

ESSAYS IN NONLINEAR MACROECONOMIC
MODELING AND ECONOMETRICS

by

BEBONCHU ATEMS

B.A., University of Maryland at College Park, 2007

M.A., Kansas State University, 2009

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Economics
College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2011

Abstract

This dissertation consists of three essays in nonlinear macroeconomic modeling and econometrics. In the first essay, we decompose oil price movements into oil demand (stock market) shocks and oil supply (oil-market) shocks, and examine the response of the stock market to these shocks. We find that when oil prices are “net-increasing”, a stock market shock that causes the S&P 500 to rise by one percentage point will cause the price of oil to rise approximately 0.2 percentage points, with a statistically significant positive effect one day after the stock market shock. On the other hand, the response of the stock market to an oil market shock is a decline of 6.8% when the price of oil doubles. For other days, the initial response of the oil market to a stock market shock is the same as in the net oil price increase case (by construction). We then analyze the response of monetary policy to the identified stock market and oil market shocks and find that short-term interest rates respond to the stock market shocks but not the oil market shocks. Finally, we evaluate the predictive power of the decomposed stock market and oil shocks relative to the change in the price of oil. We find statistically significant gains in both the in-sample fit and out-of-sample forecast accuracy when using the identified stock market and oil market shocks rather than the change in the price of oil.

The second essay revisits the statistical specification of near-multicollinearity in the logistic regression model using the Probabilistic Reduction approach. We argue that the *ceteris paribus* clause invoked with near-multicollinearity is rather misleading. This assumption states that one can assess the impact of near-multicollinearity by holding the parameters of the logistic regression model constant, while examining the impact on their standard errors and *t* – ratios as the correlation (ρ) between the regressors increases. Using the Probabilistic Reduction approach, we derive the parameters (and related statistics) of the logistic regression model and show that they are functions of ρ ,

indicating the *ceteris paribus* clause in the traditional account of near multicollinearity is unattainable. Monte carlo simulations in the paper confirm these findings. We also show that traditional near-multicollinearity diagnostics, such as the variance inflation factor and condition number can fail to detect near-multicollinearity. Overall, the paper finds that near-multicollinearity in the logistic model is highly variable and may not lead to the problems indicated by the traditional account. Therefore, unexpected, unreliable or unstable estimates and inferences should not be blamed on near-multicollinearity. Rather the modeler should return to economic theory or statistical respecification of their model to address these problems.

The third essay examines the correlations between income inequality and economic growth using a panel of income distribution data for 3,109 counties of the U.S. We examine the non-spatial dynamic correlations between county inequality and growth using a System GMM approach, and find significant negative relationships between changes in inequality in one period and growth in the subsequent period. We show that this finding is robust across different sample sizes. We further argue that because the space-specific time-invariant variables that affect economic growth and inequality can differ significantly across counties, failure to incorporate spatial effects into a model of growth and inequality may lead to biased results. We assume that dependence among counties only arises from the disturbance process, hence the estimation of a spatial error model. Our results indicate that the bias in the parameter for inequality amounts to about 2.66 percent, while that for initial income amounts to about 21.51 percent.

ESSAYS IN NONLINEAR MACROECONOMIC
MODELING AND ECONOMETRICS

by

BEBONCHU ATEMS

B.A., University of Maryland at College Park, 2007

M.A., Kansas State University, 2009

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Economics
College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

2011

Approved by:

Major Professor
Lance Bachmeier

Copyright

Bebonchu Atems

2011

Abstract

This dissertation consists of three essays in nonlinear macroeconomic modeling and econometrics. In the first essay, we decompose oil price movements into oil demand (stock market) shocks and oil supply (oil-market) shocks, and examine the response of the stock market to these shocks. We find that when oil prices are “net-increasing”, a stock market shock that causes the S&P 500 to rise by one percentage point will cause the price of oil to rise approximately 0.2 percentage points, with a statistically significant positive effect one day after the stock market shock. On the other hand, the response of the stock market to an oil market shock is a decline of 6.8% when the price of oil doubles. For other days, the initial response of the oil market to a stock market shock is the same as in the net oil price increase case (by construction). We then analyze the response of monetary policy to the identified stock market and oil market shocks and find that short-term interest rates respond to the stock market shocks but not the oil market shocks. Finally, we evaluate the predictive power of the decomposed stock market and oil shocks relative to the change in the price of oil. We find statistically significant gains in both the in-sample fit and out-of-sample forecast accuracy when using the identified stock market and oil market shocks rather than the change in the price of oil.

