_2))' \beta$ and $S_\varepsilon(\varepsilon) = \partial \log f_\varepsilon(\varepsilon) / \partial \varepsilon$. 
Proof: included in Appendix A.

It follows from Theorem 2.1 that the semiparametric efficiency bound for the independence case is the inverse of

\[ E[S(z) \cdot S(z')] = E[S_e(z)^2] \cdot E[(x_i - E(x_1 | x_2))(x_i - E(x_1 | x_2))]. \]

There is a related example in Newey (1990a) without the nonlinear and unspecified part. The efficient score derived for that purely linear model with residual \( \varepsilon \) independent of regressors is \( S(z) = -S_e(z)[x - E(x)] \). Assuming knowledge of the conditional expectations, we can simply transform, following Robinson, (2.1) to obtain:

\[ \bar{y} = y - E(y | x_2) = (x_i - E(x_1 | x_2))' \beta + \varepsilon = \bar{x}' \beta + \varepsilon \quad (2.6) \]

and simply apply Newey's result to the transformed model. Note that the Robinson estimator will attain the stated bound under independence only if \( E[S_e(z)^2] = 1/E[\varepsilon^2] \), as will occur with normality.

### 2.3.3 Semiparametric efficiency bound for conditional symmetry

We now turn to the case where the disturbances are conditionally symmetric given the exogenous variables. In this case, assumptions (A2) and (A3) for the independence case are replaced by the following,

**Assumption A2':** \( F(\varepsilon < \mu | X) = 1 - F(-\varepsilon < \mu | X) \), where \( F \) is the CDF of \( \varepsilon \) and

**Assumption A3':** \( E[x_i] = 0. \)

Note that the disturbances have conditional zero mean given the exogenous variables (provided the moment exists), since \( \varepsilon \) is symmetric about zero. Accordingly, any nonzero mean is now included as part of the nonparametric function \( g(x_2) \).
Under the new assumptions, the tangent set/projection approach used in the independent case yields the following theorem.

**Theorem 2.3**  Under (A1), (A2'), and (A3'), the efficient score associated with model (2.1) is

\[
S(z) = -S_{elx}(\varepsilon, x) \left( x_1 - \frac{E[S^2_{elx}(\varepsilon, x) x_1 | x_2]}{E[S^2_{elx}(\varepsilon, x) | x_2]} \right)
\]  

(2.7)

where \( S_{elx}(\varepsilon, x) = \partial \log f_{elx}(\varepsilon, x) / \partial \varepsilon \).

Proof: included in Appendix A.

By Theorem 2.1, it follows that the lower bound for semiparametric estimation under the conditional symmetry assumption is given by the inverse of

\[
E[S(z) \cdot S(z)'] = E[E[S_{elx}(\varepsilon, x)'^2 | x_1] \cdot (x_1 - E(w(z) x_1 | x_2))(x_1 - E(w(z) x_1 | x_2)'\]  

where \( w(z) = S_{elx}(\varepsilon, x)'^2 x_1 / E[S^2_{elx}(\varepsilon, x) | x_2] \).

Note that the function \( S_{elx}(\varepsilon, x) \) is antisymmetric (odd) in \( \varepsilon \) under the conditional symmetry assumption. In the event that the disturbances also turn out to be independent, then the efficient score given by (2.7) simplifies to (2.5), except that \( S_{\varepsilon}(\varepsilon) \) is now restricted to be an odd function. And the Robinson estimator will attain the stated bound only if \( E[w(z)] = E[x_1 | x_2] \) and \( E[S_{elx}(\varepsilon)^2 | x] = 1/E[^2] \), which will typically only happen under independence and normality.
2.3.4 Semiparametric efficiency bound for conditional zero mean

The stochastic assumption for the partially linear model that has received the most attention in the semiparametric econometrics literature is the assumption of conditional zero mean given the exogenous variables. For this case, the assumptions are the same as the conditional symmetry case, except that assumption (A2') is replaced by the following,

*Assumption A2'":* $E(e|X)=0$.

As with the conditional symmetry case, any nonzero mean will be included in the nonparametric function $g(x_2)$. Note that both independence and conditional symmetry imply conditional zero mean if the mean exists and are hence special cases. And both the conditional symmetry and conditional zero mean cases allow for conditional heteroskedasticity.

A similar approach to that used in the previous two cases can be applied to the current case to obtain the following efficient score result.

**Theorem 2.4** Under (A1), (A2'"), and (A3'), the efficient score associated with model (2.1) is

$$S(z) = \left( \frac{e / \sigma^2(x)}{\frac{E[(x_1 / \sigma^2(x)) x_2]}{E[1 / \sigma^2(x)] x_2}} \right) \left( x_1 - \frac{E[(x_1 / \sigma^2(x)) x_2]}{E[1 / \sigma^2(x)] x_2} \right)$$  

(2.8)

where $\sigma^2(x) = E[e^2|x]$.

Proof: included in Appendix A.

The lower bound is then given, as above, by Theorem 2.1 as the inverse of

$$E[S(z) \cdot S(z)'] = E\left[ \left( 1 / \sigma^2(x) \right) \cdot (x_1 - E(w(z)x_1 | x_2))(x_1 - E(w(z)x_1 | x_2)) \right],$$
where now \( w(z) = \left( \frac{x_i}{\sigma^2(x)} \right) \frac{1}{E[1/\sigma^2(x)|x_2]} \). This is the lower bound obtained by Chamberlain (1992) for this case, using a multinomial approximation approach.

The efficient score and lower bound for this model are deceptively similar to those that would be obtained by applying weighted least squares to the transformed model (2.6). However, the equivalence to weighted least squares applies only if the conditional variance \( \sigma^2(x) \) is a function only of the secondary regressors \( x_2 \). And the Robinson estimator will generally attain the lower bound for this case only if the model is conditionally homoskedastic.

2.3.5 Semiparametric efficiency bound for partial conditional zero mean case

The minimal assumption under which least square is appropriate for estimation of the standard linear model, indeed the assumption under which it is semiparametric efficient, is that the regressors only be uncorrelated with the disturbances. In a partial movement toward this case for reasons that will become obvious below, we assume the disturbances are uncorrelated with respect to the primary regressors but are conditional zero mean given the secondary regressors. Specifically, the assumptions are the same as the conditional symmetry case, except assumption (A2') is replaced by the following,

Assumption A2''': (1) \( E(\varepsilon \cdot x_i) = 0 \), and (2) \( E(\varepsilon | x_2) = 0 \).

Again, any nonzero mean will be included in the nonparametric function.

If we utilize a similar approach to the previous cases, we find the following rather remarkable result:
Theorem 2.5 Under (A1), (A2*'), and (A3*), the efficient score associated with model (2.1) is

\[ S(z) = -E[\bar{x}_i \bar{x}_i'] \{ E[\epsilon^2 \bar{x}_i \bar{x}_i'] \}^{-1} \bar{x}_i \epsilon \]  

(2.9)

where \( \bar{x}_i = x_i - E(x_i | x_2) \).

Proof: included in Appendix A.

The lower bound for this problem, by Theorem 2.1 is the inverse of

\[ E[S(z)S(z)'] = E[\bar{x}_i \bar{x}_i'] \{ E[\epsilon^2 \bar{x}_i \bar{x}_i'] \}^{-1} E[\bar{x}_i \bar{x}_i'] \].

But this covariance is just the heteroskedastic consistent covariance matrix of the transformed model (2.6). And the Robinson estimator will generally attain this bound under unspecified conditional heteroskedasticity. Thus the assumptions for this case define the class of models for which the Robinson estimator is generally semiparametric efficient.

2.4 Feasible and Efficient Estimation

We have derived the efficient score and the semiparametric efficiency bound for each of the alternative stochastic assumptions regarding the degree of dependence between the disturbances and regressors. In this section we will utilize the efficient scores developed in the previous section to obtain the feasible estimators that asymptotically obtain the lower bounds. The partially conditional zero mean case will not be treated explicitly, since the Robinson estimator attains the bound for that case. The basic approach is to use nonparametric estimators to estimate the unknown components of the scores, other
than $\beta$, and proceed as if the scores were known. The regularity conditions imposed to obtain the limiting behavior of the estimators are based on Newey (1994).

2.4.1 Basic approach for feasible estimation

The expectation of the efficient score at the true parameters $(\beta_0, \eta_0)$ is equal to zero for any parametric submodel, i.e., $E(S(z; \beta_0, h(\eta_0))) = 0$ where $h(\eta)$ indicates that the scores may depend on the nuisance parameters through unknown function such as conditional expectations and densities. If we knew the value of the nuisance parameters and the functions for a parametric model, the obvious estimator would be obtained as the \( \hat{\beta} \) solving the following sample moment analog,

$$0 = n^{-1} \sum_{i=1}^n S(z_i, \hat{\beta}, h(\eta_0)).$$  \hspace{1cm} (2.10)

Under weak regularity conditions, such an estimator should have the efficient score as an influence function and attain the lower bound for the problem of interest. Since we don’t know $h(\bullet)$ or $\eta_0$, our approach is to estimate the unknown functions nonparametrically and obtain $\hat{\beta}$ as the solution to the feasible moment condition

$$0 = n^{-1} \sum_{i=1}^n S(z_i, \hat{\beta}, \tilde{h})$$  \hspace{1cm} (2.11)

where $\tilde{h}$ is a preliminary estimator of the infinite dimensional function. The form of $S(\bullet)$, $h(\bullet)$, and the estimators $\tilde{h}$, will obviously depend on the stochastic assumption.

The important issue is whether the estimators resulting from this feasible procedure are also consistent and asymptotically efficient. We first introduce and discuss regularity conditions that assure consistency. Let $\bar{S}_n(\beta, h) = n^{-1} \sum_{i=1}^n S(z_i, \beta, h)$ and
\( \overline{S}_0(\beta) = E[S(z, \beta, h_0)] \). Let \( \|h\| \) denote a (e.g. Sobolev) norm for the function \( h \). Given this notation, we assume the following regularity conditions are satisfied:

**Assumption C1:** The parameter space of \( \beta \), denoted \( \mathcal{B} \), is compact;

**Assumption C2:** \( \overline{S}_0(\beta) = 0 \) for \( \beta \in \mathcal{B} \) implies \( \beta = \beta_0 \).

**Assumption C3:** For all \( \beta \in \mathcal{B} \), (i) \( S(z, \beta, h_0) \) continuous; (ii) \( \|S(z, \beta, h_0)\| \leq b(z) \), for \( b(z) > 0 \); and (iii) \( \|S(z, \beta, h) - S(z, \beta, h_0)\| \leq \bar{b}(z)\|h - h_0\|^{\bar{\epsilon}} \), for \( \epsilon > 0 \), \( \bar{b}(z) > 0 \).

Under these conditions, we have the following consistency result, due to Newey (1994b):

**Theorem 2.6:** If Assumptions (C1) - (C3) are satisfied and \( \|\tilde{h} - h_0\| \rightarrow 0 \), then \( \tilde{\beta} \rightarrow \beta_0 \).

**Proof:** included in Appendix A.

Except for (C3.iii), the conditions are relatively primitive. The identification condition (C2) is needed to rule out secondary solutions to (2.11) that will converge to the wrong value. In reality, we only need local identification, since we can then restrict \( \mathcal{B} \) to a close enough neighborhood of \( \beta_0 \) to satisfy (C2). Moreover, since we have a consistent preliminary estimator from Robinson, we are assured of starting in this neighborhood with probability approaching one. For the same reason, the compactness assumption (C1) is not restrictive, provided that \( \beta_0 \) in not a boundary point, since the smaller neighborhood can always be chosen as compact. The condition (C) guarantees
that $\bar{S}_n(\beta, \hat{h})$ converges uniformly in probability to $\bar{S}_0(\beta)$. The continuity condition (C3.i) is introduced to assure continuity of $\bar{S}_n(\beta, h_0)$ and, hence by (C3.ii) and (C3.iii), $\bar{S}_0(\beta)$. The boundedness condition (C3.ii), is a standard assumption for uniform convergence and easily verified. The final condition (C3.iii), guarantees that the remainder from estimation of $h$ is uniformly small.

In order to establish the asymptotic normality and semiparametric efficiency of $\hat{\beta}$ yielded by (2.11), we need to introduce additional regularity conditions. Specifically, we impose the following conditions:

**Assumption N1:** There are functions $D(z, h)$, linear in $h$, and $b(z)$, such that (i) for $\|h - h_0\|$ small enough $\|S(z, h) - S(z, h_0) - D(z, h - h_0)\| \leq b(z)\|h - h_0\|^2$ and (ii) $E[b(z)\sqrt{n}\|\hat{h} - h_0\|^2] \to 0$.

**Assumption N2:** $\sum_{i=1}^{n} \left[ D(z, \hat{h} - h_0) - \int D(z, \hat{h} - h_0) dF_0 \right] / \sqrt{n} \to 0$.

**Assumption N3:** For $\|\hat{h} - h_0\|$ small enough, $\int D(z, \hat{h} - h_0) dF_0 = 0$.

**Assumption N4:** (i) $\beta \in$ interior($B$); (ii) there is $\|h\|, \varepsilon > 0$, and a neighborhood $N$ of $\beta_0$ such that for all $\|h - h_0\| > \varepsilon$, $S(z, \beta, h)$ is differentiable in $\beta$ on $N$; (iii) $M = E[\partial S(z, \beta_0, h_0) / \partial \beta']$ nonsingular; (iv) $E[\|S(z, \beta_0, h_0)\|^2] < 0$; and (v) Assumption N1 is satisfied with $S(z, \beta, h)$ there equal to each row of $\partial S(z, \beta, h) / \partial \beta'$.

Under these conditions, we have the following asymptotic normality result, due to Newey (1994a).
Theorem 2.7: If (C1)-(C3) and (N1)-(N4) are satisfied and \( \| \hat{h} - h_0 \| \to 0 \), then
\[
\sqrt{n}(\hat{\beta} - \beta_0) \to N(0, V_\beta^*),
\]
where \( V_\beta^* = (E[S(z, \beta_0, h_0)S(z, \beta_0, h_0)]^{-1} \).

Proof: included in Appendix A.

According to Newey, Assumptions (N1) and (N2) are “second-order”, and hence can be viewed as regularity conditions that should be satisfied if \( S(z, \beta_0, h) \) is sufficiently smooth and \( \hat{h} \) sufficiently well behaved. The condition (N1) requires that the remainder term from a linearization be small and is often straightforward to verify when the norm is Sobolev and the rate of convergence for \( \hat{h} \) is sufficiently fast. The assumption (N2) is a stochastic equicontinuity condition and follows directly when \( \hat{h} \) is a kernel estimator (see Newey (1994a) for discussion). On the other hand, Assumption (N3) is “first-order” and allows \( n^{-1/2} \sum_{i=1}^{n} S(z_i, \beta_0, \hat{h}) \) to be asymptotically normal even though \( \hat{h} \) may converge at a slow rate. Accordingly, it is particularly important to verify this assumption for the specific alternative estimators that follow. Assumption (N4) is introduced to guarantee uniform convergence of the estimated Jacobian with full rank.

In order to conduct inference using the estimator \( \hat{\beta} \), it is useful to have a consistent estimator of the covariance matrix \( V_\beta^* = (E[S(z, \beta_0, h_0)S(z, \beta_0, h_0)]^{-1} \). Define
\[
\Omega = E[S(z, \beta_0, h_0)S(z, \beta_0, h_0)] \quad \text{and} \quad \hat{\Omega} = n^{-1} \sum_{i=1}^{n} S(z_i, \hat{\beta}, \hat{h})S(z_i, \hat{\beta}, \hat{h}),
\]
then \( \hat{\Omega}^{-1} \) provides a direct estimator of the target covariance. Since
\[
M = E[\partial S(z, \beta_0, h_0)/\partial \beta] = -E[S(z, \beta_0, h_0)S(z, \beta_0, h_0)],
\]
then we can write $V_\beta^* = M^{-1} \Omega M^{-1}$ and $\hat{V}_\beta^* = \hat{M}^{-1} \hat{\Omega} \hat{M}^{-1}$ provides an alternative, more robust estimator. The following result applies to the latter approach and follows from Newey (1994a).

**Theorem 2.8:** If (i) $\hat{\beta} \xrightarrow{p} \beta_0$ and there is $\|h\|$ such that

$\|S(z,b,h) - S(z,\beta_0,h_0)\| \leq b(z)\|h-h_0\|$ and $E[b(z)^3] < \infty$ and (ii) Assumption (N4) is satisfied, then $\hat{V}_\beta^* \xrightarrow{p} V_\beta^*$.

### 2.4.2 Feasible and efficient estimation under independence

We now present the estimator for the model under the stochastic assumption of independence of the disturbances and regressors. Applying (2.11) to the efficient score for this case, as given by (2.5), yields the estimator

$$0 = n^{-1} \sum_{i=1}^{n} \hat{S}_\varepsilon \left( y_i - \hat{E}[y|x_{2i}] - (x_{ii} - \hat{E}[x|x_{2i}]) \hat{\beta} \right) x_{ii} - \hat{E}[x|x_{2i}] \right)$$

(2.12)

where the estimators of the unknown components are

$$\hat{S}_\varepsilon(\varepsilon) = -a^{-1} \sum_{j=1}^{n} K((\varepsilon_j - \varepsilon)/a) / \sum_{j=1}^{n} K((\varepsilon_j - \varepsilon)/a),$$

$$\tilde{\varepsilon}_j = (y_j - \hat{E}[y|x_{2j}]) - (y_j - \hat{E}[y|x_{2j}]) \bar{\beta},$$

$$\hat{E}[y|x_{2j}] = \sum_{t=1}^{n} \nu_t K((x_{2t} - x_{2j})/a) / \sum_{t=1}^{n} K((x_{2t} - x_{2j})/a),$$

and $\bar{\beta}$ is a preliminary consistent estimator, such as the Robinson estimator. The function $K(\bullet)$ is the Nadaraya-Watson kernel with window width $a$ that may differ for the density and conditional expectation estimators.

One attractive feature of this estimator is that it is a semiparametric maximum likelihood estimator. That is, it results from maximization of the likelihood when the
density and conditional expectations are estimated but are treated as known in the maximization. Specifically, we have

$$\hat{\beta} = \arg\max_{\beta} \left\{ n^{-1} \sum_{i=1}^{n} \hat{r}_i \left( y_i - \hat{E}[y|x_{2i}] \right) - \left( x_{ii} - \hat{E}[x_1|x_{2i}] \right) \beta \right\}.$$ 

This objective function provides an attractive way of choosing between multiple solutions to the first-order condition (2.12). And our asymptotic result states that the feasible estimator has the same asymptotic distribution as when we know the density and conditional expectations, so the estimator is asymptotically adaptive.

The most important condition from the previous section to verify is (N3), which requires that the Fréchet derivative $D(z,h-h_0)$ at $h_0$ has expectation zero. It is straightforward to show the condition is met for the ordinary derivative for any parametric submodel, which is sufficient.

A second important condition to verify is (N1), which requires $\hat{h}$ converge to $h_0$ at a rate faster than $n^{1/4}$, unless the linearization is exact. For the present problem, this requires that $q$, the number of conditioning variables $x_2$ in the conditional expectation be no more than 3, for the usual second order kernel with optimally chosen window width. This is the same result as found by Robinson for his estimator. The density estimator presents no problem since it is a univariate estimator.

2.4.3 Feasible and efficient estimation under conditional symmetry

We next present a feasible estimator of the model under the stochastic assumption of conditional symmetry of the disturbances given the regressors. The efficient score for this case is given by (2.7), which may be utilized with (2.11) to obtain an $\hat{\beta}$ as the solution to the following
\[ 0 = n^{-1} \sum_{i=1}^{n} \hat{S}_{el|x}( y_i - \hat{E}[y|x_{2i}], x_i ) - (x_i - \hat{E}[x_i|x_{2i}]) \beta_i x_i \left( x_i - \frac{E[\hat{S}_{el|x}(\bar{\varepsilon}, x_1) x_{2i}]}{E[\hat{S}_{el|x}(\bar{\varepsilon}, x) x_{2i}]} \right) \]  

(2.13)

where \( \hat{S}_{el|x}(\bar{\varepsilon}, x) = -a^{-1} \sum_{j=1}^{n} K_i((u_j - u)/a)/\sum_{j=1}^{n} K((x_j - x)/a) \)

for \( u_j = (\bar{\varepsilon}_j, x_j') \). The conditional expectation estimator is the same as for the independence case.

As with the independence case, the zero expectation of the Fréchet derivative condition (N3) is met since it holds for ordinary derivatives of all parametric submodels, which is straightforward to show. The "curse of dimensionality" is more severe for conditional symmetry than independence. This is because the conditional density estimation involves all \( k + q + 1 \) variables and not just the secondary regressors. In order to have \( \hat{h} \) to converge to \( h_0 \) at a rate faster than \( n^{1/4} \) will likely require the use of higher order kernel techniques.

Note that \( S_{el|x}(\varepsilon, x) \) is an odd function in \( \varepsilon \), whereas \( \hat{S}_{el|x}(\varepsilon, x) \) is not so restricted. This suggests that there might possibly be gains from imposing this restriction on the estimator. This is easily accomplished by using antithetic variate techniques, to obtain:

\[ \hat{S}_{el}^a(\varepsilon, x) = .5\hat{S}_{el|x}(\varepsilon, x) - .5\hat{S}_{el|x}(-\varepsilon, x). \]

Asymptotically, there would be no gain from using the restricted estimator, since all that is required is consistency of the estimator, but it might be useful in finite samples.

The fact that symmetry does not need to be imposed to attain the lower bound allows for the possibility of a general estimator that would attain the bound under both independence and symmetry. In fact, the estimator defined by (2.13) will attain the bound under both assumptions. This is because \( \hat{S}_{el|x}(\varepsilon, x) \) will converge to \( S_{el|x}(\varepsilon, x) \),
which becomes $S_\varepsilon(\varepsilon)$ under independence. Likewise, the conditional expectation estimators $E[(\hat{S}_{1x}(\varepsilon, x)x_{ij})x_{ij}]$ and $E[(\hat{S}_{2x}(\varepsilon, x))x_{ij}]$ converge to $E[S_\varepsilon^2(\varepsilon)|E|x_{ij}|x_{ij}]$ and $E[S_\varepsilon^2(\varepsilon)]$ under the independence assumption, whereupon the ratio of conditional expectations reduces to conditional expectation used in the independence case. Thus it is conceivable to estimate both cases with one estimator even though the two cases are not nested. The downside of this approach is the higher dimensionality of the conditional symmetry case relative to the independence case.

2.4.4 Feasible and efficient estimation under conditional zero mean

Finally, we present a feasible estimator of the model under the stochastic assumption of conditional zero mean. The efficient score for this case is given by (2.8), which may be substituted into (2.11) to obtain the estimator $\hat{\beta}$ as the solution to the following

$$
0 = n^{-1} \sum_{i=1}^{n} \left( y_i - \hat{E}[y|x_{ij}] - (x_{ij} - \hat{E}[x_{ij}|x_{ij}]) \hat{\beta} \right) / \hat{\sigma}^2(x_i) \left( x_{ij} - \frac{E[x_{ij}/\hat{\sigma}^2(x)]x_{ij}}{E[y/\hat{\sigma}^2(x)]x_{ij}} \right)
$$

(2.14)

where

$$
\hat{\sigma}^2(x) = \sum_{j=1}^{n} \tilde{e}_j^2 K((x_j - x)/a) / \sum_{j=1}^{n} K((x_j - x)/a).
$$

The conditional expectation estimator continues the same as for the independence case.

Again, it is straightforward to show that the zero expectation of the Frèchet derivative condition (N3) is met since it holds for ordinary derivatives of all parametric submodels. The “curse of dimensionality” problem is slightly better for this case than the conditional symmetry case but substantially less favorable than for the independence case. Specifically, the conditional variance estimator $\hat{\sigma}^2(x)$ involves all $k + q$ regressors
variables and not just the secondary regressors. In order to have \( \hat{h} \) converge to \( h_0 \) at a rate faster than \( n^{1/4} \) may require the use of higher order kernel techniques.

2.5 Specification Test

One question, often occurred in the empirical studies, is which parametric submodel should be used in estimation, which we can not always tell directly from economic theory. Even if we can construct a combined estimator that covers two parametric restrictions, like independence and conditional symmetry, dimensionality may also be a concern. That is in the conditional symmetry case, it requires higher rate of convergence for nonparametric nuisance parameter estimators and the independence case is relatively "dimension-friend". The problem asks for a specification test among the four possible stochastic restrictions.

We observe that there is a nesting structure among the parametric submodels associated with different stochastic restrictions studied in this chapter, which is described as follows,

Nesting Structure:

*Partially conditional zero mean model*

*Conditional zero mean model*

*Conditional symmetry model*

*Independent model*
Two types of testing statistic are constructed, which can be used to test any of the above nested hypotheses. They are Hausman-Wu type test and LM type test that are introduced on the next page.

1) Hausman-Wu type test:

\( \bar{\beta} \) is unrestricted estimator (e.g. conditional zero mean case)

\( \tilde{\beta} \) is restricted estimator (e.g. conditional symmetry case)

\( \tilde{V}_\beta^* \) and \( \tilde{V}_\beta^* \) are corresponding estimated covariance matrices

\[
H = n(\bar{\beta} - \tilde{\beta}) (\tilde{V}_\beta^* - \tilde{V}_\beta^*) (\bar{\beta} - \tilde{\beta}) \overset{d}{\longrightarrow} \chi^2_c \quad (2.15)
\]

where \( c = \text{rank}(\tilde{V}_\beta^* - \tilde{V}_\beta^*) \).

2) LM type test

Define \( \bar{S}_n(\beta) = n^{-1} \sum_{i=1}^n \hat{S}_n(z_i, \beta, h) \), which is the average score for unrestricted model at \( \beta \).

\[
LM = n \cdot \bar{S}_n(\tilde{\beta}) (\tilde{V}_\beta^{*-1} - \tilde{V}_\beta^{*-1} \tilde{V}_\beta^* \tilde{V}_\beta^{*-1}) \bar{S}_n(\tilde{\beta}) \overset{d}{\longrightarrow} \chi^2_c \quad (2.16)
\]

If specification test is needed in empirical studies, either of the above two statistics can be used and they are equivalent asymptotically. A possible problem that may arise in doing the specification test is pretest bias. A detailed study on pretest bias is beyond the scope of this research. Interested readers can refer to Becker et al. (1990), Judith (1991), and Wong (1997).
2.6 Sampling Results

In the previous section, the asymptotical behavior of our estimator is studied. To gain some idea of the finite sample performance of the estimator and the impact of factors like the dimensionality of $X_2$ and the order of the kernel, a Monte Carlo simulation is conducted using GAUSS. To be able to compare our results with Robinson’s, we try to be as close as possible to his simulations in terms of the underlying model and the distribution assumptions. Three types of partially linear model are chosen for our simulations.

Model (I): \[ Y = \alpha + \beta x_1 + \gamma x_2 + \varepsilon \quad (2.17) \]

Model (II): \[ Y = \alpha + \beta x_1 + \gamma x_2^2 + \varepsilon \quad (2.18) \]

Model (III): \[ Y = \alpha + \beta' x_1 + \sum_{j=1}^{q} \gamma_{j} x_{2, j \rho}^2 + \varepsilon \quad (2.19) \]

Four sample size, denoted as $n$, are picked: 50, 100, 200, 300 and the number of replications, $r$, are from 1000 to 5000. A trimming parameter, $m$, is also employed in simulation to avoid the tail observations in density estimation. There is no serious attempt at an exact optimal choice (cross validation) for the smoothing parameter $h$, instead $h$ is picked according to the optimal asymptotical order from balancing bias and variance in a general fashion that reflects the assumptions (N1)-(ii) in Theorem 2.7.

The results under independence assumption are reported for all three models using three different kernels, under two types of disturbance assumptions: standardized chi squared with mean subtracted (denoted as $\chi_0^2$) and standard normal. Besides, a small simulation is also conducted under conditional zero mean assumption for model (I). In
(2.17) and (II), we take $x_1$ and $x_2$ to be scalar random variables from a bivariate normal population with zero means, variances 4 and 3, and covariance 2 (exactly same as the Robinson's simulation). In simulations for (III), two values of dimensional parameter $q$ are tried. The true values of $\alpha, \beta,$ and $\gamma$ are set to be one. In the following each table, we report the simulation biases of the estimated parameter of interest for both $\bar{\beta}$ (the Robinson estimator) and $\hat{\beta}$ (the semiparametric efficient estimator), and their corresponding efficiency indices, defined by Robinson as the ratio of OLS estimators' simulation standard deviation to $\hat{\beta}(i)$'s, where $i$ stands for different normal kernel $K(i)$: $i = 1$ - standard normal kernel, $i = 2$ - fourth order kernel, and $i = 3$ - sixth order kernel.
Table 2.1: Simulation Results under Independence: Model (I)
\( \varepsilon \sim \chi^2_0 \) and \( m = 0.0005 \)

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>( n = 50, r = 1000 )</th>
<th>( n = 100, r = 1000 )</th>
<th>( n = 200, r = 1000 )</th>
<th>( n = 300, r = 2000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\beta} (1) )</td>
<td>0.020599</td>
<td>0.014269</td>
<td>0.011160</td>
<td>0.005094</td>
</tr>
<tr>
<td>( \hat{\beta} (2) )</td>
<td>0.002918</td>
<td>0.003337</td>
<td>0.003968</td>
<td>-0.00053</td>
</tr>
<tr>
<td>( \hat{\beta} (3) )</td>
<td>0.002471</td>
<td>0.003157</td>
<td>0.003826</td>
<td>-0.00070</td>
</tr>
<tr>
<td>( \hat{\beta} (1) )</td>
<td>0.019815</td>
<td>0.014310</td>
<td>0.010086</td>
<td>0.005600</td>
</tr>
<tr>
<td>( \hat{\beta} (2) )</td>
<td>0.001636</td>
<td>0.002480</td>
<td>0.003039</td>
<td>-0.00012</td>
</tr>
<tr>
<td>( \hat{\beta} (3) )</td>
<td>0.002340</td>
<td>0.002062</td>
<td>0.002854</td>
<td>-0.00044</td>
</tr>
</tbody>
</table>

Table 2.2: Simulation Results under Independence: Model (I)
\( \varepsilon \sim \mathcal{N}(0, I) \) and \( m = 0.0005 \)

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>( n = 50, r = 1000 )</th>
<th>( n = 100, r = 1000 )</th>
<th>( n = 200, r = 1000 )</th>
<th>( n = 300, r = 2000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\beta} (1) )</td>
<td>0.025088</td>
<td>0.016275</td>
<td>0.009434</td>
<td>0.006497</td>
</tr>
<tr>
<td>( \hat{\beta} (2) )</td>
<td>0.007575</td>
<td>0.004455</td>
<td>0.002040</td>
<td>0.000848</td>
</tr>
<tr>
<td>( \hat{\beta} (3) )</td>
<td>0.007882</td>
<td>0.003755</td>
<td>0.001649</td>
<td>0.000837</td>
</tr>
<tr>
<td>( \hat{\beta} (1) )</td>
<td>0.024993</td>
<td>0.015916</td>
<td>0.009553</td>
<td>0.006577</td>
</tr>
<tr>
<td>( \hat{\beta} (2) )</td>
<td>0.009408</td>
<td>0.004029</td>
<td>0.001856</td>
<td>0.001102</td>
</tr>
<tr>
<td>( \hat{\beta} (3) )</td>
<td>0.008561</td>
<td>0.003367</td>
<td>0.001161</td>
<td>0.000857</td>
</tr>
</tbody>
</table>
Table 2.3: Simulation Results under Independence: Model (II)
\[ \varepsilon \sim \chi^2_0 \text{ and } m = 0.0005 \]

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>$n = 50, r = 1000$</th>
<th>$n = 100, r = 1000$</th>
<th>$n = 200, r = 1000$</th>
<th>$n = 300, r = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h = 0.79207558$</td>
<td>$h = 0.68954185$</td>
<td>$h = 0.60028104$</td>
<td>$h = 0.55352390$</td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>$\bar{\beta}$ (1)</td>
<td>Bias</td>
<td>$\bar{\beta}$ (2)</td>
<td>Bias $\bar{\beta}$ (3)</td>
</tr>
<tr>
<td>$\bar{\beta}$ (1)</td>
<td>-0.00626</td>
<td>0.721105</td>
<td>-0.00397</td>
<td>0.716003</td>
</tr>
<tr>
<td></td>
<td>0.004768</td>
<td>0.544855</td>
<td>-0.00357</td>
<td>0.707571</td>
</tr>
<tr>
<td></td>
<td>0.003930</td>
<td>0.544855</td>
<td>-0.00324</td>
<td>0.607098</td>
</tr>
<tr>
<td></td>
<td>0.00343</td>
<td>0.650132</td>
<td>-0.00175</td>
<td>0.781283</td>
</tr>
<tr>
<td></td>
<td>0.000541</td>
<td>0.771305</td>
<td>-0.00157</td>
<td>0.847749</td>
</tr>
<tr>
<td></td>
<td>0.004253</td>
<td>0.584347</td>
<td>-0.00174</td>
<td>0.762529</td>
</tr>
</tbody>
</table>

Table 2.4 Simulation Results under Independence: Model (II)
\[ \varepsilon \sim \mathcal{N}(0, \eta) \text{ and } m = 0.0005 \]

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>$n = 50, r = 1000$</th>
<th>$n = 100, r = 1000$</th>
<th>$n = 200, r = 1000$</th>
<th>$n = 300, r = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h = 0.79207558$</td>
<td>$h = 0.68954185$</td>
<td>$h = 0.60028104$</td>
<td>$h = 0.55352390$</td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>$\bar{\beta}$ (1)</td>
<td>Bias</td>
<td>$\bar{\beta}$ (2)</td>
<td>Bias $\bar{\beta}$ (3)</td>
</tr>
<tr>
<td>$\bar{\beta}$ (1)</td>
<td>0.002477</td>
<td>0.718499</td>
<td>-0.00194</td>
<td>0.725048</td>
</tr>
<tr>
<td></td>
<td>0.002403</td>
<td>0.701719</td>
<td>-0.00166</td>
<td>0.695437</td>
</tr>
<tr>
<td></td>
<td>0.006809</td>
<td>0.446495</td>
<td>-0.00192</td>
<td>0.634937</td>
</tr>
<tr>
<td></td>
<td>0.002951</td>
<td>0.618605</td>
<td>-0.00028</td>
<td>0.670143</td>
</tr>
<tr>
<td></td>
<td>0.003832</td>
<td>0.684241</td>
<td>-0.00079</td>
<td>0.710679</td>
</tr>
<tr>
<td></td>
<td>0.009302</td>
<td>0.445948</td>
<td>0.000190</td>
<td>0.670346</td>
</tr>
</tbody>
</table>

| Bias                  | $\bar{\beta}$ (2)  | Bias                 | $\bar{\beta}$ (3)  | Bias $\bar{\beta}$ (3) |
| $\bar{\beta}$ (2)    | 0.758443            | -0.00115             | 0.782068            |
|                      | 0.749624            | -0.00070             | 0.760836            |
|                      | 0.724890            | -0.00067             | 0.727692            |
|                      | 0.725548            | -0.00044             | 0.749261            |
|                      | 0.725413            | -0.00012             | 0.761982            |
|                      | 0.694334            | 0.00033              | 0.728004            |
For both model (I) and (II) with the $\chi^2$-distributed disturbance terms, the simulation results show that the semiparametric efficient estimator (SEE) in general dominates Robinson’s in terms of both bias and efficiency, especially after the sample size becomes large. In (I), the SEE is even better than the OLS estimator when the unknown function $g(.)$ is correctly specified under chi squared distribution. But, when residual is normally distributed, the above result is not obvious, which is however not inconsistent with the observation that the Robinson’s estimator is efficient under independence when the error term follows a Normal distribution and the OLS estimator is equivalent to MLE under Normal distribution. It is also possible that $\widehat{\beta}$ performs relatively better in small sample than $\hat{\beta}$ given the independence assumption and the normally distributed disturbances. We also observe that the efficiency indices in the normal distribution case are smaller than those in the chi squared case. This is due to the fact that under normality, $\widehat{\beta}_{OLS}$ is the same as MLE and hence efficient, but not under chi squared distributed residuals.

Next, we test the both estimation methods against $X_2$ with higher dimension ($q$). Model (III) is an extension of model (II), in which we make exact the same assumptions about the parameters and the distributions of the regressors. That is $\alpha$, $\beta$, and all $\gamma_j$ are set equal to one; and $X_1$ and $X_2$ are equicorrelated identically distributed $N(1,3)$ variables with correlation 2/3.
Table 2.5: Simulation Results under Independence: Model (III)

\[ \varepsilon \sim \chi^2_0, \ q = 3, \text{ and } m = 0.0005 \]

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>( n = 50, \ r = 1000 )</th>
<th>( h = 0.99049121 )</th>
<th>( n = 100, \ r = 1000 )</th>
<th>( h = 0.89711133 )</th>
<th>( n = 200, \ r = 1000 )</th>
<th>( h = 0.81254396 )</th>
<th>( n = 300, \ r = 2000 )</th>
<th>( h = 0.76680717 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tilde{\beta} (1) )</td>
<td>0.060298</td>
<td>0.463973</td>
<td>0.036033</td>
<td>0.546648</td>
<td>0.025067</td>
<td>0.646859</td>
<td>0.020827</td>
<td>0.670462</td>
</tr>
<tr>
<td>( \tilde{\beta} (2) )</td>
<td>0.036577</td>
<td>0.402905</td>
<td>0.022339</td>
<td>0.512099</td>
<td>0.001450</td>
<td>0.616549</td>
<td>0.010564</td>
<td>0.629704</td>
</tr>
<tr>
<td>( \tilde{\beta} (3) )</td>
<td>0.045454</td>
<td>0.343369</td>
<td>0.028282</td>
<td>0.444142</td>
<td>0.018411</td>
<td>0.527178</td>
<td>0.013676</td>
<td>0.541526</td>
</tr>
<tr>
<td>( \hat{\beta} (1) )</td>
<td>0.052143</td>
<td>0.426764</td>
<td>0.024585</td>
<td>0.573346</td>
<td>0.017171</td>
<td>0.719592</td>
<td>0.012655</td>
<td>0.758059</td>
</tr>
<tr>
<td>( \hat{\beta} (2) )</td>
<td>0.033962</td>
<td>0.385851</td>
<td>0.019988</td>
<td>0.504607</td>
<td>0.012776</td>
<td>0.613078</td>
<td>0.009448</td>
<td>0.644822</td>
</tr>
<tr>
<td>( \hat{\beta} (3) )</td>
<td>0.043643</td>
<td>0.326771</td>
<td>0.024551</td>
<td>0.429993</td>
<td>0.015493</td>
<td>0.529658</td>
<td>0.001172</td>
<td>0.559758</td>
</tr>
</tbody>
</table>

Table 2.6: Simulation Results under Independence: Model (III)

\[ \varepsilon \sim \mathcal{N}(0, \tau), \ q = 3, \text{ and } m = 0.0005 \]