The second essay revisits the statistical specification of near-multicollinearity in the logistic regression model using the Probabilistic Reduction approach. We argue that the *ceteris paribus* clause invoked with near-multicollinearity is rather misleading. This assumption states that one can assess the impact of near-multicollinearity by holding the parameters of the logistic regression model constant, while examining the impact on their standard errors and *t* – ratios as the correlation (ρ) between the regressors increases. Using the Probabilistic Reduction approach, we derive the parameters (and related statistics) of the logistic regression model and show that they are functions of ρ ,

indicating the *ceteris paribus* clause in the traditional account of near multicollinearity is unattainable. Monte carlo simulations in the paper confirm these findings. We also show that traditional near-multicollinearity diagnostics, such as the variance inflation factor and condition number can fail to detect near-multicollinearity. Overall, the paper finds that near-multicollinearity in the logistic model is highly variable and may not lead to the problems indicated by the traditional account. Therefore, unexpected, unreliable or unstable estimates and inferences should not be blamed on near-multicollinearity. Rather the modeler should return to economic theory or statistical respecification of their model to address these problems.

The third essay examines the correlations between income inequality and economic growth using a panel of income distribution data for 3,109 counties of the U.S. We examine the non-spatial dynamic correlations between county inequality and growth using a System GMM approach, and find significant negative relationships between changes in inequality in one period and growth in the subsequent period. We show that this finding is robust across different sample sizes. We further argue that because the space-specific time-invariant variables that affect economic growth and inequality can differ significantly across counties, failure to incorporate spatial effects into a model of growth and inequality may lead to biased results. We assume that dependence among counties only arises from the disturbance process, hence the estimation of a spatial error model. Our results indicate that the bias in the parameter for inequality amounts to about 2.66 percent, while that for initial income amounts to about 21.51 percent.

Contents

List of Figures	xii
List of Tables	xv
Acknowledgements	xvi
Dedication	xviii
1 Asymmetric Oil Price Shocks and the Response of the Stock Market to Oil Shocks	1
1.1 Introduction	1
1.2 Empirical Analysis of the Effects of Oil Price Shocks on the Stock Market	4
1.2.1 Data	4
1.2.2 Methodology	5
1.3 Empirical Results	11
1.3.1 Coefficient Estimates	11
1.3.2 Historical Decomposition of the Stock Market and Oil Market Shocks	15
1.4 Macroeconomic Analysis	17
1.4.1 Monetary Policy	17
1.4.2 Forecasts of Macroeconomic Variables	20

1.4.2.1	In-Sample Analysis	21
1.4.2.2	Out-of-Sample Analysis	22
1.5	Conclusion	32
2	Revisiting the Statistical Specification of Near-Multicollinearity in the	
	Logistic Regression Model	34
2.1	Introduction	34
2.2	Multicollinearity in the Linear Regression Model	37
2.2.1	Multicollinearity in the Linear Regression Model: The Traditional	
	Story	37
2.2.1.1	Effects of Near-Multicollinearity	38
2.2.1.2	Detecting Near-Multicollinearity	39
2.2.2	Revisiting the Problem of Near-multicollinearity in the Linear	
	Regression Model	41
2.3	Multicollinearity in the Logistic Regression Model	43
2.3.1	The Logistic Regression Model	43
2.3.2	Multicollinearity in the Logistic Regression Model: The Tradi-	
	tional Story	45
2.4	The Probabilistic Reduction Approach and the Logistic Regression Model	
	47	
2.4.1	A Re-examination of Near-Multicollinearity in Logistic Regres-	
	sion using the PR Approach	48
2.4.2	The Logistic Regression Model with Two Correlated Normal Co-	
	variates with Homogeneous Covariance Matrix	51
2.4.3	The Variance and Standard Error of β	53
2.5	Simulation and Graphics	55
2.5.1	Simplifications	55