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>( n = 50, \ r = 1000 )</th>
<th>( h = 0.99049121 )</th>
<th>( n = 100, \ r = 1000 )</th>
<th>( h = 0.89711133 )</th>
<th>( n = 200, \ r = 1000 )</th>
<th>( h = 0.81254396 )</th>
<th>( n = 300, \ r = 2000 )</th>
<th>( h = 0.76680717 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tilde{\beta} (1) )</td>
<td>0.049217</td>
<td>0.452004</td>
<td>0.034329</td>
<td>0.528032</td>
<td>0.022421</td>
<td>0.626612</td>
<td>0.018602</td>
<td>0.673661</td>
</tr>
<tr>
<td>( \tilde{\beta} (2) )</td>
<td>0.027315</td>
<td>0.404305</td>
<td>0.018899</td>
<td>0.494793</td>
<td>0.010593</td>
<td>0.581486</td>
<td>0.007724</td>
<td>0.641026</td>
</tr>
<tr>
<td>( \tilde{\beta} (3) )</td>
<td>0.036690</td>
<td>0.344582</td>
<td>0.025045</td>
<td>0.427848</td>
<td>0.013705</td>
<td>0.507200</td>
<td>0.010922</td>
<td>0.555942</td>
</tr>
<tr>
<td>( \hat{\beta} (1) )</td>
<td>0.035835</td>
<td>0.404738</td>
<td>0.024494</td>
<td>0.509140</td>
<td>0.015334</td>
<td>0.643383</td>
<td>0.013818</td>
<td>0.686857</td>
</tr>
<tr>
<td>( \hat{\beta} (2) )</td>
<td>0.026513</td>
<td>0.380068</td>
<td>0.018975</td>
<td>0.476293</td>
<td>0.007691</td>
<td>0.561785</td>
<td>0.006595</td>
<td>0.623231</td>
</tr>
<tr>
<td>( \hat{\beta} (3) )</td>
<td>0.033422</td>
<td>0.330409</td>
<td>0.023529</td>
<td>0.415487</td>
<td>0.010814</td>
<td>0.498283</td>
<td>0.009661</td>
<td>0.552958</td>
</tr>
</tbody>
</table>
Table 2.7: Simulation Results under Independence: Model (III)

\( \varepsilon \sim \chi^2_0, \ q = 5, \text{ and } m = 0.00000005 \)

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>( n = 50, r = 1000 )</th>
<th>( h = 1.6549886 )</th>
<th>( n = 100, r = 1000 )</th>
<th>( h = 1.5690575 )</th>
<th>( n = 200, r = 1000 )</th>
<th>( h = 1.4875881 )</th>
<th>( n = 300, r = 2000 )</th>
<th>( h = 1.4419069 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (1) )</td>
<td>0.228967</td>
<td>0.197775</td>
<td>0.170455</td>
<td>0.158780</td>
<td>0.23390</td>
<td>0.220640</td>
<td>0.256982</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.23390</td>
<td>0.220640</td>
<td>0.256982</td>
<td>0.23390</td>
<td>0.220640</td>
<td>0.256982</td>
<td>0.23390</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (2) )</td>
<td>0.098902</td>
<td>0.072648</td>
<td>0.052520</td>
<td>0.042928</td>
<td>0.245274</td>
<td>0.280803</td>
<td>0.372409</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.245274</td>
<td>0.280803</td>
<td>0.372409</td>
<td>0.245274</td>
<td>0.280803</td>
<td>0.372409</td>
<td>0.245274</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (3) )</td>
<td>0.103430</td>
<td>0.079202</td>
<td>0.059365</td>
<td>0.045890</td>
<td>0.194540</td>
<td>0.229240</td>
<td>0.306183</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.194540</td>
<td>0.229240</td>
<td>0.306183</td>
<td>0.194540</td>
<td>0.229240</td>
<td>0.306183</td>
<td>0.194540</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (1) )</td>
<td>0.193047</td>
<td>0.149982</td>
<td>0.111721</td>
<td>0.097801</td>
<td>0.193150</td>
<td>0.217211</td>
<td>0.285130</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.176759</td>
<td>0.217211</td>
<td>0.285130</td>
<td>0.176759</td>
<td>0.217211</td>
<td>0.285130</td>
<td>0.176759</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (2) )</td>
<td>0.089957</td>
<td>0.070361</td>
<td>0.050013</td>
<td>0.037377</td>
<td>0.231747</td>
<td>0.278176</td>
<td>0.372390</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.231747</td>
<td>0.278176</td>
<td>0.372390</td>
<td>0.231747</td>
<td>0.278176</td>
<td>0.372390</td>
<td>0.231747</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (3) )</td>
<td>0.101390</td>
<td>0.080893</td>
<td>0.054140</td>
<td>0.041944</td>
<td>0.187551</td>
<td>0.222193</td>
<td>0.300008</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.187551</td>
<td>0.222193</td>
<td>0.300008</td>
<td>0.187551</td>
<td>0.222193</td>
<td>0.300008</td>
<td>0.187551</td>
</tr>
</tbody>
</table>

Table 2.8: Simulation Results under Independence: Model (III)

\( \varepsilon \sim M_0, \ h, \ q = 5, \text{ and } m = 0.00000005 \)

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>( n = 50, r = 1000 )</th>
<th>( h = 1.6549886 )</th>
<th>( n = 100, r = 1000 )</th>
<th>( h = 1.5690575 )</th>
<th>( n = 200, r = 1000 )</th>
<th>( h = 1.4875881 )</th>
<th>( n = 300, r = 2000 )</th>
<th>( h = 1.4419069 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (1) )</td>
<td>0.233379</td>
<td>0.191886</td>
<td>0.173623</td>
<td>0.154121</td>
<td>0.211313</td>
<td>0.252643</td>
<td>0.251556</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.211313</td>
<td>0.252643</td>
<td>0.251556</td>
<td>0.211313</td>
<td>0.252643</td>
<td>0.251556</td>
<td>0.211313</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (2) )</td>
<td>0.119892</td>
<td>0.064972</td>
<td>0.047444</td>
<td>0.042834</td>
<td>0.226801</td>
<td>0.333427</td>
<td>0.360676</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.226801</td>
<td>0.333427</td>
<td>0.360676</td>
<td>0.226801</td>
<td>0.333427</td>
<td>0.360676</td>
<td>0.226801</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (3) )</td>
<td>0.139320</td>
<td>0.073359</td>
<td>0.054932</td>
<td>0.048947</td>
<td>0.179803</td>
<td>0.261194</td>
<td>0.292603</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.179803</td>
<td>0.261194</td>
<td>0.292603</td>
<td>0.179803</td>
<td>0.261194</td>
<td>0.292603</td>
<td>0.179803</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (1) )</td>
<td>0.206786</td>
<td>0.153545</td>
<td>0.117589</td>
<td>0.097801</td>
<td>0.183884</td>
<td>0.250505</td>
<td>0.290001</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.183884</td>
<td>0.250505</td>
<td>0.290001</td>
<td>0.183884</td>
<td>0.250505</td>
<td>0.290001</td>
<td>0.183884</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (2) )</td>
<td>0.109992</td>
<td>0.060300</td>
<td>0.044868</td>
<td>0.038474</td>
<td>0.219498</td>
<td>0.327983</td>
<td>0.361277</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.219498</td>
<td>0.327983</td>
<td>0.361277</td>
<td>0.219498</td>
<td>0.327983</td>
<td>0.361277</td>
<td>0.219498</td>
</tr>
<tr>
<td>Bias</td>
<td>( \hat{\beta} (3) )</td>
<td>0.137356</td>
<td>0.070750</td>
<td>0.052978</td>
<td>0.045276</td>
<td>0.173967</td>
<td>0.253948</td>
<td>0.286279</td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td>0.173967</td>
<td>0.253948</td>
<td>0.286279</td>
<td>0.173967</td>
<td>0.253948</td>
<td>0.286279</td>
<td>0.173967</td>
</tr>
</tbody>
</table>
As $q$ increases to 3 and 5, the rate of convergences of both the Robinson and the efficient estimators decrease, which is due to the so-called "Curse of Dimensionality" problem. We find that the efficiency of $\hat{\beta}$ improves against $\tilde{\beta}$ as sample size becomes large, especially when disturbance is non-normally distributed. We also observe that the bias is reduced when the kernel is switched from to $K(1)$ to $K(2)$ or $K(3)$, which is consistent with the asymptotic theory of high order kernel. We observe in the case that $q = 3$, the bias reduction is not obvious when kernel is changed from $K(2)$ to $K(3)$ and high order kernel estimators have not improved efficiency. This is probably related to the fact that the window-width here is not optimally chosen for $K(3)$. An empirical evidence for the argument is when optimal window width for fourth order kernel, $K(2)$, is chosen for simulations when $q = 5$ (simply for satisfying the regularity condition for nuisance parameter estimators), we find that the estimators associated with $K(2)$ have smaller sample variance compared with those using $K(1)$.

Finally, we conduct simulations under the assumption of conditional zero mean. To maintain the simulation simple, we only do this for model (I). We follow the exact assumptions made previously, except for that the distribution of $\varepsilon$ is now assumed to have two independent parts, i.e., $\varepsilon = \mu \xi(X)$. $\mu$ follows a standard normal distribution and is independent of $x$, and $\xi(X) = X\lambda$, where $\lambda = [0.5, 0.5]'$. The results are reported in the following tables. This time, when we calculate the relative efficiency, we use the sample variance of weighted least square (WLS) estimator instead of the variance of OLS estimator, simply for WLS is efficient given the information about the nuisance parameters.
Table 2.9 Simulation Results under Conditional Zero: Model (I)

\( \varepsilon \sim \mathcal{N}(0, n) \times \lambda \), and \( m = 0.000005 \)

<table>
<thead>
<tr>
<th>Simulation Parameters</th>
<th>( n = 50, r = 1000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( h = 0.79207558 )</td>
</tr>
<tr>
<td>( \hat{\beta} (1) )</td>
<td>0.023493</td>
</tr>
<tr>
<td>Bias</td>
<td>0.015060</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.424373</td>
</tr>
<tr>
<td>( \hat{\beta} (2) )</td>
<td>0.003650</td>
</tr>
<tr>
<td>Bias</td>
<td>0.001874</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.428726</td>
</tr>
<tr>
<td>( \hat{\beta} (3) )</td>
<td>0.005364</td>
</tr>
<tr>
<td>Bias</td>
<td>0.000853</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.397439</td>
</tr>
<tr>
<td>( \hat{\beta} (1) )</td>
<td>-0.01617</td>
</tr>
<tr>
<td>Bias</td>
<td>-0.01652</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.702372</td>
</tr>
<tr>
<td>( \hat{\beta} (2) )</td>
<td>0.005087</td>
</tr>
<tr>
<td>Bias</td>
<td>0.001125</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.455830</td>
</tr>
<tr>
<td>( \hat{\beta} (3) )</td>
<td>0.003715</td>
</tr>
<tr>
<td>Bias</td>
<td>0.001059</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.420139</td>
</tr>
</tbody>
</table>

The above four simulations show that the semiparametric efficient estimator has smaller sample variance than the Robinson's in almost every case under conditional zero mean assumption. All the simulations conducted demonstrate that \( \hat{\beta} \) in general performs better than \( \tilde{\beta} \), especially when sample size becomes larger.

2.7 Conclusion

In this chapter we proposed a semiparametric efficient estimation method for partially linear model under three types of statistical restrictions. We demonstrate that the estimator is consistent, follows an asymptotic normal distribution, and attains the semiparametric efficiency bound in the limit. A combined efficient estimator that covers both the independence and conditional symmetry cases is suggested. Monte Carlo simulation results suggest that finite sample efficiency gain can be substantial compared
with the Robinson estimator, especially with non-normally distributed error term. A general class of parametric models in which the Robinson's estimator is always semiparametrically efficient is also provided. Two types of specification test are constructed for possible applications.

Our results may also be extended to some other econometric models with partially linear format in a similar way as discussed by Robinson (1988), such as seemingly uncorrelated regression and simultaneous equations.
Chapter 3

Optimal Dynamic Production Policy:
The Case of a Large Oil Field in Saudi Arabia

3.1 Introduction

Modeling the international oil market, and more specifically OPEC 's oil production has been one of the major concerns in energy economics since the Arab oil embargo of 1973-74 that resulted in a four-fold increase in the oil price. Saudi Arabia, the largest supplier in OPEC, produces close to 30% of total OPEC output. It also has roughly 30% of the world's proven oil reserves and a maximum sustainable production capacity of a little less than 10 million barrels per day. The influence of Saudi oil policy on the world oil price is widely acknowledged.

Oil pricing and production decisions in Saudi Arabia are determined at the highest level of government. The Saudi Arabia government expenditure is heavily financed by oil revenues and oil-related activities such as petrochemicals and refining, as shown in Table 3.1.

It is also well known that Saudi oil policy is motivated by more than the maximization of profits or market share. For example, Askari (1991, p. 29) wrote: "Saudi Arabia's oil policy has been historically motivated by broad political and economic factors." According to Askari (1991), the political concerns of Saudi Arabia have included Saudi's role in the world, Arab solidarity, and regional politics. Economic factors apart from profits have included a desire for diversification away from oil revenue
in the long term and a desire to meet fluctuations in both short-term domestic and foreign expenditures.

Table 3.1: Oil Revenues and Expenditure of Saudi Arabia Government

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Oil Revenues</td>
<td>103,880</td>
<td>61,195</td>
<td>39,636</td>
<td>60,157</td>
<td>48,400</td>
<td>78,850</td>
<td>118,142</td>
<td>135,300</td>
</tr>
<tr>
<td>Government</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure</td>
<td>216,363</td>
<td>184,004</td>
<td>141,022</td>
<td>185,317</td>
<td>134,850</td>
<td>150,048</td>
<td>210,430</td>
<td>261,570</td>
</tr>
<tr>
<td>% of oil revenues in</td>
<td>48%</td>
<td>33.3%</td>
<td>28.1%</td>
<td>32%</td>
<td>35.9%</td>
<td>52.5%</td>
<td>56.1%</td>
<td>51.7%</td>
</tr>
</tbody>
</table>

Note: the above oil revenues do not include the transfers from petroleum sector.

It is difficult to discern how the above factors affect Saudi's oil production decisions and the world oil market. A detailed study of these questions is beyond the scope of this research. We focus instead on finding the dynamically optimal (profit maximization) oil production rate in a typical Saudi light oil field under different theoretical assumptions. Our analysis allows us to asses the potential for forgone profits in the event that oil production decisions are based on criteria other than the maximization of the expected present value of profits. Our analysis also provides a measure of the extent to which long-run value maximization is being followed. This in turn allows for the calculation of lost government revenue and lost foreign exchange associated with current oil policies in Saudi Arabia. Even if other goals take precedence over value maximization it is still worthwhile for the Saudi Arabia government to have an estimate of what those policies
are costing. Only then can a rational decision be made about whether the benefits from those policies are sufficient to compensate for the costs.

Many studies have addressed the oil production policies of OPEC, and Saudi Arabia in particular, in a dynamic setting. Quandt (1982) provides a detailed qualitative discussion of Saudi oil policy (in the 70s and 80s) and its possible motivations based on two criteria: optimization of the long-term value of reserves and the attainment of political goals. Khadduri (1996) discusses oil policy in the context of current developments in the Middle East. Powell (1990) gives an excellent review of two major strands of economic research on OPEC oil production. Intertemporal optimization, traditionally attributed to Hotelling (1931), constitutes one of the two strands. Recent research on OPEC or Saudi oil production using dynamic value optimization includes paper: Wirl (1990), Suranovic (1993), Benkerouf (1994), and Lohrenz and Bailey (1995).

Our research differs from these studies in that we consider not only the economic factors that affect Saudi oil policy, like demand and costs, but also relevant reservoir engineering variables, such as the amount of water injected, the cumulative production of the field, and the number of oil wells. Engineering software (The WorkBench Black Oil Simulator) is used to simulate the effects of these reservoir conditions on the production cost and the short-run field capacities in an oil field whose production characteristics mirror those of Ghawar - the largest light oil field in Saudi Arabia. Specifically, the engineering production process thus is modeled by recursively solving a system of homogeneous difference equations that describe the fluid dynamics within and among a set of three-dimension grids that partition the oil field and whose joint behavior describes the production dynamics of the overall field. A short-term dynamic production function for the field is then estimated using the simulation results. This dynamic production is
employed in the dynamic optimization model as an inequality constraint qualifying the intertemporal profit maximization problem.

Our approach contrasts with the standard approach taken in the economics of production literature of either specifying the production function based on some generic property of scale or substitution (e.g., the constant elasticity of production function), or based on approximating forms (such as the translog and Leontief). It is related to Griffin (1977, 1978) who constructs approximations to technology utilizing pseudo-data, but differs from that approach in that the production technology is fundamentally dynamic.

Section 2 discusses some general issues regarding developing a stylized version of our dynamic programming model. Section 3 discusses data sources and the methodology used to estimate the revenue function, the cost functions, and the dynamic production function. Section 4 addresses theoretical issues regarding the solutions to the dynamic programming model. Optimal solutions to our model under a number of scenarios are presented in section 5. Section 6 includes a conclusion.

3.2 Dynamic Modeling of Oil Production Decisions

Modeling the optimal oil policy, particularly of Middle East countries like Saudi Arabia, is not a simple matter. It involves not only modeling the interactions among the major players inside and outside OPEC but also requires a thorough understanding of the country's specific economic and political situations. While acknowledge the complexity, we believe that there is much to be gained by focusing on profitability. In particular, even when the expected present value of profits is not the primary goal, it is of interest to know how much the other objectives cost in terms of foregone profits.
In order to build a representative model for Saudi Arabia's oil production decisions, a number of issues need to be addressed. First, the demand equation for Saudi Arabian oil should be based on the structure of the world oil market. Second, Saudi's oil production cost function needs to be correctly specified, including exploration, development, and production operation costs, which accounts for part of the opportunity costs of oil production. Third, current production affects the reservoir conditions and hence future production costs and capacity (Figure A6 – Figure A10, in the Appendix, provides a clear picture of this idea). Each of the three elements could be developed into an independent paper. However, to maintain a simple model, we utilize the results of other researchers and software packages to formalize these ideas. These sources are described in detail in the next section. In addition, we recognize that oil fields in Saudi Arabia are not homogeneous in their reservoir environments, which may further complicate the picture if we try to model the whole Saudi oil production process. Our research is hence focused on a hypothetical light oil field that is similar to Ghawar in field properties (Ghawar accounts about 60% - 70% of Saudi oil reserves).

We model the optimal production policy for the hypothetical light oil field using the Bellman equation

\[ v(N_t, CP_t) = \max_{X_t, dN_t} \{ r(X_t) - c(X_t, dN_t, W_t, N_t) + \beta v(N_{t+1}, CP_{t+1}) \} \]  \hspace{1cm} (3.1)

subject to

\[ N_{t+1} = N_t + dN_t \]
\[ CP_{t+1} = CP_t + 365X_t \]
\[ W_t = w(X_t, N_t, dN_t) \]

\[ ^{1} \text{It is not difficult to see the form of SP (sequence problem) from which (3.1) is deduced. Under certain general and nonrestrictive conditions, the two problems are equivalent For these conditions, see Theorem 4.2 and 4.3 in Stokey, Lucas, and Prescott (1989).} \]
\[ 0 \leq X_t \leq f(N_t, dN_t, W_t, CP_t) \text{ and } dN \geq 0 \]

Here, \( X, dN, W, N, \) and \( CP \) stand correspondingly for daily oil production from the field, the number of new oil wells drilled during the period, water injected to maintain the reservoir liquid pressure, the number of existing producing wells in the beginning of a period, and cumulative production of the field. \( X \) and \( dN \) are policy variables and \( N \) and \( CP \) are state variables. \( \beta \) is the discount factor. \( r \) denotes the revenue function, which can be formulated as \( p(X)X \) where \( p \) stands for the inverse demand equation that relates the oil price to Saudi output. We also assume that increasing output requires additional water injection as specified by the function \( w \). Finally, the extraction rate is also constrained by the short-term field capacity function \( f \).

Powell (1990, p30) points out that the solutions to intertemporal optimization models for OPEC countries will be "\textit{strongly influenced by the future cost of oil substitutes and the discount rate(s) of the decision-makers.}" Since these factors are difficult to estimate and/or forecast, we also examine the robustness of our solutions to variations in the discount rate and other assumptions about the evolution of oil demand over the coming decades, like the timing of new sources of alternative energy supply.

Two scenarios associated with different discount rates (10% and 30%) are studied in the numerical simulations for the optimal policy. Our choices are consistent with the results of Adelman (1993) (Chapter 21, p457), in which he discuss the oil-producing countries’ discount rates. According to his research, 10 percent is a standard discount rate for countries like the United States and others that can diversify their income to a significant degree. On the other hand, for some OPEC countries, like Saudi Arabia, the discount rate may exceed 20 percent or even three times the standard rate. The reason is that the Saudi government relies heavily on oil incomes (See Table 3.1), which implies
that there is substantial risk associated with the income stream provided by exploiting oil resources.

The other two simulations examine the effects on Saudi’s long term oil production policy of future cost reductions in one oil substitute (solar energy). The difference between these two simulations is the timing of a breakthrough in producing solar energy on a massive scale. In the solar energy literature, there are different predictions of the likely cost reduction of solar energy. Among them, Cody and Tiedje (1992) predict a fall in the cost of photovoltaic electricity from 40 cents per kwh to 7 – 12 cents per kwh in 2010. Dracker and De Laquill (1996) forecast that the cost of electricity from solar energy will fall to 3.5 – 10.6 cents per kwh by 2005 – 2010. A research paper done by the Department of Energy (paper number: DE 90000322) argues that with intensified R&D, the cost of photovoltaics will fall below the cost of electricity from fossil fuel around 2026. We adopted more conservative expectations in our simulations. In one scenario, the cost of photovoltaic electricity starts to fall below that of electricity from fossil fuel in 2036, and in the other it starts a little earlier in 2026. A ten-year transition period is assumed in both cases during which much of the demand for oil gradually switches to solar energy. The detailed description of the transition from oil to solar energy is in section 5. Another issue that could be investigated in more detail is that different components of the energy market could abandon fossil fuel at different times. Fuel cell technology could also displace oil from transportation applications before solar energy does (either through electric vehicles, the use of methanol as a fuel or some other technology). Much more work could be done developing alternative substitution scenarios. In this respect, our work should be regarded as no more than indicative of the possibilities.
3.3 Data and Estimations

This section deals with the first three issues mentioned in the last section: the revenue function, the cost function, and the dynamic production function. One of the contributions of this chapter is the detailed modeling of the three functions. Accordingly, we have avoided excessive detail in other parts of overall problem, but these could be included to provide a more accurate picture.

3.3.1 The Revenue Function

The revenue function specifies the revenue as a function of the rate of extraction from the oil field. Revenue is equal to price times the level of extraction. We base our revenue function on the Oil Market Simulation Model (OMS) developed by the Energy Information Administration. The OMS is an annual model that projects the world oil market through the year 2010 from database that begins in 1979. Its geographic coverage includes almost all major market economies. Net imports from the formerly centrally planned economies are taken as exogenous. The model estimates the effects of price changes on oil supply and demand and computes an oil price path over time that allows supply and demand to remain in balance within the market economies as a whole. OMS has been extensively used in the EIA analyses since 1978 and has been used for every issue of the Agency’s annual report to Congress. The key assumptions in the OMS model are that the current oil price, GDP growth rates, exchange rates, and last year’s supplies and demands are the only determinants of supply and demand from non-OPEC and non-Communist countries. All oil exporting countries except those in OPEC are assumed to be price takers. OPEC countries are assumed to follow a price-reaction function that increases price with increased capacity utilization, where capacity is defined as maximum
sustainable production. The resulting OMS model consists of twelve equations (5 supply, 5 demand, OPEC production, and world price) and two identities. OMS projects a market clearing price for each year that equates world oil demand and the sum of oil productions from all sources, including the inventory changes. Using the OMS model, we simulated twenty-five different production ($X_i$) and price ($P_i$) trajectories over the simulation period (1986-2010) and the corresponding price and production level for the residual supplier but price setter – the OPEC countries. Utilizing these simulation results, the following inverse demand equation is estimated which describes the relation between the demand for Saudi oil production and its corresponding pricing policy.

$$\log p_t = \alpha_0 + \alpha_1 x_t + \alpha_2 T + \varepsilon$$  

(3.2)

where $T$ is the time trend index defined as: $T = t/60$ if $t \leq 60$ and $T = 1$ if $t > 60$, and $t$ starts at 1 in year 1986 and $t$ equals to 60 in year 2045. While we understand the specification most likely excludes some other relevant variables, we want to keep the format simple and still maintain the basic relation between quantity and price. The time index is included to capture the effects of the omitted variables that might affect demand.

Regression results are listed in Table 3.2.

Table 3.2: Estimation of the Inverse Demand Equation

<table>
<thead>
<tr>
<th></th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates</td>
<td>3.5323</td>
<td>-0.0398</td>
<td>3.9656</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.0520</td>
<td>0.0023</td>
<td>0.1557</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.7476</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Therefore the annual revenue function of the field can be expressed as,

$$R(X_t) = (365X_t) \cdot e^{3.5323-0.0398X_t+3.9656T}$$  

(3.3)
3.3.2 The Production Cost Function

Studies of optimal oil production in OPEC countries typically forgo presenting a production cost function. This is no doubt explained by the difficulty of obtaining accurate information. In general, oil production cost consists of three parts: (1) exploration cost, (2) development cost, and (3) production operation cost. We use information from different sources to construct a complete cost function for the hypothetical Saudi light oil field.

Exploration cost includes the costs incurred in geological and geophysical surveys for discovery of new fields, and the drilling of exploration wells, which represents 65-80% of exploration cost. While we have (and use below) data on the cost of drilling producing wells, the cost of an exploration well is still not available for the Saudi light oil field. We do not have an exact form of the exploration expenditure as a function of oil production due to the difficulty of getting operable data for estimation. However, according to Maseron (1990, p.98), exploration expenditure accounts for about 10-20% of the total production cost. Therefore, after estimating the other two parts of the cost, 20% is added to account for the exploration expenditure.

Development costs can be separated into two general categories: (1) surface installation costs (which may also be called infrastructure cost), e.g., trunk line and production facilities, and (2) oil well costs, which include the investment for new wells and the maintenance cost for old ones. Due to data limitations, only a weighted annual average infrastructure cost per barrel (bbl) is calculated for Arabian light fields (like Abu adriyah, Ghawar, etc.), and Arabian medium fields (like Zuluf, Aatif, etc.), which together usually account for about 70% of total Saudi output. The calculation is based on information contained in “Oil Production Capacity in the Gulf, vol II: Saudi
Arabia" (1993) and the weights are the relative production ratio of each type of field in 1992. For adding one daily barrel of capacity, a $1459.3 annual surface installation cost is required, while infrastructure maintenance cost is about $157 per year, or $0.44 per daily barrel in average. In our simulations, only the maintenance part is directly used. The capacity expansion part is represented by a cost of drilling a new producing well, which is, for Arabian light crude field, about two million dollars per well. Clear information on well maintenance cost is hard to obtain. In the Black-Oil simulation, a 20-year depreciation life is assumed for a producing well. In our numerical simulation, we assume that an oil well could not perform normally in producing oil if its present value is too low compared with its original value. Accordingly, to maintain a well's production, certain amount of investment must be paid to cover the depreciation of the well. To keep the calculation simple, we assume that the well maintenance cost is evenly distributed over the life of a well and the depreciation rate is 10%. This gives rise to an approximate 0.1783 million dollar annual maintenance cost per well (under 10% discount rate), including the necessary replacement of the surface equipment. Another way to interpret this calculation is that a new well was drilled using the maintenance costs when the old one expires. In other words, twenty years after a well was drilled and used in oil production, its remaining present value plus the present value of the total maintenance costs is about equal to the present cost of drilling a new oil well.

Production costs can also be separated into two parts: operation costs and reservoir engineering costs. Operation costs denote the expenditures on manpower, and the other variable production costs. According to "Oil production Capacity Expansion Costs for the Persian Gulf", a variable operating expense per barrel ranges from $0.25 to $1.00, depending on the extraction rates. The following estimated functional form (using the
petroconsultants' *Estimator* database and software\(^2\) by EIA.) is used for variable operation cost.

\[
\text{Var. Oper. Cost} = 0.7714Q^{-0.2423}
\]

(3.4)

where \(Q\) denotes the annual field extraction rate.

We use the cost of water injected into the reservoir in order to maintain the reservoir fluid pressure as representative of reservoir engineering costs. Following the suggestion of a petroleum expert we consulted, a $0.2 per barrel cost is used in the cost function to represent water injection.

Summing these components, the estimated oil production cost for a typical oil field in Saudi Arabia is,

\[
\text{Production Cost} = 1.2[(0.44 \times 365)X + (0.7714(365X)^{-0.2423}) \times (365X) + 365 \times (0.2W) + 0.1783N + 2dN]
\]

(3.5)

where \(X, W, N, dN\) follow the same notation as in (3.1), and the estimated water injection function \(W\) is modeled as a function of production rate \(X\) and number of producing wells \(N\).

### 3.3.3 The Dynamic Production Function

It is beyond the scope of our study to simulate an entire field in a country such as Saudi Arabia. This is due to the sheer size of the major Saudi fields, the need for a complex reservoir description specific to the chosen field, the absence of individual well production history, and the time, manpower, and computing resources needed for such an undertaking even if the data were available.

---

\(^2\) The *Estimator* database contains field and production characteristics for eight geological plays (defines as a group of discovered and/or undiscovered fields with similar geological, geographic, and temporal characteristics). Three field sizes (ultimate recovery) provided by *Estimator* for each play are the basis for the low-, mid-, and high-cases.
It is possible, however, to create a Black Oil Simulator utilizing properties that are typical of a large Saudi oil reservoir. The majority of the Arab Light oil production is derived from the Jurassic Arab-D formation. This is a calcarenitic limestone more akin to a highly porous and permeable sandstone than a fractured limestone. This formation is the predominant producing reservoir in the giant Ghawar field.

The simulator model is roughly 5 kilometers from the crest to the oil/water contact. The dip of the formation varies from 0 degrees at the crest to about 5 degrees on the flank. The wells are drilled in a square pattern with 1 kilometer between nearest neighbors. The reservoir gross thickness is about 250 feet. The simulation model, then, is a 3-dimensional wedge perpendicular to a line along the crest with its sides passing through the adjacent lines of wells (0.707 kilometers apart). The model extends downdip to below the oil/water contact. Water injection wells are located below the oil/water contact for peripheral water injection. The rock properties, the fluid properties, and the fluid/rock interaction properties were set to values that are typical for reservoirs in the neighborhood of the Ghawar field. Figure A1 (in Appendix) shows a cross section perpendicular to a line along the crest and through a line of wells. The vertical scale has been exaggerated relative to the horizontal scale in order to display details in the layering. This figure shows the initial positions for the fluids in the well, green corresponds to oil and blue corresponds to water.

The laws of physics control how fluids are distributed in oil reservoirs that are undisturbed for geologic time and how those fluids move in a reservoir once they are disturbed by the initiation of production and/or injection. Those laws of physics can be expressed as partial differential equations. The Black Oil Simulator starts with the fluids in place at the time of discovery, consistent with an equilibrium attained over geologic
time. This becomes the set of initial conditions for the partial differential equations. The mathematical equivalent of wells then are added to the description and the partial differential equations are stepped through time to predict the movement of the fluids through the reservoir.

Wells can be used to produce fluid from the reservoir or to inject fluid into the reservoir. Technology can control only what happens at the wells and even then not absolutely. Nature controls everything in the reservoir between the wells. Also while technology can control what fluid is injected into a well, a producing well can produce only those fluids in its immediate vicinity in the reservoir. The Black Oil Simulator predicts how the fluids move in the reservoir based on the reservoir properties and how hard the wells are produced.

For example, Figure A2 (in Appendix) shows what would happen if the reservoir were produced with no injection, resulting in the loss of reservoir pressure. In this figure, blue represents oil or water and yellow represents gas. In this case, oil production is reduced considerably, because gas is produced preferentially over oil or water. Figure A3 (in Appendix) shows what would happen if water injection and fluid production is balanced. In this figure green represents oil, dark blue represents 100% water and light blue represents mostly water surrounding irreducible oil. In this case, much more oil is produced because water effectively displaces and takes the place of oil.

The results of the Black Oil Simulator consist of time series of the pressure and the amount of oil, gas, and water present at selected points in the reservoir and the production rates of oil, gas, and water out of, and injection rates of water into, each of the wells. Each of these properties can be presented as tables of numbers or in graphical displays. The graphical displays aid understanding of the physical processes occurring in the
reservoir. The mountain of tabular data can be used for quantitative economic calculations. The pressure and the amount of oil, gas, and water present in the reservoir describe the "state" of the reservoir. The pressure and the amount of oil, gas, and water present in the reservoir around the wells and the pressure in the wells determines the ability of the wells to produce or inject fluids. Thus, the performance of the wells can be taken as a measure of the "state" of the reservoir. However, the performance of the wells does not include the "state" of the entire reservoir, just the area around the wells. For this reason, correlating the performance of the producing wells against the performance of the injection wells is not always successful.

The typical operational procedure in Saudi Arabia is to drill producing wells down dip of the crest and injection wells below the oil/water contact. Only enough producing wells are drilled so as to meet the targeted production rate. The injected water travels across the reservoir until it encroaches upon the producing wells. When water reaches the producing wells, the oil production rate falls. Eventually, the wells are not able to meet the production target and new wells have to be drilled up dip to maintain capacity. Successive wells are drilled until the wells at the crest of the formation water out. A suite of simulations was run to investigate the performance of the reservoir for a range of production and injection levels. The amount of production and injection was balanced in each run. Any changes to the rates were made to the producers and injectors at the same time.

The scheduling of the wells in the model takes place in the following fashion. There are seven wells in the model: five producing wells and two injection wells. The sequencing of the wells begins with the producer lowest on the structure and ends with the fifth producer, which is drilled on the crest. The injectors are both drilled below the
oil/water contact. Both of the injectors, as well as the first producing well, are drilled at the same time and thus represent the initial investment. Subsequent capital investments are made as the additional producers are drilled in response to the watering out of previous wells. Figure A4 (in Appendix) shows a projection from above of the wedges of the reservoir and the placement of the five producing wells and the two injector wells. Figure A5 (in Appendix) shows the placement of these wedges as they are horizontally stitched together side by side to approximately cover the Ghawar reservoir’s northern portion of Aindar. The eighty-four wedges in the model cover an area of about 10 × 40 kilometers which is comparable to Aindar’s 10 × 30 kilometer size. The main portion of Ghawar is about 180 kilometers wide. Reservoirs that are close by, and which are comparable to the representative reservoir we model, include Abuaiq (15 × 40 kilometers) and Harmaliyah (8 × 15 kilometers).

Results of five of those simulations are shown below. The production and injection rates were held constant throughout each of the runs. At the lowest level of production target as shown in Figure A6, only one well is needed to meet the target rate. As each of the wells waters out, one new well is drilled. However, the deliverability of the next to highest well is insufficient to meet the target and the highest well has to be drilled at the same time. At the highest level of production target as shown in Figure A10, only the first well is able to meet the target rate by itself. When the first well waters out, all of the remaining wells have to be drilled. At the intermediate production rate targets, shown in Figures A7-A9, the results lie in between these two extremes. In general, the higher the target production rate, the quicker the wells have to be drilled. The capital investment schedule for each of these cases would be different, because of the different schedule for
drilling the wells. There is a trade-off between deferring capital investment at lower production rates against the consequent loss of revenue.

Twenty-four additional sets of simulations were performed to investigate the effect of production rates on well requirements. Each simulation was started at one of the 5 levels of production target just described. After a length of time at that initial rate, the rate was either increased to the next higher, or decreased to the next lower, level of production target and then maintained at that rate for the remainder of the run. Runs were made with the rate change at 2, 4, and 6 years. In all, 29 simulations were run. Based on these sets of simulations, we then accumulated the temporal production, water injection, and well drilling schedules. Using the simulation results, we estimated two important functions that comprise our dynamic production function (at the full capacity) for the oil field – a water injection function and a short-term capacity function.

The water injection rate is modeled as a function of oil output and the number of producing wells as follows:

\[
\log W_i = \delta_0 + \delta_1 \log X_i + \delta_2 \log (N^*_i) \tag{3.6}
\]

where \( N^*_i = N_i + dN_i \), which is the total number of producing oil wells.

The regression results are reported in Table 3.3.

<table>
<thead>
<tr>
<th>Table 3.3: Estimation of the Water Injection Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\delta}_0 )</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Estimates</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
<tr>
<td>( R^2 )</td>
</tr>
</tbody>
</table>

Given the results, we can rewrite the water injection function as,
\[ W_i = e^{0.7999 X_i^{0.9509}} N_i^{0.0306} \] (3.7)

The Black Oil Simulation program essentially applies the results from field dynamics to facts about the geological structures involved. The estimated relation (3.7) could thus be viewed as revealing the physical relationship between oil extraction and water injection in the field.

The dynamic\(^3\) production function (at the full capacity) - of an average oil well in the field is modeled in the following semi-log form.

\[ \max(X_i) = \lambda_0 + \lambda_1 \log W_i + \lambda_2 \log W_i \log CP_i + \lambda_3 \log CP_i + \epsilon_i \] (3.8)

where \(\max(.)\) denotes the maximum producing capacity in the neighborhood of an oil well in the reservoir. The estimation of (3.8) yields an approximated feasible set for oil production dependent on the reservoir engineering conditions.

Regression results are reported in Table 3.4.

<table>
<thead>
<tr>
<th></th>
<th>(\lambda_0)</th>
<th>(\lambda_1)</th>
<th>(\lambda_2)</th>
<th>(\lambda_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates</td>
<td>0.0451</td>
<td>0.0362</td>
<td>-0.0038</td>
<td>-0.0044</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.00023</td>
<td>0.00053</td>
<td>0.000064</td>
<td>0.000027</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.7950</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The estimated coefficients are consistent with the hypotheses that short-term overproduction will jeopardize the producing environment of a particular well and also that water injection generally has a linear and positive relation to the amount of oil that can be extracted. However, if the cumulative production of a well is too high, it is

\(^3\) 'Dynamic' refers to the effects of current production on future capacity and choices of policy variables.
possible that water injection could further reduce the short-term capacity as described previously.

Since the dynamic production function is a function of the state variables in our optimization model and describes the feasible production set, equation (3.8), extended to the entire field, is used as a short-term capacity constraint in our model, which is function \( f \) in (3.1).

\[
X_t \leq f(W_t, N_t^*, CP_t)
\]

\[
= \left\{ 0.0451 + 0.0362 \log(W_t) - 0.0038 \log(W_t) \log(CP_t) - 0.0044 \log(CP_t) \right\} \cdot N_t^* \quad (3.9)
\]

Note that the short-term field production capacity is not simply the function \( f \) due to the water injection equation (3.6). However, equation (3.8), along with (3.7) implicitly determine the maximum amount of oil that can be produced from a typical Saudi oil field as a function of those reservoir state variables. The reason why we also call (3.8) the 'short-term' field capacity function is not only because the state variables (\( CP \) and \( N \)) are functions of decision-makers' choices and thus evolve over time but also because \( W_t \) is a time-varying indicator for the reservoir fluid pressure and is affected by the choice variables directly and indirectly.

3.3.4 Summary

The results derived previously in this section can be summarized by restating the dynamic programming model (3.1) as:

\[
V_t(N_t, CP_t) = \max_{X_t, dN_t} \left\{ 365X_t \cdot e^{3.5323 - 0.0398 \times 37 \times X_t + 3.9656} - 1.2((0.44 \times 365)X_t + \right\}
\]
\[(0.7714(365X_t)^{-0.243}) \times (365X_t) + 365 \times (0.2W_t) + 0.1783N_t + 2dN_t + \beta V_{t+1}(N_{t+1}, CP_{t+1})\] 

subject to 

\[N_{t+1} = N_t + dN_t\]  
\[CP_{t+1} = CP_t + 365X_t\]  
\[W_t = e^{0.7999} X_t^{0.9509} N_t^{-0.0306}\]  
\[X_t \leq \left\{ 0.0451 + 0.0362 \log(W_t) - 0.0038 \log(W_t) \log(CP_t) - 0.0044 \log(CP_t) \right\} \cdot N_t^{*}\] 

and both \(X_t\) and \(dN_t\) are nonnegative.