2.5.1.1	Admissable Parameter Values	55
2.5.1.2	<i>Simplification 1</i>	56
2.5.1.3	<i>Simplification 2</i>	56
2.5.2	Simulation Design	57
2.5.3	Graphics	59
2.5.3.1	The β_0 surface with varying ρ and different mean vectors of $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$	59
2.5.3.2	The β_1 Surface with varying ρ and different mean pairs (μ_{11}, μ_{21})	62
2.5.3.3	The β_2 Surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$	65
2.5.3.4	The standard error of β_0 surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$	69
2.5.3.5	The standard error of the β_1 Surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$	71
2.5.3.6	The standard error of the β_2 surface with varying ρ and different mean pairs (μ_{1i}, μ_{2i})	72
2.5.3.7	The $\tau(\beta_0)$ surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$	74
2.5.3.8	The $\tau(\beta_1)$ surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$	77
2.5.3.9	The $\tau(\beta_2)$ surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$	79
2.5.3.10	The R^2 , <i>Condition Number</i> , and <i>VIF</i> surfaces with varying ρ and different Mean Pairs (μ_{1i}, μ_{2i})	81
2.5.3.11	Marginal Effects	85

2.6	Extensions	88
2.6.1	The Logistic Regression Model with Continuous Normal Covariates, Different Mean Vectors and a Heterogeneous Variance-Covariance Structure	88
2.6.2	The Logistic Regression Model with Binary Covariates	91
2.6.3	The Multivariate Normal k -regressor Case with Homogeneous Covariance	96
2.7	Conclusion	98
3	The Spatial Dynamics of Growth and Income Inequality: Empirical Evidence Using U.S. County Level Data.	101
3.1	INTRODUCTION	101
3.2	Background and Literature Review	104
3.2.1	Trends in U.S. County-Level Growth and Inequality	104
3.2.2	Literature Review	110
3.3	Data and Methodology	114
3.4	Estimation and Results	118
3.4.1	The Non-Spatial Dynamic Model	118
3.4.2	Robustness of the Non-Spatial Model	121
3.5	The Role of Space in the Panel Growth Model	123
3.5.1	The Spatial Weights Matrix	124
3.5.2	The Dynamic Spatial Panel Growth Model	126
3.5.3	The Likelihood Function.	127
3.6	Conclusion	132
	Bibliography	145

List of Figures

- 1.1 Response of the S&P 500 Return and WTI Crude Oil Prices to Each Structural Shock: $\delta_s = 0$ 12
- 1.2 Impulse Response Functions of the S&P 500 Return and WTI Crude Oil Prices to Each Structural Shock: $\delta_s = 1$ 14
- 1.3 Evolution of Stock and Oil Market Shocks 15
- 1.4 One-Step Ahead Forecasts and Forecast Errors - Core CPI Inflation . . 26
- 1.5 One-Step Ahead Forecasts and Forecast Errors - CPI Inflation 26
- 1.6 One-Step Ahead Forecasts and Forecast Errors - Federal Funds Rate: 1998:7-2011:2 27
- 1.7 One-Step Ahead Forecasts and Forecast Errors - Change in Federal Funds Rate: 1998:7-2011:2 27
- 1.8 One-Step Ahead Forecasts and Forecast Errors - Industrial Production: 1998:7-2011:2 28
- 1.9 One-Step Ahead Forecasts and Forecast Errors - M2 Money Supply: 1998:7-2011:2 28
- 1.10 One-Step Ahead Forecasts and Forecast Errors - PPI Inflation (All Commodities): 1998:7-2011:2 29
- 1.11 One-Step Ahead Forecasts and Forecast Errors - Core PPI: 1998:7-2011:2 29