The final piece of information needed in order to derive the optimal policy path is the functional form of the value function \(V\). In the literature, different iterative methods have been used to approximate the value function. For example, Hartley (1996) proposes polynomial approximation of the value function in models with inequality constraints as we have here. Our case is a little different, however, in so far as we have multiple state variables. We use tensor splines to approximate the value function on two-dimensional grids and a different approach to handle the inequality constraint. These issues are discussed fully in Section 5.

### 3.4 Theoretical Issues

Before we start to numerically approximate the solutions to the dynamic programming model in (3.10), we need address the existence and uniqueness of the optimal solution in the time-invariant state to ensure the iterative process used to approximate the value function is valid.

If any function in the dynamic programming model is non-stationary, there would be no hope of obtaining a solution using iterative technique. The optimal policy in period \(t\) in general depends on the optimal policy in all subsequent periods. We need to
characterize the non-stationarity in terms of a small number of parameters and then solve the model as a function of those parameters. The approach we take is to hypothesize that the oil market eventually settles down to a time-invariant state in terms of the price and quantity relationship (see equation (3.2)). In two of the four simulation scenarios, a competitive backstop technology (solar energy) is also assumed to control the demand for fossil fuel in the time-invariant state. Thus, we can solve for a "terminal value" of the oil revenues in the time-invariant state using iterative techniques. In the remainder of this section, we focus on the feasibility of the iterative procedure we use to solve for the optimal policy and value in the terminal competitive environment (time-invariant state). It is understood, however, that we will then use the solution to the terminal optimization problem to solve the problem for the optimal policy and value functions in the earlier periods.

First we introduce some notation. Let $\pi$ denote all the policy variables and $\sigma$ the state variables. $\Pi$ is used to represent the set of all the possible values for the policy variables (i.e., $\pi \in \Pi$), $\Sigma$ stands for the set of all possible values for the state variables (i.e., $\sigma \in \Sigma$), and let $\Phi: \Sigma \to \Pi$ be the correspondence describing the feasibility constraint. Notice that the correspondence $\Phi$ here is implicitly defined by the short-term capacity function (3.9) as an upper bound and nonnegativity of policy variables as the lower bound. $\hat{B}$ denotes the profit function (hat $\wedge$ stands for the estimated function in this chapter) that is equal to $r - c$ in (3.1). Hence $\hat{B}$ is a function mapping from $\Pi \times \Sigma$ to $\mathbb{R}^1$, where $\mathbb{R}^i$ denotes the $i$-dimension real number space. $M$ stands for the state transition function, $M: \Pi \times \Sigma \to \Sigma$. 
Given the notation defined as above, the dynamic programming model in (3.10) can be expressed like (3.11) or (3.12), as an operator \( T \) on the space of continuous and bounded real functions.

\[
(T \cdot \nu)(\sigma_t) = \max_{\pi_t \in \Phi(\sigma_t)} \left\{ \bar{B}(\pi_t, \sigma_t) + \beta \nu(\sigma_{t+1}) \right\} 
\] (3.11)

or

\[
(T \cdot \nu)(\sigma_t) = \max_{(\sigma_{t+1}, \sigma_t) \in \Phi(\sigma_t)} \left\{ \bar{B}(\sigma_{t+1}, \cdot, \sigma_t) + \beta \nu(\sigma_{t+1}) \right\} 
\] (3.12)

where \(-,\) indicates that the subtraction follows the state transition function \( M \).

Theorem 3.1 Let \( \Phi(\sigma) \) be the space of bounded continuous functions \( g: \Sigma \to \mathbb{R}^1 \), with sup norm. Then operator \( T \) defined in (3.11) has a unique fixed point \( \nu^* \in \Phi(\sigma) \), and for all \( \nu \in \Phi(\sigma) \),

\[
\left\| T^n \nu_0 - \nu^* \right\| \leq \beta^n \left\| \nu_0 - \nu^* \right\|, \quad n = 0, 1, 2, \ldots
\]

Moreover, given \( \nu^* \), the optimal policy correspondence \( P: \Sigma \to \Pi \), defined as

\[
P(\sigma) = \left\{ \pi \in \Phi(\sigma): \nu^*(\sigma) = \bar{B}(\pi, \sigma) + \beta \nu^*(M(\pi, \sigma)) \right\}
\]

is compact-valued and upper hemi-continuous \((u.h.c.)\).

Proof: included in Appendix A.

Theorem 3.1 shows the existence of the fixed point for the operator \( T \) and hence allows us to use iterative methods to approximate the value function in the time-invariant state that is introduced in the next section. The theorem also provides a bound on the rate of convergence. The following corollary examines the monotonicity of the value function \( \nu^* \) in the time-invariant state. We first need to rewrite one of the state variables so that standard methodology can be employed. We denote \( CP^- \) as \( CP^- = -CP \) and substitute \( CP^- \) back into the dynamic programming model (3.10). All the functions involving
CP are redefined as a function of CP⁻. This transformation obviously has no effect on \( v^* \) or the optimal policy.

**Corollary 3.1** Let \( v^* \) be the unique solution to our dynamic programming model (3.11) in the time-invariant state. Then \( v^* \) is strictly increasing in both state variables \( N \) and \( CP^- \).

**Proof:** included in Appendix A.

Hence, \( v^* \) should be strictly increasing in \( N \) and strictly decreasing in \( CP \). The second part of the result follows from the transformation of \( CP \) to \( CP^- \). The simulated value functions, which are presented in the next section, are consistent with Corollary 4.1.

### 3.4 Simulation Results

Numerical approximations to the value function in the time-invariant state, and the corresponding optimal extraction path and the number of new wells, are presented in this section for four scenarios. The first scenario is the one described in (3.10) with a discount rate equal to 0.9 and the time-invariant states beginning in 2045. The second scenario sets the discount rate to 0.7 to capture the heavy reliance of Saudi Arabia on its short-term oil income. In effect, oil revenues as an asset are quite risky and hence the implicit rent from leaving them in the ground has to yield a substantial risk premium. The optimal extraction policies under the two discount rates thus reflect the possible effects on long-term oil policy of relying on an undiversified income structure. In the third scenario, a technology breakthrough significantly reduces the cost of solar energy. The technological shock is assumed to happen in 2036 and continue to improve, so there is a ten-year period
for solar energy to totally replace crude oil in energy consumption. In the meantime, the demand for oil from the petrochemical industry is assumed unchanged. Only the revenue function is affected by such a shock to demand. The transition pattern is purely hypothetical. We simply assume that, after technology shock, the intercept term in (3.2) starts to decline by ten percent each year and ends at five percent\(^4\) of its original value in the end of transition period. The focus here is not on how the transition would happen, but on how the expectation of such a transition could affect Saudi's oil production policy. The last case is simply a repetition of scenario 3 with a different timing of the solar technology shock. We assume the shock will occur in 2026 instead of 2036, in order to check - by comparing the two simulated optimal production paths - how the impacts of such a technology shock on Saudi's oil production decisions might depend upon the expected timing of the shock.

The numerical approximations to the value function in the terminal time-invariant state for each scenario are presented in the graphs on the next two pages.

\(^4\) 5% is used to account for the remaining demand from the petrochemical industry after the shock.
Figure 3.1: Approximated Value Function in the Time-Invariant State

- Figure 3.1: The value function in the time-invariant state is approximated in the scenario 1 with no assumed solar technology shock and the discount rate $\beta = 0.9$.

Figure 3.2: Approximated Value Function in the Time-Invariant State

- Figure 3.2: The value function in the time-invariant state is approximated in the scenario 2 with no assumed solar technology shock and the discount rate $\beta = 0.7$. 

Figure 3.3: The value function in the time-invariant state is approximated in scenario 3 and 4.

Notice that changing the timing of solar technology shock from 2036 to 2026 would have no effect on the value function $v^*$ in the time-invariant state, though the values of the state variables at the time of the shock would, of course, be different. In the following figures, we present the corresponding production paths and the trajectories of new wells drilled in the field associated with the optimal oil production policy for each scenario. The initial conditions are set as: $N_0 = 57$ and $CP_0 = 2648$ million barrels. $CP_0$ is calculated using the Saudi’s production data 1976-1985. We have no exact data on the total number of on-shore producing oil wells in 1986. We instead approximated it using the information that Saudi has about 555 producing wells in the mid-80's. The paths for value function were derived by solving the optimization problem one period at a time moving backwards to the middle 1980's.

---

5 The information was obtained from a paper at the following web set: http://lcweb2.loc.gov/cgi-bin/query/D?cstdy;2:/temp/~frd_FXn3::; and the original data source is The Library of Congress Country Studies.
Figure 3.6: Optimal Extraction Rates in Scenario II

Figure 3.7: Optimal Number of New Wells in Scenario II
Figure 3.8: Optimal Extraction Rates in Scenario III

Figure 3.9: Optimal Number of New Wells in Scenario III
In the next figure, the simulated optimal production paths (1986 – 1995) in different scenarios are compared with 10% of the Saudi true daily extraction rates during this period. The proportion ten percent is based the size of the reservoir studied in the Black-Oil simulation and the fact that the initial conditions were chosen under the assumption that the field accounts for about 10% of total Saudi oil reserves and output.

![Figure 3.12: Comparison of Simulated Optimal Productions with the Real Path](image)

A Comparison of the optimal oil production policy paths under different simulation scenarios indicates that the discount rate does have a significant effect on the optimal extraction rates and the number of new wells drilled for capacity expansion. A lower discount rate like 0.7, which might indicate a high demand for short-term oil income, would lead to a higher production rate in the short-run but a lower capacity in the future compared with the outcome when \( \beta = 0.9 \). Figure 3.12 suggests that Saudi oil production is relatively better approximated by scenario I \( (\beta = 0.9) \) before 1990. After Gulf crisis occurred in 1989 - 1990, scenario II \( (\beta = 0.7) \) approximates the real path better, and
actually in 1990 the two paths coincide. This is not difficult to explain as we recognize that there was a large increase in Saudi government expenditure (especially military expenditure) around 1990 (see Table 3.1).

Comparing the simulation results from scenario I with those of scenario III tells us that the expectation of a future technology breakthrough in solar energy will act like a smaller discount factor and tend to increase production for a short period and reduce capacity subsequently. The similarity arises because in both these scenarios the opportunity cost of current production in terms of future oil income is relatively reduced compared with scenario I. However, the two scenarios have different impacts on the number of new oil wells drilled for expansion. The decision makers in scenario II are much more concerned about short-run oil profits and do not care about the future production as much as the decision makers in scenario III. Furthermore, under scenario IV, where Saudi decision makers expect the technology shock to occur even earlier, production would be even higher in short-run while production capacity would be maintained or even expand (for a period) by drilling more oil producing wells. Under this scenario, the decision-makers foresee a sharp decline in the return on oil reserves due to the (earlier) substitution to an alternative energy.

Another interesting question, illustrated in Figure 3.13, is what happens to the oil price paths in each of these four scenarios and how they compare to the actual Saudi price path.
The actual price path is more volatile than any of the four simulated price paths. Of course the simulated price paths should be thought of as the mean of the potential sample paths. Even so, the volatility of the actual path might reflect other factors affecting Saudi decision-makers but not taken into account in our model. The apparent flat trend in actual prices compared with the rising trend in the simulated paths is more difficult to explain\(^6\). Quite possibly it reflects a slower growth in the world economy, and hence energy demand, than the OMS model had anticipated.

3.6 Conclusion

We have proposed and illustrated a detailed economic and engineering-based methodology to model the dynamic production decisions of an idealized oil field representing, in many ways, the largest light oil field in Saudi Arabia - Ghawar. Our

\(^6\) The rising trend of the simulated price path, however, is not surprising given the estimated demand. It is consistent with the prediction of Hotelling's condition (see Prato (1998), p. 140, or Neher (1990), p. 95).
analysis incorporates a game theoretical structure of the world oil market through OMS simulation data. The results of the optimal production model approximate actual Saudi extraction rates. Perhaps of more interest, however, a comparison of the results under different scenarios helps our understanding of the possible effects of Saudi dependence on risky oil income and possible future demand reductions resulting from the development of alternative energy.

It is well known that there are other objectives embedded in Saudi oil policy. The simulation results allow us to estimate the potential costs of pursuing those other objectives in terms of the foregone profits. Since production cost is low relative to oil revenue in Saudi Arabia, we present only the simulated revenue losses in Table 3.5, which indicates the magnitude of profit losses.

Table 3.5: Comparison of Simulated Oil Revenues (million dollars) with the True Ones

<table>
<thead>
<tr>
<th></th>
<th>Scenario I</th>
<th>Scenario II</th>
<th>Scenario III</th>
<th>Scenario IV</th>
<th>Real Rev.</th>
<th>Losses*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>32004</td>
<td>33171</td>
<td>33016</td>
<td>33315</td>
<td>28489</td>
<td>-11%</td>
</tr>
<tr>
<td>1987</td>
<td>34612</td>
<td>35456</td>
<td>34899</td>
<td>35498</td>
<td>25739</td>
<td>-26%</td>
</tr>
<tr>
<td>1988</td>
<td>37207</td>
<td>37891</td>
<td>37877</td>
<td>37788</td>
<td>23720</td>
<td>-36%</td>
</tr>
<tr>
<td>1989</td>
<td>39924</td>
<td>40493</td>
<td>40473</td>
<td>40618</td>
<td>33526</td>
<td>-16%</td>
</tr>
<tr>
<td>1990</td>
<td>42744</td>
<td>43268</td>
<td>43112</td>
<td>43397</td>
<td>55188</td>
<td>29%</td>
</tr>
<tr>
<td>1991</td>
<td>45710</td>
<td>46236</td>
<td>46116</td>
<td>45978</td>
<td>47345</td>
<td>4%</td>
</tr>
<tr>
<td>1992</td>
<td>48844</td>
<td>49406</td>
<td>49327</td>
<td>47664</td>
<td>51574</td>
<td>6%</td>
</tr>
<tr>
<td>1993</td>
<td>52119</td>
<td>52784</td>
<td>52707</td>
<td>52588</td>
<td>35787</td>
<td>-31%</td>
</tr>
<tr>
<td>1994</td>
<td>55665</td>
<td>56390</td>
<td>56297</td>
<td>56188</td>
<td>46985</td>
<td>-16%</td>
</tr>
<tr>
<td>1995</td>
<td>59545</td>
<td>60159</td>
<td>60137</td>
<td>60044</td>
<td>52242</td>
<td>-12%</td>
</tr>
</tbody>
</table>

* Losses in percentage are calculated compared with the revenue under Scenario I, the lowest one in all simulations.

We observe a substantial potential for revenue (and hence profit) loss from pursuing what appears to be less than optimal policies from a purely profit maximizing perspective. However, there are a couple of years (1990-1992) in which the actual revenues are higher than any of the simulated ones. This could be attributed to the exogenous shock of the Gulf War, which resulted in a rise in oil price (for a short period) as well as Saudi oil
sales, as can be seen in Figure 3.12 and Figure 3.13. Our model does not take into account the effects of such shocks.

Although the model does quite well in approximating Saudi production decisions in its largest light oil field, we acknowledge there are still many simplifications and approximations in our modeling. Some of these would be worth rectifying in a more detailed study aimed at informing policy. Some of the defects may also direct us toward future improvements of the model. First, as pointed out by Powell (1990), the assumption of perfect knowledge and foresight may appear to be unrealistic. We would like to extend our model in the future to encompass the stochastic demand environment while randomizing the timing of a solar technological breakthrough. Second, instead of modeling the Saudi production decision purely as maximization of its dynamic profits, we could apply a multi-level optimization (MLO) approach to incorporate profit maximization by local oil firms along with other motivations of the Saudi Arabia government. This type of approach might approximate the oil production decision more accurately. Islam (1998) provides a good example of using MLO to model energy plans involving both the private sector and the government. Last, but not the least, while the OMS model allows us to simplify the model, a more ambitious model might attempt to improve the OMS model of the world oil market and incorporate strategic behavior directly into the dynamic optimization model.
Chapter 4

The Effects of Oil Price Volatility on Technical Change

4.1 Introduction

In his 1983 paper, James Hamilton convincingly put forth the idea that oil price has a significant negative effect on output. Noting the empirical fact that virtually every recession during the post-World War II period was preceded by a shock to the price of oil, he effectively demonstrated that oil prices Granger-cause output. Since then the observation of a negative correlation between crude oil prices and output has become widely accepted, but the notion of a causal relationship from oil price to output has been the subject of much debate. In particular, the idea that oil prices are directly responsible for causing changes in output has not been universally accepted, and many different possible explanations for the observed negative correlation have been proposed.

Recently, a great deal of attention has been given to the effect that oil price volatility has on aggregate output. The work of Bernanke (1983), Pindyck (1991), Dixit and Pindyck (1994), and Hamilton (1988), to name a few, have extended theories which have paved the way for this line of research. These theories' applicability to the oil market relies on two phenomena which were virtually non-existent in the regulated market environment that predominated prior to the 1980's: unexpected oil price changes and high levels of volatility of the price of oil. Several authors have attempted to test these theories using empirical methods. For example, Lee, Ni and Ratti (1995) employ a GARCH measure of variability of the forecast error of oil price to find that unexpected changes in oil prices are negatively correlated with output. Ferderer (1996) uses a vector
autoregressive analysis to find that oil price volatility has a greater negative effect on industrial production than does absolute changes in oil price.

While it has been demonstrated by previous authors that oil price volatility is significant in determining output, we ask the question: is it possible that increased price volatility is causing a delay in the implementation of production technologies which could potentially increase a country's productive capacity? If so, this would suggest that the innovation displayed by a specific country is highly correlated with its dependence on energy resources, or in the context of this research, oil resources. We investigate the possibility that displayed innovation, or technical change, is driven by reactions to changes in the oil market. More specifically, we are interested in determining whether or not it is the price of oil, the volatility of the price of oil, or other economic factors that drive movements in displayed innovation.

We will begin with a brief discussion of the different proposed channels of transmission from oil prices to the macroeconomy in order to further motivate our discussion. We then embark on the technical aspects of our analysis by specifying a measure of volatility and quantifying technical progress using a measure of innovation developed by Färe et al. (1994a,b). A vector autoregressive (VAR) analysis is used to investigate the existence of any significant relationships between the included variables. We next present our results, and conclude with some remarks about the implications of our findings.
4.2 Channels of Transmission and Oil Price Volatility

While the impacts which oil prices have on aggregate output are not the focus of this study, given that output is a component of any productivity measure, it is not unreasonable to predict that oil prices will effect productivity in a similar fashion as they effect output. Since the measure of technical change to be used in our analysis is decomposed from a measure of productivity, it is important to understand the proposed channels of transmission through which oil prices effect the macroeconomy. These channels are elegantly summarized by Ferderer (1996), so they only need a brief exposition here.

Two demand-side propositions for the observed negative correlation between oil prices and output focus on the effect that oil price shocks have on consumer expenditures. First, the real balance channel claims that increases in the price of oil lead to inflation thereby lowering the quantity of real balances in the economy. This generates an economic recession by suppressing aggregate demand. Second, the income transfers channel posits that oil price increases cause income transfers from oil importing countries to oil exporting countries. This, in turn, causes rational agents in the oil importing countries to reduce their consumption thereby depressing total output.

Authors such as Berndt and Wood (1979), Fischer (1985), and Rasche and Tatom (1981) suggest a supply-side argument, which we will call the complements channel. The complements channel suggests that if oil and capital are complements in the production process then oil price increases will induce a reduction in the utilization of both capital and energy. This would have a suppressing effect on aggregate output.
With the more recent findings of Mork, Olsen, and Mysen (1994), it has become evident that oil price increases and oil price decreases effect the economy in an asymmetrical manner. Output responds negatively to oil price increases, but does not respond positively in a proportional manner to oil price decreases. Other authors have since verified this conclusion (see Ferderer (1996)), and, in fact, some studies have gone so far as to investigate the cause of the asymmetry (see Huntington (1998)). Therefore, the validity of the real balances channel, the income-transfers channel, and the complements channel is called into question as these would presumably each produce symmetric effects on output. The investigation of this issue is beyond the scope of this research, however, we will use the results of previous authors to narrow the focus of our study. Specifically, given the inability of the above channels to explain the asymmetry puzzle, we will not investigate their implications. Instead, we will focus on three channels that can potentially explain the asymmetry puzzle.

The first of these channels, the monetary policy channel, has been advocated by authors such as Darby (1982) and Bohi (1991). This channel stipulates that counter-inflationary monetary policy that is enacted in response to oil price increases drives the observed declines in output. Hence, shocks to the price of oil are only indirectly responsible for recessions.

The remaining two channels that have been suggested to explain the negative correlation between oil price and output place little importance on the price of oil itself. The sectoral shocks channel assumes that it is costly to shift specialized labor and capital between sectors, therefore oil price increases can decrease output by decreasing factor employment. For example, given a recession is not unreasonably long, the high costs of training will cause a specialized labor force to wait until conditions improve rather than
seek employment in another sector. Hamilton (1988) has discussed this avenue by constructing a multi-sector model in which there is some cost to labor reallocation. He demonstrates that relative price shocks in his model will generate reductions in employment. An implication of this theory is that increases in price volatility (characterized by a high frequency of relative price changes) will cause aggregate unemployment to increase.

The uncertainty and investment channel asserts that in the face of high uncertainty about factor prices it is optimal for firms to postpone irreversible investment expenditures. Bernanke (1983) develops a model that demonstrates precisely this point, and the theory is discussed in extensive detail in Dixit and Pindyck (1994). Oil price shocks generate some degree of uncertainty about the future movement of prices creating an environment in which this channel could explain declines in output. Investments are said to be irreversible when they are firm or industry specific. Firm specific examples include advertising and marketing expenditures because these are by-and-large not recoverable. An industry specific example is the construction of a steel plant. Although the plant can be sold if times are bad, if the steel industry is reasonably competitive bad times for one manufacturer will be bad for another. Thus, the probability of resale would be low because the plant would be equally useless to a competitor.

Mork (1989) demonstrates that the results obtained by Hamilton concerning the relationship between oil prices and output are weakened if his data set is extended to include the early to mid 1980's. An explanation for this result is that prior to the early 1980's market regulation limited the unpredictability of price. In fact, Hamilton claims that the unique environment that existed allowed for dramatic one-time movements in price.
...the discrete, dramatic pattern of crude oil price changes... is explained by the specific regulatory structure of the oil industry over 1948-1972. Each month the Texas Regulatory Commission (TRC), and other state regulatory agencies like it, would forecast demand for the subsequent month and would set allowable production levels... the commission was generally unwilling or unable to accommodate sudden disruptions in supply, preferring instead to exploit these events to realize the dramatic price increases.... (p 230).

The end result was the purging of any cyclically endogenous component of the price series resulting in a highly predictable environment. However, with the relaxation of these types of controls, it can be argued that the price of oil has been subject to a higher degree of uncertainty. Hence, it has been asserted that Mork’s findings are the result of increased volatility in the price of oil. In other words, because the price series in Hamilton’s analysis is characterized by low volatility, reactions to increases in price volatility are by-and-large unobservable. Only in periods where oil price shocks occur will there be any significant levels of price volatility. This would explain why Hamilton's focus on price changes would lead to the conclusion that oil price changes Granger-cause output for the pre-1980's. Periods in which prices changed exactly coincided with periods in which there was any substantial price volatility. In other words, price shocks created some uncertainty about the future movements of the price path.

The extended oil price series in Mork’s analysis is characterized by a relatively higher degree of volatility. This can be tied to the fact that prices move more freely in the deregulated market, so the price path must undergo some adjustment as shifts in its determining factors are realized. Therefore, a continued focus on oil price changes will lead to a reversal of Hamilton’s conclusion because there is an increase in the frequency
of price changes but not in output movements thus weakening the correlation between the two. Recent theoretical and empirical research (discussed above) suggests that it is more appropriate to focus on oil price volatility as it is the better predictor of movements in output. It is stipulated that price shocks generate uncertainty over future price movements. Thus, an increasing frequency of price changes will generate increasing uncertainty. As posited by both the sectoral shocks channel and the uncertainty and investment channel, increasing uncertainty will have adverse effects on output. The degree of uncertainty that is generated in the oil market is best approximated by oil price volatility. Hence, evaluating the macroeconomic impacts of oil price volatility is one way of analyzing the effects of increasing uncertainty.

4.3 A Measure of Volatility

Evaluating the uncertainty that is faced over the path of oil prices requires an adequate measure of volatility, and there currently exist capable means of calculating volatility within the econometric framework. One example is the use of GARCH models in financial markets. Forecasts of asset price volatility can be obtained in order to evaluate the risk of said asset. It is precisely the parametric nature of these measures that permit such analysis. It should be noted that the Ljung-Box Q-statistics for the squared residuals of the ARMA(1,1) process are not indicative of any ARCH or GARCH errors. Similarly, results obtained using the LM test that was suggested by Engle (1982) do not permit the rejection of the null hypothesis of no ARCH or GARCH errors. Hence, it would seem that the use of such a model is not appropriate for the oil price series.
We suggest a nonparametric measure of volatility designed not to facilitate forecasting, but to evaluate past movements of a particular series. In this regard, it can be viewed as a property of the series, much like the mean or standard deviation. The measure is constructed based upon an intuitive notion of the concept of volatility. We wish to capture the idea that volatility is a description of both the rate and magnitude of the changes realized by the series in question. Therefore, it is necessary to obtain a continuous time approximation of the oil price path. This will enable the evaluation of derivatives of the oil price series at each observed data point, giving us the rate of change. In order to do this, we employ a cubic smoothing spline to the series and evaluate the derivatives at the discrete data points. The measure is defined as:

\[
\text{Price Volatility} = \left( \left| \frac{\partial P(t)}{\partial t} \right|_{t+1} - \left| \frac{\partial P(t)}{\partial t} \right|_t \right)
\]  

(4.1)

Thus, we define volatility as the absolute value of the change in the rate of change.

An illustration of this measure is seen in Figure 4.1 along with oil price. We see spikes in volatility when there are rapid changes in the oil price path. Furthermore, consistent with numerous previous claims, the average level of oil price volatility has increased since the beginning of the 1980’s. In fact, the mean volatility has approximately quadrupled. Given a priori expectations of oil price volatility, it is apparent that the measure presented above performs rather adequately.

---

7 The ARMA(1,1) specification is used because the Akaike Information Criterion and the Schwartz Bayesian Criterion each indicate that it is the most appropriate structure for the first differenced oil price series.
8 A cubic smoothing spline is a nonparametric method of polynomial approximation applied to discrete data observations with a smoothing parameter that is chosen so that the smoothing spline contains as much of the information and as little of the supposed noise in the data as possible.
9 On an intuitive level, if there is a perceived entropy to the system, one can naively approximate the path of the series tomorrow with the path of the series today. Therefore, a simplistic point of view is that this measure is the absolute value of the deviation from the expected path.
In order to prevent criticism of the results of the remainder of this study based solely on the measure of volatility used, other measures of volatility have been analyzed. In particular, the smoothed variance of the residuals of an ARMA(1,1) specification of the first difference of oil price has also been used as a measure of volatility. As can be seen from Figure 4.2, this method produces a very similar measure to our own with the major difference being the relative absence of volatility in periods when there is not a shock to the price of oil. It turns out that using this alternative specification do not alter the results presented herein. Furthermore, alternative measures of volatility presented by Federer (1996) and Fleming and Ostdiek (1998)\footnote{Federer uses daily averages of standard deviations in monthly frequencies. He points out that this has the advantage of improved accuracy due to increased sampling of the stochastic process. Fleming and Ostdiek develop a measure that is similar to an EGARCH model, but is stochastic rather than conditional on past values. They point out that “this feature is attractive because the information flow to financial markets is unpredictable and it is information that generates volatility (p 9).”} demonstrate very similar patterns as the measure presented here.
4.4 A Measure of Innovation (Technical Change)

We begin by analyzing a measure of productivity presented by Färe et al. (1994a,b). They develop an output-based measure of productivity growth that is the geometric mean of two Malmquist productivity indexes. A reference technology (or frontier) for each time period is constructed from the output/input data for all countries in the sample, and then each country is compared to it. The vehicle for comparison is the output distance function, defined here as the inverse of the Farrell (1957) measure of technical efficiency. By evaluating distance functions using different time period input and output mixes, the productivity index can be constructed and decomposed into efficiency change and technical change, as is illustrated in Figure 4.3. For example, $D_0^* t (x^t, y^t)$ is the output distance function evaluated using inputs from time period $t$ and outputs from time period $t + 1$. Evaluation of similar measures defined, as in Figure 4.3, allows for the construction of the relevant measures. These can be used to ascertain the direction that each country is moving relative to the frontier, and whether changes in efficiency or in technology are motivating changes in productivity.
The diagram in Figure 4.3 is constructed assuming a constant returns to scale technology for a simple two-dimensional case. $S(t+1)$ and $S(t)$ are the production frontiers in period $t$ and $t+1$ respectively. They represent input/output combinations of the best practice firm, or country (as in our case), within the sample. For illustrative purposes, assume that we are interested in the country that has output $Y(t)$ and input $X(t)$ for time period $t$. The distance function is expressed as $D'(x', y')$ and can be described by the ratio of the line segments $\frac{0a}{0b}$. Any value less than unity indicates that the country is operating off the frontier. We can construct the other distance functions in a similar fashion in order to construct the desired index measures, as given in Figure 4.3. A value of unity for the index indicates that there is no change in productivity from the previous period. Values exceeding unity indicate an increase in productivity, and values less than unity indicate a decrease in productivity. The same holds for the derived measures of efficiency change and technical change. Solutions of the output distance functions required for index construction are obtained for each period by solving the appropriate linear programming problem. The interested reader is referred to Färe et al. (1994a,b) for extensive detail. An example of how this measure can be applied can be found in Alam and Scickles (1998).  

11 They construct a methodology to test the hypothesis that competitive pressures enhance technical efficiency in the airline industry.
Figure 4.3: The Malmquist-based Index and Decomposition by Distance Function

Productivity:

\[ M_0(x^{t+1}, y^{t+1}, x', y') = \left( \frac{D_0'(x^{t+1}, y^{t+1})}{D_0'(x', y')} \times \frac{D_0''(x^{t+1}, y^{t+1})}{D_0''(x', y')} \right)^{1/2} = \left\{ \frac{Oe/Oc}{Oa/Ob} \right\}^{1/2} \]

Efficiency Change:

\[ EC_0(x^{t+1}, y^{t+1}, x', y') = \frac{D_0''(x^{t+1}, y^{t+1})}{D_0'(x', y')} = \frac{Oe/Of}{Oa/Ob} \]

Technical Change:

\[ TC_0(x^{t+1}, y^{t+1}, x', y') = \left( \frac{D_0'(x^{t+1}, y^{t+1})}{D_0''(x^{t+1}, y^{t+1})} \times \frac{D_0'(x', y')}{} \right)^{1/2} = \left\{ \frac{Oe/Oc}{Oe/Oe} \times \frac{Oa/Ob}{Oa/Od} \right\}^{1/2} \]
The appealing aspect of this construction is that we do not have to assume a functional form for the production process. The data enveloping technique used here generates the indices purely from the data for outputs and inputs; hence, the index is a 'primal' measure. In this sense, this index has a distinct advantage over other productivity measures because there is no prerequisite specification of an underlying production technology. This type of restrictive assumption is usually made when deriving other measures of productivity\textsuperscript{12}. For example, the Tornqvist, or discrete Divisia, index is built upon the assumptions of a translog functional form for production as well as the calculation of revenue or expenditure shares. Furthermore, unlike both the Tornqvist index and growth accounting measures of total factor productivity (TFP) the Malmquist-based index allows for allocative inefficiencies in production. It is implicitly assumed that production is efficient in the Tornqvist and TFP index measures. Thus, the Malmquist-based index allows for the differentiation of changes in allocative efficiency and technical innovation.

We allow for variable (VRS), non-increasing (NIRS), and constant returns to scale (CRS) in our construction, and calculate the index measures for a group of seven OECD countries (United Kingdom, Canada, Japan, Italy, Germany, United States, and France). We use capital stock, labor force, and energy consumption as inputs\textsuperscript{13}, and real GDP as the measure of output. The time period covered extends from the first quarter of 1966 to the third quarter of 1996.

\textsuperscript{12} An extensive discussion of index measures of productivity is found in Good, Nadiri, and Sickles (1997).
\textsuperscript{13} Glasure and Lee (1997) empirically demonstrate bidirectional causality between GDP and energy consumption for the cases of Singapore and S. Korea, which gives empirical support to the treatment of energy as an input to production.
Quarterly data for GDP is obtained from the OECD Quarterly National Accounts. To enable the cross-country comparison that is required for the index construction, we apply the purchasing power exchange rates given in the PENN World Tables 5.6 (PWT5.6). Quarterly values for labor force are obtained from various issues of OECD Main Economic Indicators. In order to obtain capital stock data, quarterly values are approximated. Quarterly capital formation data from OECD Quarterly National Accounts are converted to international units using the appropriate exchange rates from the PWT5.6, and combined with annual data on capital stocks, which are retrieved from annual data on capital stock per worker in PWT5.6, to calculate annual depreciation rates\textsuperscript{14}. The quarterly capital stocks are then approximated by applying the calculated depreciation rate to the existing inventories. This is done in order to preserve the integrity of the perpetual inventory method used in the estimation of the annual capital stocks given in the PWT5.6. The energy consumption data are approximated using quarterly petroleum consumption data and annual data for the percentage of petroleum in total energy. It is assumed that the annual percentages are constant for each quarter within that year.

The average values for the Malmquist-based productivity index, efficiency change index and technical change index for the constant returns to scale case are reported in Table 4.1. Also reported are the average index values for the time periods before and after the first quarter of 1980. We only report the values for the case of constant returns to scale (CRS) technology as they are the values used in the remainder of the analysis. The CRS results are chosen primarily based upon the comment of Ray and Desli (1997) and reply of Färe, Grosskopf and Norris (1997). In Färe et al.'s reply, they point out that

\textsuperscript{14} It is interesting to note that the average depreciation rate for all countries was found to be about 5%.
technical change is the change in the maximal average product from period to period, and that the "CRS technology captures maximal average product to specify and compute technical change (p 1041)". This holds true even if the underlying technology is VRS. They also note that the CRS construction is consistent with Solow's (1957) notion of technical change. Hence, the CRS results are attractive because they have a well-founded corollary in the economic literature. Furthermore, they allow for a comparison to the results reported in Färe et al. (1994b)\textsuperscript{15}.

\textsuperscript{15} The results for the VRS and NIRS cases are available upon request.
A cross-reference of the output distance functions (not reported) with the technical change index can indicate which country is shifting the frontier for a given period. The criterion are:

\[ \text{TechnicalChange} > 1 \]

\[ D'_0(x_{t+1}, y_{t+1}) > 1 \]

and \[ D'_0(x_{t+1}, y_{t+1}) = 1 \].
If all three of these conditions are met for a particular country, then that country contributed to a shift in the frontier between period $t$ and period $t + 1$. It is interesting to note that Färe et al. (1994b) report that the United States is the sole innovator in all periods. The inclusion of the energy consumption variable in our analysis actually shifts Italy to the frontier along with the U.S. for the post-1980 period\textsuperscript{16}. This is a by-product of the relative energy efficiencies of the countries included in our sample. For example, Italy has the highest level of output per unit energy input in the sample. Alternatively, the U.S. is ranked first and second in output per worker and output per unit capital, respectively, but almost last in output per unit energy. The relative rankings of these variables will greatly impact the relative rankings of each country in terms of productivity. Given the relative increases of output per worker and output per capital in Italy, coupled with a superior energy efficiency, it is not surprising that the efficiency index for Italy improves substantially in the latter half of the sample period. As a point of reference, the omission of the energy variable will produce a result very similar to that reported in Färe et al. in which the U.S. is the sole innovator in virtually every period.

4.4 Vector Autoregression

The next step of our analysis is to place the technical change measure for the US in a VAR framework with an indicator of monetary policy, the real oil price, and the oil price volatility. As explained above, we are interested in determining whether it is the price of

\textsuperscript{16} It is interesting to note that under VRS the UK, Japan, and Italy all join the US on the frontier.
oil, the volatility of the price of oil, or monetary policy that drives or hinders advances in the implementation of technologies that increase productive capacity\footnote{It should be noted that the technical change measure is exactly equal to the productivity measure for the US. This is due to the fact that efficiency change is unity for all periods indicating that the US is on the production frontier for the entire sample period.}.

Following the findings of Bernanke and Blinder (1992), who tested various interest rates in the US to determine which is the best indicator of monetary policy, we use the Federal funds rate as the indicator of monetary policy. The oil price series is constructed using monthly data, which are averaged to obtain quarterly figures. We use the Refiner's Acquisition Cost (RAC) for all periods after 1973 as it is not available for earlier periods. For all observations prior to this, we use the price for Saudi Light Crude. The two are linked around the OPEC oil embargo of 1974. The oil price volatility data was constructed using the methodology outlined above.

In order to determine the form in which to include the data in the VAR, we test each series for the presence of a unit root using the Augmented Dickey-Fuller test. The Federal funds rate and the oil price series exhibit that they possess one unit root. Oil price volatility and technical change exhibit no unit roots. The Likelihood ratio test suggested by Johansen (1988) was used to determine that there is no evidence of any cointegrating relationships. The results of these diagnostic tests are given in Table 4.2.
Table 4.2: Tests for Integration and Cointegration

<table>
<thead>
<tr>
<th>Augmented Dickey-Fuller (ADF) tests</th>
<th>levels</th>
<th>first differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>op</td>
<td>-1.686</td>
<td>-4.558</td>
</tr>
<tr>
<td>vl</td>
<td>-4.371</td>
<td>....</td>
</tr>
<tr>
<td>ff</td>
<td>-1.938</td>
<td>-5.914</td>
</tr>
<tr>
<td>tc</td>
<td>-4.651</td>
<td>....</td>
</tr>
</tbody>
</table>

Critical Values: 10% = -3.487, 5% = -2.886, 1% = -2.580.

Johansen cointegration test

Null Hypothesis: there exist, at most, h cointegrating relationships
System: OP, FF

<table>
<thead>
<tr>
<th>h</th>
<th>Likelihood Ratio test</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>12.520</td>
<td>15.410</td>
<td>20.040</td>
</tr>
<tr>
<td>1</td>
<td>2.795</td>
<td>3.760</td>
<td>6.650</td>
</tr>
</tbody>
</table>

The Likelihood Ratio test reveals no cointegrating relationships.

Given the findings of the diagnostic tests, the Federal funds rate and oil price is included in first differences, and oil price volatility and technical change are included in level form. Some selected variable properties are illustrated in Table 4.3. These give some intuition as to the nature of the relationship between oil price and innovation. The fact that the interest rate fell on average in the post-1980 period, coupled with an average decline in the price of oil, suggests that the Fed's policies are to some extent dictated by oil prices.