1.12	One-Step Ahead Forecasts and Forecast Errors - 3-Month T-Bill Rate:	
	1998:7-2011:2	30
1.13	One-Step Ahead Forecasts and Forecast Errors - Unemployment Rate:	
	1998:7-2011:2	30
2.1	The β_0 surface with varying ρ , and varying $\bar{\mu}_{2,1}(\bar{\mu}_{1,1} = 2)$	60
2.2	β_0 with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	61
2.3	β_0 with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	62
2.4	The β_1 Surface with varying ρ and varying $\mu_{2,1}(\mu_{1,1} = 2)$, $p = 0.6$	63
2.5	β_1 with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	64
2.6	β_1 with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	65
2.7	The β_2 surface with varying ρ and varying $\bar{\mu}_{2,1}(\bar{\mu}_{1,1} = 2)$	66
2.8	β_2 with Varying ρ and different mean pairs (μ_{1i}, μ_{2i}) : $p = 0.6$	67
2.9	β_2 varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	68
2.10	The $se(\beta_0)$ with varying ρ and different pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	69
2.11	The $se(\beta_0)$ with varying ρ and different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	70
2.12	The $se(\beta_1)$ with varying ρ for different (μ_{1i}, μ_{2i}) : $p = 0.6$	71
2.13	The $se(\beta_1)$ surface with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	72
2.14	The $se(\beta_2)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	73
2.15	The $se(\beta_2)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	74
2.16	The $\tau(\beta_0)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	75
2.17	The $\tau(\beta_0)$ with varying ρ and different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	76
2.18	The Surface of the β_1 <i>t</i> -statistics With Varying ρ for Different (μ_{1i}, μ_{2i}) :	
	$p = 0.6$	77
2.19	The $\tau(\beta_1)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	78
2.20	The $\tau(\beta_2)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	79
2.21	$\tau(\beta_2)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	80

2.22	The McFadden pseudo R^2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	81
2.23	The McFadden pseudo R^2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	81
2.24	The <i>Condition Number</i> with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	83
2.25	The <i>Condition Number</i> with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	83
2.26	The Variance Inflation Factors (<i>VIF</i>) of x_1 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	84
2.27	The Variance Inflation Factor (<i>VIF</i>) of x_1 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	84
2.28	The Variance Inflation Factor (<i>VIF</i>) of x_2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$	85
2.29	The Variance Inflation Factor (<i>VIF</i>) of x_2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	85
2.30	A: The Marginal Effects of x_1 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $P = 0.6$	86
2.31	A: The Marginal Effects of x_2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$	87
3.1	Spatial Variation in U.S. County-Level Per Capita Income Growth: 1970 – 1980	105
3.2	Spatial Variation in U.S. County-Level Per Capita Income Growth: 2000 – 2007	107
3.3	Spatial Variation in U.S. County-Level Income Inequality: 1970	108
3.4	Spatial Variation in U.S. County-Level Income Inequality : 2000	109

List of Tables

1.1	Reduced Form VAR Estimates	11
1.2	Response of Interest Rates to Oil and Stock Market Shocks	19
1.3	Response of Interest Rates to Oil Price Change	20
1.4	Variables, Adjustments and Transformations	21
1.5	In-Sample <i>J</i> -Test Results: Our Model Vs Change in Oil Price Model . .	24
1.6	One-step-ahead Forecasts	25
1.7	Multi-Step Ahead Forecasts	31
3.1	Summary Statistics and Data Sources	117
3.2	Estimation Results for the Non-spatial Model	120
3.3	Sensitivity Analysis	122
3.4	Estimation Results for the Spatial Dynamic Model	131

Acknowledgements

I would like to express my profound gratitude to my advisor, Dr. Lance Bachmeier for his help, support and guidance throughout my graduate school career. I am thankful for his subtle, yet unforgiving insistence on quality and accuracy in research. He helped me become a better student, and developed my passion for research far beyond this project. He is my economic role model and for that, I owe him a debt of appreciation. I want to thank Dr. Lloyd Thomas who believed in my ability to teach from the first day I gave a practice lecture. Even when I questioned my readiness and ability to teach, his belief in me was steadfast. I am also grateful to him for being a great mentor. Thank you to Dr. Cassou for being a great teacher, for developing my love for macroeconomics, and for making helpful comments throughout this dissertation process. Dr. Bergtold worked closely with me in the econometrics area, particularly on issues relating to multicollinearity. I learned a lot from him. I also want to thank him for being a great friend. Dr. Weixing Song was always willing to take time off of his busy schedule to help me with “R”, and for that, I am truly grateful indeed. To the rest of the economics faculty, particularly Dr. Babcock, Dr. Al-Hamdi, and Dr. Kuester, you made the department fun for me.

I especially want to thank my dad, George Atems, and my mom, Agnes Atems whose thoughts, prayers, good wishes and unshakable love kept me going every day. Thank you to Tatangha, Ekokobe, Nkolaka and Zinkeng Atems for testing me, challenging me, making me laugh, and bringing out the best in me. To my Aunt, Grace Anyike and her husband, I cannot express to you how grateful I am for your kindness. You all are the most awesome family I know and I love you all.

My immense gratitude goes to my colleagues in the economics department with whom I have shared thoughts and ideas. First with Vladimir Bejan, and later with Patrick Scott, we formed the most effective tag team partnership I could ever imagine.

To Abhinav who stressed on a philosophy of “together, we win”, yes we did! Dr. Ojede, Eddery Lam, Chris and Laura Youderian, Mercy Palamuleni, Crystal Strauss, Jenny Wu, Dan Lensing, Rashmi Dhankar, and others I have not mentioned have all been part of this incredible journey.

Throughout this journey, I have been accompanied by an extraordinary band of brothers who stood and fought for me when necessary. To Edwin Sama and Dr. Orlandrew Danzell for being big brother figures to me in Manhattan. Céline Apercé was always willing to lend a listening ear when I vented during times of frustration. To all my team mates of my championship-winning soccer teams - Global Athletic FC and Too Big to Beat FC - for letting me captain and lead the team for three long years, thank you!

Dedication

To my parents, George and Agnes Atems for their endless love and support.

Chapter 1

Asymmetric Oil Price Shocks and the Response of the Stock Market to Oil Shocks

1.1 Introduction

Despite the constant development of alternative sources of energy, oil still accounts for the largest fraction of primary energy worldwide. Data from the BP Statistical Review of World Energy (2009), shows that oil accounted for 35 percent of global energy use in 2008, much higher than coal (29 percent), natural gas (24 percent), nuclear (6 percent) and hydropower (5 percent). As a consequence of this heavy dependence on oil, and based on past experience with oil shocks, a large body of research has attempted to estimate the effect of an oil price shock on macroeconomic variables including GDP (Hamilton 1983, 2003), inflation (Blanchard and Gali, 2009, and Bachmeier and Cha, 2011), and monetary policy (Bernanke, Gertler and Watson, 1997).¹ A key paper

¹See Hamilton (2003) and Kilian (2008) for reviews of the literature.

by Hamilton (1983) documented a clear negative relationship between oil prices and output. In that paper, Hamilton provided strong evidence that oil price changes could be treated as exogenous to the U.S. economy. A variety of tests supported the assertion that oil price increases are followed by decreases in output. Other studies, including Rotemberg and Woodford (1996), Warnock and Warnock (2000), and Sensier and Van Dijk (2004) have shown that this result extends to other macroeconomic variables such as aggregate unemployment, wages and prices.

To the extent that oil price changes affect the macroeconomy, and thus firm cash flows, they should also affect stock prices. Papers by Jones and Kaul (1996), Ewing and Thompson (2007), Bachmeier (2008), and Miller and Ratti (2009) all confirm this view. Driesprong, Jacobsen, and Maat (2008) provide evidence that stock markets overreact to oil price changes, in the sense that there is a reversal of stock price changes in the days after an oil price shock.