However, declining interest rates would not portend the reported reduction in the mean level of technical change. Therefore, something other than monetary policy must be driving the decline in the mean level of displayed innovation.
Table 4.3: Selected Variable Properties

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Deviation</td>
<td>Mean</td>
<td>Std Deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Differenced Oil Price</td>
<td>0.039</td>
<td>2.072</td>
<td>0.382</td>
<td>1.382</td>
<td>-0.243</td>
</tr>
<tr>
<td>Volatility</td>
<td>1.424</td>
<td>2.172</td>
<td>0.591</td>
<td>1.208</td>
<td>2.121</td>
</tr>
<tr>
<td>Differenced Fed Funds Rate</td>
<td>0.004</td>
<td>2.421</td>
<td>0.053</td>
<td>1.394</td>
<td>-0.0361</td>
</tr>
<tr>
<td>Technical Change</td>
<td>100.2</td>
<td>1.603</td>
<td>100.232</td>
<td>2.149</td>
<td>100.164</td>
</tr>
</tbody>
</table>

A relatively low level of volatility characterizes the period prior to 1980. In fact, the mean volatility level quadruples in the period after 1980. At the same time, the mean value of innovation, or technical change, falls from the pre- to post-1980 period. Also, the standard deviation of technical change falls by more that half suggesting that, prior to 1980, the absence of sustained volatility created periods of lumpy innovation. This is perfectly in line with the uncertainty and investment literature. According to the theory, a period of high volatility will push investments into the future so that more information can be gathered about the environment in which a firm operates. If periods such as these are not continuous, rather, they are infrequent, then one could expect to see spikes in the level of innovation. From examination of Figure 4.1, it is apparent that there was not only a sustained higher level of price volatility after 1980, there is also a higher frequency of spikes in the series. This would cause a continuous series of delays in the
implementation of innovating technologies, thereby lowering the standard deviation, as well as the mean value, of the technical change series.

Table 4.4: VAR Estimation Results

<table>
<thead>
<tr>
<th></th>
<th>DOP</th>
<th>VL</th>
<th>DFF</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOP(-1)</td>
<td>0.0708</td>
<td>0.2613</td>
<td>0.2540</td>
<td>-0.0092</td>
</tr>
<tr>
<td>DOP(-2)</td>
<td>-0.1791</td>
<td>0.0903</td>
<td>0.0254</td>
<td>-0.0741</td>
</tr>
<tr>
<td>DOP(-3)</td>
<td>0.1980</td>
<td>-0.1138</td>
<td>-0.0201</td>
<td>-0.0983</td>
</tr>
<tr>
<td>DOP(-4)</td>
<td>-0.0613</td>
<td>-0.0202</td>
<td>0.2479</td>
<td>-0.0296</td>
</tr>
<tr>
<td>VL(-1)</td>
<td>0.0827</td>
<td>0.1599</td>
<td>0.1508</td>
<td>-0.0233</td>
</tr>
<tr>
<td>VL(-2)</td>
<td>-0.0832</td>
<td>0.6017</td>
<td>0.0359</td>
<td>-0.1549</td>
</tr>
<tr>
<td>VL(-3)</td>
<td>-0.0878</td>
<td>0.0529</td>
<td>-0.2640</td>
<td>-0.0700</td>
</tr>
<tr>
<td>VL(-4)</td>
<td>-0.0865</td>
<td>-0.2048</td>
<td>0.0703</td>
<td>-0.0442</td>
</tr>
<tr>
<td>DFF(-1)</td>
<td>0.0050</td>
<td>0.1694</td>
<td>-0.6383</td>
<td>-0.0184</td>
</tr>
<tr>
<td>DFF(-2)</td>
<td>-0.0784</td>
<td>0.1544</td>
<td>-0.6291</td>
<td>-0.0961</td>
</tr>
<tr>
<td>DFF(-3)</td>
<td>0.0273</td>
<td>0.0618</td>
<td>-0.3330</td>
<td>-0.0710</td>
</tr>
<tr>
<td>DFF(-4)</td>
<td>-0.0019</td>
<td>-0.0211</td>
<td>-0.0846</td>
<td>-0.0469</td>
</tr>
<tr>
<td>TC(-1)</td>
<td>-0.1226</td>
<td>0.0621</td>
<td>-0.0175</td>
<td>-0.1533</td>
</tr>
<tr>
<td>TC(-2)</td>
<td>-0.1196</td>
<td>-0.0641</td>
<td>0.2590</td>
<td>-0.1230</td>
</tr>
<tr>
<td>TC(-3)</td>
<td>-0.0595</td>
<td>0.0230</td>
<td>-0.0344</td>
<td>-0.0468</td>
</tr>
<tr>
<td>TC(-4)</td>
<td>0.0410</td>
<td>-0.0033</td>
<td>0.0665</td>
<td>0.2376</td>
</tr>
<tr>
<td>Constant</td>
<td>26.4013</td>
<td>-1.1837</td>
<td>-27.4157</td>
<td>109.1745</td>
</tr>
</tbody>
</table>

The appropriate lag lengths to include in the VAR is determined using the Akaike Information Criterion (AIC). The AIC test reveals that a common lag length of four is best for estimating the system. The estimation results are reported in Table 4.4. As expected, all of the included variables have a negative impact on technical change.
Furthermore, oil price changes have a positive effect on the Federal funds rate indicating that interest rates do respond in a fashion consistent with anti-inflationary behavior.

Table 4.5: Hypothesis Test Results

<table>
<thead>
<tr>
<th>Likelihood Ratio test</th>
<th>( \chi^2(4) )</th>
<th>Null Hypothesis: All lags (denoted XL) = 0 in equation Y.</th>
<th>Critical Values: 10% - 7.779, 5% - 9.488, 1% - 13.277.</th>
<th>DOPL</th>
<th>VLL</th>
<th>DFFL</th>
<th>TCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOP</td>
<td>5.936</td>
<td>3.722</td>
<td>1.293</td>
<td>2.335</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VL</td>
<td>14.889</td>
<td>51.740</td>
<td>5.090</td>
<td>0.937</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFF</td>
<td>20.614</td>
<td>8.370</td>
<td>53.119</td>
<td>7.250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>3.403</td>
<td>11.291</td>
<td>1.374</td>
<td>13.612</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The lagged values in each equation read along the top. The dependent variable is in the left-hand column. A LR value of 11.291 indicates that the lags of oil price volatility (VLL) are significant at the 5% level in explaining technical change (TC).

<table>
<thead>
<tr>
<th>Pair-wise Granger Causality tests</th>
<th>F(4,118)</th>
<th>Null Hypothesis:</th>
<th>F-Statistic</th>
<th>Probability that null is true</th>
</tr>
</thead>
<tbody>
<tr>
<td>VL does not Granger Cause DOP</td>
<td>0.445</td>
<td>0.776</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOP does not Granger Cause VL</td>
<td>2.886</td>
<td>0.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFF does not Granger Cause DOP</td>
<td>0.203</td>
<td>0.936</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOP does not Granger Cause DFF</td>
<td>3.131</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC does not Granger Cause DOP</td>
<td>0.201</td>
<td>0.937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOP does not Granger Cause TC</td>
<td>1.091</td>
<td>0.365</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFF does not Granger Cause VL</td>
<td>0.591</td>
<td>0.670</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VL does not Granger Cause DFF</td>
<td>1.144</td>
<td>0.340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC does not Granger Cause VL</td>
<td>0.333</td>
<td>0.855</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VL does not Granger Cause TC</td>
<td>2.518</td>
<td>0.045</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC does not Granger Cause DFF</td>
<td>1.096</td>
<td>0.362</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFF does not Granger Cause TC</td>
<td>0.567</td>
<td>0.687</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of the likelihood-ratio tests which were performed to determine the significance of particular lagged variables on the contemporaneous variables are reported in Table 4.5 along with pair-wise Granger-causality tests. The former uses the
information contained in the entire system, whereas the latter is a simple bi-variate test\textsuperscript{18}. For both tests the results are identical. We find that oil price volatility has a significant negative effect on the technical change series. In other words, oil price volatility suppresses production frontier movements. This fits nicely with the notion that increased price volatility will cause a delay in the release of new technologies. This is not to say that research and development endeavors are delayed, rather that the implementation of the fruits of such endeavors is delayed. This would have a negative effect on the expansion of productive capacity thereby producing the negative relationship between volatility and changes in output which have been reported in previous studies. We also find that oil price changes are significant in explaining changes in the Federal Funds rate, but the lack of explanatory power from the Federal Funds rate to technical change indicates that this has little effect upon the level of innovation displayed in the US economy. Hence, the monetary policy channel does not explain how oil prices effect the level of innovation in the macroeconomy.

In order to have some idea about the long run effects of the included variables on technical change, we construct impulse response functions for a ten quarter time horizon. These are reported in Figure 4.4. The system is identified by employing a Choleski factorization that places oil price changes first in the ordering. This is followed by price volatility, the first-differenced Federal Funds rate, and the technical change variable. The ordering implies that each variable is independent of contemporaneous disturbances in the following variables. Thus, oil prices are independent of contemporaneous movements in any of the other variables. This assumes, probably incorrectly, that there is no lag in

\textsuperscript{18} It is recognized that the power of the pair-wise test is considerably weaker as it ignores the relationships of the other variables, but it serves as an adequate reinforcement of the other results.
the response of monetary policy. However, changing the ordering of the variables does not effect the results reported.
Figure 4.4: Impulse Response Analysis

Impulse response of technical change to a one standard deviation shock to oil price changes

Impulse response of technical change to a one standard deviation shock to Federal Funds rate changes

Impulse response of technical change to a one standard deviation shock to volatility
The impulse response analysis indicates a 0.0045 unit reduction in technical change given a one standard-deviation shock in oil price volatility after 4 quarters. The magnitude of this response in terms of reduction in the annual growth of output to a one standard-deviation shock in volatility is approximately 0.43%\textsuperscript{19}. This value was computed by solving the CRS linear programming problems for technical change backwards. Specifically, inputs are held constant across periods and the resulting output is calculated given a 0.0045 unit reduction in the technical change index for the US. In the sense that inputs are held constant, this provides only an approximation. In reality, inputs are increasing; therefore, output may in fact not decline. Thus, we use the terminology ‘reduction in the growth of output’ in order to indicate declines in the potential growth rate. For instance, the US experienced very low annual growth rates in real GDP in the period surrounding the crisis in the Persian Gulf. In fact, annual growth rates were 1.2% and –1% in 1990 and 1991, respectively, whereas they were 3.3% and 2.8% in 1989 and 1992, respectively. This decline in GDP growth followed a spike in oil price volatility in the third quarter of 1990 that is equal to approximately 7 standard deviations. According to our analysis, this would beget a total decline in the annual growth rate of GDP of about 3% after only 4 quarters, ceteris paribus. In particular, if the realized annual growth rate is about –1%, then without the shock to volatility it would have been around 2%. Therefore, the effect of the shock to volatility produced a reduction in GDP growth that is similar to that which is predicated by our analysis. In other words, if there had been no shock to oil price volatility, growth rates in 1990 and 1991 would have been more in line with the growth rates realized in 1989 and 1992.

\textsuperscript{19} This is the sum of the reductions in technical change as indicated by the impulse response analysis for the four quarters immediately following the shock to volatility.
This, of course, assumes that no changes would have occurred in any other variables that could adversely effect the economy.

The length of the effect is also of some importance. The negative impact does not dissipate until after nearly two-and-a-half years. Thus, a high frequency of shocks to oil price volatility will cause a prolonged reduction in technical change. This will have a significant impact on both short and long run growth. Not only will growth rates in the immediate future be impacted negatively, but also the compounded effect of lower levels of technical change will produce lower long run levels of real output growth.

Furthermore, the impulse response analysis indicates that changes in the Federal funds rate have only very minor impacts on technical change. The large negative spike that occurs in the impulse response to oil price changes is at the fourth quarter, and dies out rather rapidly. Hence, the net effect of an oil price change is only minor compared to the effects of shocks to volatility. However, the response of the Federal funds rate to shocks in changes in oil prices is rather substantial. Again, this is indicative of a monetary policy response to changing conditions in the oil market, but we find no evidence of significant effects on technical change.

4.5 Concluding Remarks

Two aspects of our analysis stand out. First, we have expanded the analysis in Färe et al. (1994b), albeit with a considerably smaller sample of countries (their analysis uses 17) but with a longer time horizon, to include the input energy consumption. Not only does this reveal that the US is not alone at the production frontier, but that the exclusion of significant inputs has a substantial effect on the results one obtains.
Second, we find that oil price volatility has a significant negative impact on technical change. This is new evidence by which the uncertainty caused by the volatility of oil prices affects the macroeconomy. Both the sectoral shocks channel and the uncertainty and investment channel expound theories to explain this. If increased volatility in the price of oil does indeed raise aggregate unemployment, as is stipulated, then one could expect for productivity to decline. In addition, delays in investment caused by increased price volatility will have the effect of reducing the rate of implementation of some new technologies. The combined effect of both phenomena is a reduction in production frontier expansion, or a reduction in technical change. Furthermore, we find that oil price changes have a significant positive influence on interest rates, thereby giving evidence of counter-inflationary monetary policy. However, we find that these counter-inflationary measures do not have an effect on technical change. This sheds a negative light on the ability of the monetary policy channel to explain how oil prices effect innovation.

These results suggest, from a policy perspective, that taking measures to minimize the volatility of oil price will maximize advances in the frontier production technology, ceteris paribus. This would have the effect of increasing the growth of productivity by encouraging technical change. From the standpoint of welfare maximization, this is a desirable goal.

Oil is a feedstock into the production of the many petroleum products that are used in the US on a daily basis (such as plastics, chemicals, etc.) as well as an important means of deriving energy. Hence, the effects of oil price volatility can reach into many sectors of the economy. An interesting extension of this work would be an application to individual sectors of an economy. Disaggregation at that level might give some insight
into the sources of technical change as well as identify those most sensitive to oil price fluctuations. A study of this magnitude could be beneficial in specifying what types of policies could be enacted to minimize the effect that price volatility has on innovation.
References

Chapter 1


Chapter 2


Li, Q. (1988) *working paper* presented at Econometrics Seminar, Economics Department, Rice University.


**Chapter 3**


Center for Global Energy Studies (1993) *Oil Production Capacity in the Gulf, volume II*


**Chapter 4**


Appendix A

I. Proof of Theorem 2.2 (Efficient Score for Independence Case)

Differentiating the log-likelihood function in (2.4) leads to the score function with respect to $\beta$ and $\eta$ respectively

$$S_\beta = -S_\varepsilon(\varepsilon)x_1 \quad \text{(parametric score), and}$$

$$S_\eta = \left[ -S_\varepsilon \frac{\partial g(x_2, \eta)}{\partial \eta_1} + \frac{\partial f_\varepsilon(\varepsilon)}{\partial \eta_2} \frac{\partial f_z(x)}{\partial \eta_3} \right]$$

$$= -S_\varepsilon(\varepsilon)h_1(x_2) + h_2(\varepsilon) + h_3(x) \quad \text{(nuisance score)}$$

Therefore, the tangent set, denoted as $\mathcal{S}$, is (by Definition 6 in chapter 2)

$$\mathcal{S} = \{ t(\varepsilon): E(t(\varepsilon)) = 0, t(\varepsilon) = -S_\varepsilon(\varepsilon)k_1(x_2) + t_2(\varepsilon) + t_3(x) \}$$

where $E(t_1(x_2)) = E(t_2(\varepsilon)) = E(t_3(x)) = 0$ \(^20\) \hfill (A1)

We denote the mean square closure of all linear combinations of $S_\beta$ and $S_\eta$ as $\Sigma$, which may be viewed as a closed linear subspace in Hilbert space with inner product between two elements, say $\alpha$ and $\beta$, defined as $E(\alpha \beta)$. $\mathcal{S}$ is the mean square closure of all possible linear combinations of $S_\eta$, $\mathcal{S}$ hence can also be viewed as a closed subset of Hilbert space $\Sigma$. Let's denote $\mathcal{N}$ for the span of all linear combinations of $t_2(z)$, i.e., $\mathcal{N}$ is also a closed linear subspace of $\Sigma$. Similarly, $\Phi$ is the mean-square closure of all linear combinations of functions like $S_{\varepsilon|x_1}k_1(x_2)$.

$$\Sigma = \mathcal{N} \oplus \mathcal{N}^\perp$$ \hfill (A2)

---

\(^20\) For $E(t_2(\varepsilon)) = E(t_3(x)) = 0$, the interchange of differentiation and integral is assumed.
where $\Theta$ means that any element of $\mathcal{F}$ can be viewed as the sum of two element of $\mathcal{K}$ and $\mathcal{K}^\perp$, and $\perp$ denotes for orthogonality, i.e., $[\mathcal{K}, \mathcal{K}^\perp]$ is an orthogonal decomposition of Hilbert space $\Sigma$.

Due to (A2) and (A3), $S_\epsilon(\epsilon) x_1 \in \mathcal{K}^\perp$, which leads to the following result,

$$proj(S_\beta|\mathcal{F}) = proj(S_\beta|\Phi)$$  \hspace{1cm} (A3)

Therefore, the projection of $S_\beta(z)$ onto $\mathcal{F}$, where $S_\beta(z) = -S_\epsilon(\epsilon)x_1$, should take the form: $S_\epsilon(\epsilon) h(x_2)$. By definition of projection, it can be derived that,

$$proj(S_\beta|\mathcal{F}) = -S_\epsilon(\epsilon) E(x_1|x_2)$$  \hspace{1cm} (A4)

Therefore the efficient score function, denoted as $S(z)$, is

$$S(z) = -S_\epsilon[x_1 - E(x_1|x_2)]$$

II. Proof of Theorem 2.3 (Efficient Score for Conditional Symmetry Case)

$$L(y; x; \beta, \eta) = \sum_{i=1}^{T} \left[ \log f_{\rho|x}(y - g(x_2) - x_1 \beta; \eta) + \log f_{x}(x; \eta) \right]$$  \hspace{1cm} (A5)

$$\Rightarrow \quad S_\beta = -S_{\epsilon|x}(\epsilon)x_1, \text{ (parametric score)}$$

and

$$S_\eta = \left[ -S_{\epsilon|x} \frac{\partial g(X_2, \eta)}{\partial \eta} + \frac{\partial f_{\rho|x}(\epsilon)/\partial \eta}{f_{\rho|x}} + \frac{\partial f_{x}(X)/\partial \eta}{f_{x}} \right], \text{ (nuisance score)}$$

$$= -S_{\epsilon|x}(\epsilon, X)k_1(X_2) + t_2(\epsilon, X) + t_3(X)$$
Therefore, a plausible conjecture for the tangent set, based on its definition, the assumptions, and $S_{\eta}$, is

$$\mathcal{S} = \{ S_{\epsilon|x}t_1(X_2) + t_2(\epsilon, X) + t_3(X) : S_{\epsilon|x} \text{ is an odd function of } \epsilon, \ t_2(\epsilon, X) = t_2(-\epsilon, X), \text{ and } \ E(t_2(\epsilon, X)) = E(t(X)) = 0 \}$$

The residual, $\epsilon$, now is also i.i.d. distributed but could have conditional heteroscedasticity (conditional upon the regressors). Although this parametric submodel appears similar to the one in section 4, the tangent set associated with it does have two important features. First, given the conditional symmetry (A2'), $S_{\epsilon|x}t_1(X_2) \perp t_2(\epsilon, X)$. Therefore, the tangent set, $\mathcal{S}$, can be orthogonal decomposed as: $\mathcal{S} = \mathcal{N} \oplus \Phi$. $\mathcal{N}$ is the mean-square closure of all linear combinations of functions like $t_2(z)$, and $\Phi$ stands for the set of elements that are orthogonal to $\mathcal{N}$ in $\mathcal{S}$, i.e., $\Phi \subset \mathcal{N}^\perp$, where $\mathcal{N}^\perp$ is the space orthogonal to $\mathcal{N}$ in $\Sigma$. In this case $\Phi$ is the mean-square closure of all linear combinations of functions like $S_{\epsilon|x}t_1(X_2)$. Secondly, $S_\beta \in \mathcal{N}^\perp$, that is $S_\beta \perp t_2(z)$. The above two properties are due to the fact that $S_{\epsilon|x}$ and $S_\beta$ are odd functions in $\epsilon$ and $t_2(z)$ is an even function in $\epsilon$. Therefore, the following equation is true,

$$\text{proj}(S_\beta | \mathcal{S}) = \text{proj}(S_\beta | \Phi) \tag{A6}$$

and hence $\text{proj}(S_\beta | \Phi)$ should in general take the form: $-S_{\epsilon|x}(\epsilon)q(x_2)$. Then by definition, we have,

$$\text{proj}(S_\beta | \Phi) = -S_{\epsilon|x} \frac{E(S_{\epsilon|x}^2 | x_2)}{E(S_{\epsilon|x}^2 | x_2)} \tag{A7}$$
III. Proof of Theorem 2.4 (Efficient Score for Conditional Zero Mean Case)

Similar to what we did in the previous proofs, the score functions with respect to $\beta$ and $\eta$ are,

\[ S_\beta = -S_{\varepsilon|x}(\varepsilon)x_i \] (parametric score)

and

\[
S_\eta = \left[ -S_{\varepsilon|x} \frac{\partial g(x_2, \eta)}{\partial \eta_1} \frac{\partial f_x(\varepsilon)/\partial \eta_2}{f_x} + \frac{\partial f_x(\varepsilon)/\partial \eta_3}{f_x} \right] \\
= -S_{\varepsilon|x}(\varepsilon)x_i(x_2) + t_2(\varepsilon) + t_3(x) \] (nuisance score)

We also denote the mean square closure of all linear combinations of $S_\beta$ and $S_\eta$ as $\Sigma$.