Kilian and Park (2009) offer an interesting alternative view of the response of stock prices to a change in the price of oil. They argue that a drawback of existing studies is that by treating oil price shocks as exogenous to the economy and hence stock returns, reverse causality becomes a problem in regressions that relate oil price changes to stock returns. Kilian (2009) showed that GDP responds differently to an oil price change depending on the nature of the underlying shock. If the price of oil rises because of a shock to oil supplies, GDP is expected to fall. If on the other hand the price of oil rises because of a positive shock to world output, and thus the demand for oil, GDP is expected to rise. Building on these findings, Kilian and Park (2009) argue that a similar story can be told with respect to the response of the stock market following a change in the price of oil. Oil price increases that are due to higher oil demand should cause stock prices to rise. Higher oil prices should only be expected to cause stock prices to fall if the higher oil price is due to a shock to the supply of oil. Using an identified

vector autoregressive (VAR) approach, they show that oil supply shocks are bad for the Center for Research in Security Prices (CRSP) value-weighted market portfolio, but oil price increases due to increases in world output cause the portfolio to appreciate.

This chapter addresses the same issue as Kilian and Park (2009) but uses a different dataset and a different approach to decomposing oil price movements into shocks to oil supply and oil demand. Our dataset consists of daily data on U.S. oil and stock prices. An advantage of our identification strategy is that it does not require the construction of an index of world output, as in Kilian and Park (2009). Additionally, the identified shocks can be used in a real time forecasting environment, as all of the data are available in real time.

Of course, our empirical methodology has limitations. One potential criticism is that the baseline estimates come from a bivariate VAR model of stock prices and oil prices. While bivariate VAR models have frequently been used in the existing literature (Kilian and Vigfusson, 2010; Hamilton, 2010; Bachmeier and Cha, 2011; Blanchard and Gali, 2009), it is possible that there might be more than one macroeconomic shock simultaneously affecting stock prices. Kilian (2009), which provides the basis for the work done in Kilian and Park (2009), estimated a three-variable VAR model with three shocks, and labeled one of the shocks a “precautionary” oil demand shock. The use of daily data limits the series that can be included in the VAR model. We do not feel that this limitation is too severe for the purposes of this chapter because our primary focus is on the response of the stock market to an oil price shock, and our methodology provides for a clean, valid identification of shocks to the price of oil, regardless of whether the oil price shock could be broken down further into additional types of oil shocks.

A second potential criticism is that our identification relies on Hamilton’s net oil price increase model. While we believe there is strong support for the net oil price increase model in the literature (Hamilton 1996, 2003, 2010), we are aware that some

researchers have recently questioned its use (Kilian and Vigfusson, 2010), which if correct would invalidate our identification restrictions. We make no attempt to refute the findings of Kilian and Vigfusson, as that has been done by Hamilton (2010), but we will note that though they fail to reject the null hypothesis of symmetric impulse response functions for real GDP, but for unemployment, they conclude that, “the response estimates look fairly asymmetric” at a one-year time horizon, and even state, “We certainly would not want to rule out the existence of asymmetries in all possible applications on the basis of our empirical evidence.” Thus, while we recognize that there are valid questions about model specification with respect to real GDP, we do not believe there is any evidence that would justify dismissing the net oil price increase model completely.

1.2 Empirical Analysis of the Effects of Oil Price Shocks on the Stock Market

1.2.1 Data

Our dataset consists of daily data on the percentage change in the Standard and Poor’s (S&P) 500 return and the percentage change in the price of the West Texas Intermediate (WTI) crude oil in Cushing, for the period January 2, 1986 to February 28, 2011. The data for the S&P 500 were downloaded from Yahoo! Finance, while the WTI spot crude oil prices came from the U.S. Energy Information Administration.