Differentiating condition (A2") with respect to $\beta$ leads to

\[ -x_i - E(\varepsilon S_{\varepsilon|x}(\varepsilon)x_i|x) = 0 \quad \Rightarrow \]

\[ E(\varepsilon S_{\varepsilon|x}(\varepsilon)x) = -1 \] \hspace{1cm} (A8)

In addition, differentiating (A2") with respect $\eta_2$ (note: $\eta_2$ can be viewed as an arbitrary distribution function of $\varepsilon$) yields,

\[ E(\varepsilon S_{\varepsilon|x}(\varepsilon)x) = 0 \] \hspace{1cm} (A9)

Therefore, the tangent set is

\[ \mathcal{S} = \{ t(z) : E(t(z)) = 0, t(z) = -S_{\varepsilon}(\varepsilon)x_i(x_2) + t_2(z) + t_3(x) \} \]

where $E(\varepsilon S_{\varepsilon|x}|x) = -1, E(\varepsilon t_2(z)|x) = 0, E(t_2(z)) = E(t_3(x)) = 0$ \}

Since $\mathcal{S}$ is mean-square closure of all possible linear combinations of $S_\eta$, $\mathcal{S}$ can also be viewed as a closed subset of Hilbert space $\Sigma$. Let's denote $\mathcal{R}$ for the mean
square closure of all linear combinations of \( t_2(z) \), i.e., \( \mathcal{K} \) is also a closed subspace of Hilbert space \( \Sigma \). Then we know the following is true,

\[
\Sigma = \mathcal{K} \oplus \mathcal{K}^\perp
\]  

(A10)

where \( \oplus \) means that any element of \( \mathcal{S} \) can be viewed as the sum of two element of \( \mathcal{K} \) and \( \mathcal{K}^\perp \), and \( \perp \) denotes for orthogonality, i.e., \([\mathcal{K}, \mathcal{K}^\perp]\) is an orthogonal decomposition of Hilbert space \( \Sigma \).

Due to (A9), for any arbitrary function of \( z \) in \( \mathcal{S} \), that is \( \forall \, R(z) \in \Sigma \), \( R(z) - \text{proj}(R(z)|\mathcal{K}) \) should in general take a general functional form: \( q(x)e \), i.e., \( q(x)e \not\in \mathcal{K} \) or \( q(x)e \in \mathcal{K}^\perp \). It is not difficult to show that \( q(x)e \) can be expressed as a linear combination of the above three elements of \( \mathcal{S} \), i.e., \( q(x)e \in \mathcal{S} \), and given (A10), we know \( \text{proj}(R(z)|\mathcal{K}^\perp) \) should also take the form: \( q(x)e \). By definition of mean-square projection, it can be derived that,

\[
q(x) = \frac{E(R(z)e|x)}{E[e^2|x]} 
\]  

(A11)

\[
\Rightarrow \text{proj}(R(z)|\mathcal{K}^\perp) = R(z) - \text{proj}(R(z)|\mathcal{K}) = \frac{E(R(z)e|x)}{E[e^2|x]} e 
\]  

(A12)

\[
\Rightarrow \text{proj}(R(z)|\mathcal{K}) = R(z) - \frac{E(R(z)e|x)}{E[e^2|x]} e 
\]  

(A13)

Apply (A9) and (A12) to \(-S_{elx}(e)\xi(x_z) \), which is one part of the tangent set,

\[
\text{proj}(-S_{elx}(e)\xi(x_z)|\mathcal{K}^\perp) = -S_{elx}(e)\xi(x_z) - \text{proj}(-S_{elx}(e)\xi(x_z)|\mathcal{K}) = \frac{t_i(x_z)}{\sigma^2(x)} e 
\]  

(A14)

Then the tangent set can be rewritten as the summation of two orthogonal subsets, i.e., in the following orthogonal decomposition form,
\[ \mathcal{S} = \{ t(z) : E(t(z)) = 0, t(z) = \frac{t_1(x_2)}{\sigma^2(x)} + t_2(\varepsilon, x) \} \]

where \( E(t_2(z)|x) = 0, E(t_2(z)) = 0 \) \( \text{21} \)

Now the condition: \( \left( \frac{t_1(x_2)}{\sigma^2(x)} \varepsilon \right) \perp t_2(z) \) holds and the linear space spanned by functions like \( \frac{t_1(x_2)}{\sigma^2(x)} \varepsilon \) is denoted as \( \Phi \). It is then not difficult to see \( \mathbb{R} \perp \Phi \) (i.e., \( \Phi \subset \mathbb{R}^\perp \)) and

\[
proj(S_\beta | \mathbb{R}) = proj(S_\beta | \mathbb{R}) + proj(S_\beta - proj(S_\beta | \mathbb{R}) | \Phi) \quad \text{(A16)}
\]

Applying (A12) and (A13) to \( S_\beta \),

\[
proj(S_\beta | \mathbb{R}) = S_\beta - \left[ S_\beta - proj(S_\beta | \mathbb{R}) \right] = S_\beta - proj(S_\beta | \mathbb{R}^\perp)
\]

\[ = S_\beta - \frac{x_1 \varepsilon}{\sigma^2(x)} = -S_{\text{est}}(\varepsilon)x_1 - \frac{x_1 \varepsilon}{\sigma^2(x)} \quad \text{(A17)} \]

The next thing is to calculate \( proj(S_\beta - proj(S_\beta | \mathbb{R}) | \Phi) = proj\left( \frac{x_1 \varepsilon}{\sigma^2(x)} | \Phi \right) \), which in general will have the form \( \frac{h(x_2) \varepsilon}{\sigma^2(x)} \), (i.e., \( \Phi \) is the mean-square closure of linear combinations of functions like \( \frac{t_1(x_2)}{\sigma^2(x)} \varepsilon \). Therefore, simply by definition of mean-square projection, we have

\[
proj(S_\beta - proj(S_\beta | \mathbb{R}) | \Phi) = \frac{E\left( \frac{x_1}{\sigma^2(x)} | x_2 \right) \varepsilon}{E\left( \frac{1}{\sigma^2(x)} | x_2 \right) \sigma^2(x)} \quad \text{(A18)}
\]

\[ \text{21} \quad \therefore t_2(\varepsilon, x) \text{ and } t_2(x) \text{ can be any continuous functions that satisfies the conditions stated in the above definition of } \mathcal{S}, \text{ } t_2(\varepsilon, x) \text{ in general includes } t_2(x). \]
\[ \Rightarrow \text{proj}(S_\beta | \mathcal{S}) = -S_{\varepsilon|x} x_1 - \frac{x_1 \varepsilon}{\sigma^2(x)} + \frac{E \left( \frac{x_1}{\sigma^2(x)} | x_2 \right) \varepsilon}{E \left( \frac{1}{\sigma^2(x)} | x_2 \right) \sigma^2(x)} \] 

\[ \Rightarrow S = S_\beta - \text{proj}(S_\beta | \mathcal{S}) = \frac{\varepsilon}{\sigma^2(x)} \left( x_1 - \frac{E \left( \frac{x_1}{\sigma^2(x)} | x_2 \right)}{E \left( \frac{1}{\sigma^2(x)} | x_2 \right) \sigma^2(x)} \right) \] 

IV. Proof of Theorem 2.5 (Efficient Score for Partial Uncorrelatedness Case)

Given that

\[ S_\beta = -S_{\varepsilon|x} (\varepsilon) x_1 \] (parametric score)

and

\[ S_\eta = \left[ -S_{\varepsilon|x} \frac{\partial g(x_2, \eta)}{\partial \eta_1} + \frac{\partial f_\varepsilon(\varepsilon) / \partial \eta_2}{f_\varepsilon} + \frac{\partial f_x(x) / \partial \eta_3}{f_x} \right] \]

\[ = -S_{\varepsilon|x} (\varepsilon) x_1 (x_2) + t_2(\varepsilon) + t_3(x) \] (nuisance score)

We also denote the mean square closure of all linear combinations of \( S_\beta \) and \( S_\eta \) as \( \Sigma \).

Differentiating condition (1) in Assumption (A2'') with respect to \( \beta, \eta_1 \), and \( \eta_2 \) leads correspondingly to

\[ E \left( x_1 \left( 1 + \varepsilon S_{\varepsilon|x} (\varepsilon) \right) x_1 \right) = 0 \] (A21)

\[ E \left( t_1(x_2) (x_1 + x_1 \varepsilon S_{\varepsilon|x} (\varepsilon)) \right) = 0 \] (A22)

and \[ E \left( x_1 \varepsilon t_2(z) \right) = 0 \] (A23)
Similarly, differentiating condition (2) in Assumption (A2‴′) with the same set of parameters leads to

\[
E\left( x_1 \left( 1 + \varepsilon S_{elx} (\varepsilon) \right) x_2 \right) = 0 \tag{A24}
\]

\[
E\left( t_1 (x_2) \left( 1 + \varepsilon S_{elx} (\varepsilon) \right) x_2 \right) = 0 \tag{A25}
\]

\[
E\left( \varepsilon x_2 (x_2) \right) = 0 \tag{A26}
\]

Therefore, given the nuisance score and conditions (A21) – (A26), the tangent set can be defined as the following,

\[
\mathcal{S} = \{ t(z) : E(t(z)) = 0, t(z) = -S_{elx} t_1 (x_2) + t_2 (z), \}
\]

where \( S_{elx} \), \( t_1 (x_2) \), and \( t_2 (z) \) satisfy (A21) – (A26) \( \}
\]

Given (A23) and (A26), it appears that functions contained in \( \Sigma \) are orthogonal to \( \mathcal{N} \) [the mean square closure of \( t_2 (z) \)] should take the form: \( \alpha (x_1 + \lambda h (x_2)) \varepsilon \), where \( \alpha \) and \( \lambda \) are scalars. That is the elements of \( \mathcal{N}^\perp \) should have this form. The objective here is, for any arbitrary function \( R(z) \) contained in \( \Sigma \), to find its projection onto \( \mathcal{S} \), i.e. the general form of functions contained in \( \mathcal{N}^\perp \cap \Phi^\perp \), where \( \Phi \) stands for the mean-square closure of \( S_{elx} t_1 (x_2) \). It can be easily verified that such function should in general assume the form:

\[\alpha (x_1 - E(x_1 | x_2)) \varepsilon \] under (A22) or (A25). Hence, for \( \forall R(z) \in \Sigma \), \( \text{proj}(R(z)|\mathcal{S}^\perp) \) should take the form of \( \alpha (x_1 - E(x_1 | x_2)) \varepsilon \).

\[
\Rightarrow \quad \text{proj}(R(z)|\mathcal{S}) = R(z) - \text{proj}(R(z)|\mathcal{S}^\perp)
\]

\[= R(z) - E(R(z) \bar{x}_i \varepsilon)(E(\varepsilon^{2} \bar{x}_i \bar{x}_i'))^{-1} \bar{x}_i \varepsilon \tag{A28}
\]

where \( \bar{x}_i = x_i - E(x_i | x_2) \).
\[ \Rightarrow \quad \text{proj}(S_\beta | S) = S_\beta - E(S_\beta \tilde{x}_i \epsilon)(E(\epsilon^2 \tilde{x}_i \tilde{x}_i'))^{-1} \tilde{x}_i \epsilon \]

\[ = S_\beta - E(\tilde{x}_i \tilde{x}_i')(E(\epsilon^2 \tilde{x}_i \tilde{x}_i'))^{-1} \tilde{x}_i \epsilon \quad \text{(due to (A21) and (A24))} \]

\[ \Rightarrow \quad \text{Then the efficient score } S(z) = E(\tilde{x}_i \tilde{x}_i')(E(\epsilon^2 \tilde{x}_i \tilde{x}_i'))^{-1} \tilde{x}_i \epsilon. \]

V. Proof of Theorem 2.6 (Consistency)

Under (C1) – (C3), Theorem 2.6 follows directly from Lemma 5.2 in Newey (1994a).

VI. Proof of Theorem 2.7 (Normality)

Given \( \hat{\beta} \xrightarrow{p} \beta_0 \) and (N1) – (N4), Theorem 2.7 follows from Lemma 5.3 in Newey (1994a).

VII. Proof of Theorem 3.1

We first need to show that our estimated dynamic programming model (3.11) satisfies the following two assumptions.

(A1) \( \Sigma \) is a convex subset of \( S^2 \), the correspondence \( \Phi: \Sigma \rightarrow \Pi \) is nonempty, compacted-valued, and continuous.

(A2) The function \( \hat{B}: \Pi \times \Sigma \rightarrow S^1 \) is bounded and continuous, and \( 0 < \beta < 1 \).

The first part of (A1) is not restrictive for our simulations\(^ {22} \). The second part requires the maximum number of new oil wells that can be drilled within one period to be bounded, which is reasonable because in reality the number of new wells that can be drilled during

\(^ {22} \) It is assumed in the numerical optimization that the number of wells can be any nonnegative real numbers. However, after the optimal solution is obtained in each step of the simulation, the number of new wells is rounded up in approximating the optimal new wells drilled.
a period has to be bounded by economic and engineering constraints\textsuperscript{23}. Then given that \( \hat{f}(\pi, \sigma) \), defined by (3.9), is a bounded and continuously differentiable function, it is not difficult to verify that the short-term capacity function is nonempty, compact valued, and continuous by the Implicit Function Theorem, and hence so is the correspondence \( \Phi: \Sigma \rightarrow \Pi \), given \( \Sigma \) is compact. The continuity of function \( \hat{B}(\pi, \sigma) \) follows directly from the linear regression results used in the estimated dynamic programming model (3.10). With (A1) and the assumption that \( \Sigma \) is compact (bounded and closed), \( \hat{B}(\pi, \sigma) \) is also bounded. Therefore, (A2) holds in general.

Next we demonstrate that in the model (3.11) can be modified so that the state variables in the next period are the choice variables. In particular, there exists a correspondence \( \hat{\Phi}: \Sigma \rightarrow \Sigma \) that also satisfies assumption (A1). For any elements \( \sigma_0 \in \Sigma \), there is a set in \( \Pi \) defined by correspondence \( \Phi \) and denoted as \( \Lambda \), which is nonempty and compact due to assumption (A1). We can further define a set in \( \Sigma \) denoted as \( \Delta \) (i.e., \( \Delta \in \Sigma \)) such that

\[
\Delta = \{ \sigma \in \Sigma: \sigma \text{ can be expressed as a summation of } \sigma_0 \text{ and any element of } \Pi \}
\]

Given \( \Delta \), choosing \( \pi (\pi \in \Pi) \) is equivalent to choosing an element in \( \Delta \) as the state variable of the next period and hence a correspondence \( \Phi: \Sigma \rightarrow \Sigma \) can be set up such that it assigns to each element of \( \Sigma \) an image set similar to \( \Lambda \) defined for \( \sigma_0 \).

Correspondence \( \Phi \) satisfies (A1) simply due to the way it is constructed. Then by Theorem 4.6 in Stokey, Lucas, and Prescott (1989), Theorem 4.1 holds for model (3.12),

\textsuperscript{23} In all the simulations, we assumed an upper bound for \( dN \) that is equal to forty and the simulation results did not get even close to that number. Empirically, the feasible set for the number of new oil wells is bounded and closed.
and hence (3.11). The basic idea of the theorem in Stokey, Lucas, and Prescott is to show under the assumption (A1) and (A2), $T$ satisfies the hypothesis of Blackwell's sufficient conditions for a contraction mapping.

VIII Proof of Corollary 3.1

The proof requires us to verify assumptions (A1) and (A2), and the following two new assumptions.

(A3) $\hat{B}(\cdot, \cdot)$ as defined in (4.2) is strictly increasing in each of its state variables of the current period (i.e., $\sigma_t$).

(A4) $\Phi$ is monotone in each of the state variables in the sense that $\sigma \leq \sigma'$ implies $\Phi(\sigma) \leq \Phi(\sigma')$.

(A4) is easy to check for (3.11) (or (3.10)). The idea is that an increase in the number of oil producing wells or a decrease in the cumulative production, raises short-term capacity. Hence the image set of $\Gamma(\tau)$ enlarges. This is consistent with the estimated short-term capacity function in (3.9), as a function of $N$ and $CP^-$. (A3) is relatively more complicated to check because it requires us to use the next period state variables ($\sigma_{t+1}$) as the choice variable as modeled in (3.12). However, if we substitute the state transition equations (3.10) into the corresponding Bellman equation to eliminate the choice variables ($X_t, dN_t$) as required by (3.12), it is not difficult to demonstrate that the partial derivatives of the profit function $\hat{B}$ with respect to $N_t$ and $CP_\tau$ are in general positive.

Therefore, with (A1), (A2), (A3), and (A4), Corollary 4.1 follows directly from our Theorem 4.1 and Theorem 4.7 in Stokey, Lucas, and Prescott (1989).
Figure A1
Base/Case 02: J-Slice 1/Gas Saturation(FRACTION)/10957.0 Days

Figure A2

Gas

Water or Oil

0. 0.16 0.32 0.48
Base/Case 03: J-Slice 1/Water Saturation (FRACTION) / 10957.0 Days

Figure A3

Water and/or Irreducible Oil

Oil
Figure A4: Geometry of a reservoir wedge (as seen from above)
Figure A5: Geometry of the stitched wedges for entire reservoir (as seen from above)
Typical Well Production Schedule
Rates at 5% of Reserves Per Year

Figure A6

Oil Production Rate
Thousands

Date

Oct-43  Oct-54  Sep-65  Aug-76

- Well 1  - Well 2  - Well 3  - Well 4  - Well 5
Typical Well Production Schedule

Rates at 6% of Reserves Per Year

Figure A7

Oil Production Rate

Thousands

Oct-43  Oct-54  Sep-65  Aug-76

Date

Well 1  Well 2  Well 3  Well 4  Well 5
Typical Well Production Schedule
Rates at 8% of Reserves Per Year

Figure A9

Oil Production Rate
Thousands

Date
Oct-43 Oct-54 Sep-65 Aug-76

Well 1 — Well 2 — Well 3 — Well 4 — Well 5
Typical Well Production Schedule
Rates at 9% of Reserves Per Year

Figure A10

Oil Production Rate
Thousands

- Well 1 - Well 2 - Well 3 - Well 4 - Well 5

Date
Oct-43 Oct-54 Sep-65 Aug-76