Following the convention in the literature, we pretested the variables for stationarity, carrying out an augmented Dickey-Fuller (ADF) test for both variables. The conclusion for both variables is that they are nonstationary in levels but stationary in first differences.² Our conclusion that both variables are nonstationary in their levels

²The ADF test statistics were -1.2 in levels and -61.3 in differences for the S&P 500, and -0.7 in levels and -59.2 in differences for WTI, against a 5% critical value of -2.86.

opens up the possibility that they are cointegrated. A Johansen test for cointegration indicated that the levels of the S&P 500 and the price of oil are not cointegrated.³ Based on these results, we proceed using a VAR model in differences. A VAR model in differences is consistent with all of the papers cited in this chapter.

1.2.2 Methodology

Denote the percentage change in the S&P 500 return on day t by s_t , and the percentage change in the WTI on day t by w_t . Then a first-order reduced-form VAR model of stock prices and oil prices⁴ can be written:

$$\begin{aligned} s_t &= \alpha_0 + \alpha_1 s_{t-1} + \alpha_2 w_{t-1} + e_{st} \\ w_t &= \beta_0 + \beta_1 s_{t-1} + \beta_2 w_{t-1} + e_{wt} \end{aligned} \tag{1.1}$$

This reduced-form VAR model is useful for forecasting purposes, but not for calculating impulse response functions. Assume that there are two structural shocks, one a “stock market” shock, denoted ε_{mt} , and the other an “oil market” shock, denoted ε_{ot} . The stock market shock captures shocks to the macroeconomy as well as other types of noise beyond the macroeconomy. This shock, representing world aggregate demand as in Kilian (2009), should have a positive effect on both stock returns (by raising expected firm cash flows) and the price of oil (by raising the demand for oil). The oil market shock, which represents current and expected future oil supply disruptions, should have a negative effect on stock returns (by lowering expected firm cash flows) but a positive effect on the price of oil (a greater supply of oil will cause the price of oil to fall).

We can then rewrite the reduced form VAR residuals in (1.1) to reflect the fact that

³The maximum eigenvalue test statistic for no cointegrating vectors versus one cointegrating vector is 6.6 against a 95% critical value of 15.67. The trace test statistic for no cointegrating vectors versus one cointegrating vector is 10.6 against a 95% critical value of 19.96.

⁴The Schwarz information criteria selected one as the optimal lag length.

they are a function of the underlying structural shocks:

$$\begin{aligned} s_t &= \alpha_0 + \alpha_1 s_{t-1} + \alpha_2 w_{t-1} + a\varepsilon_{mt} + b\varepsilon_{ot} \\ w_t &= \beta_0 + \beta_1 s_{t-1} + \beta_2 w_{t-1} + c\varepsilon_{mt} + d\varepsilon_{ot} \end{aligned} \tag{1.2}$$

Looking at the system (1.2), it is clear that in the absence of further restrictions, it is not clear what is meant by an “oil price shock”.⁵ The response of the stock market following a change in the price of oil will depend on the underlying source of the oil price movement.

The identification problem is thus to identify the coefficients a , b , c and d . The coefficient a is the response of the S&P 500 return to a “one-unit” stock market shock, b is the response of the S&P 500 return to a “one-unit” oil shock, c is the response of the oil market to a “one-unit” stock market shock, and d is the response of the oil market to a “one-unit” oil shock. We normalize the shocks by imposing $a = d = 1$. It is standard to normalize VAR shocks in this way (see e.g. Blanchard and Quah (1989)). This leaves two parameters to identify:

$$\begin{aligned} s_t &= \alpha_0 + \alpha_1 s_{t-1} + \alpha_2 w_{t-1} + \varepsilon_{mt} + b\varepsilon_{ot} \\ w_t &= \beta_0 + \beta_1 s_{t-1} + \beta_2 w_{t-1} + c\varepsilon_{mt} + \varepsilon_{ot} \end{aligned} \tag{1.3}$$

Hamilton (1996) introduced the “net oil price increase” model. He argued that changes in the price of oil should only affect the economy if the price of oil is high relative to recent experience. His logic was as follows:

If one wants a measure of how unsettling an increase in the price of oil is likely to be for the spending decisions of consumers and firms, it seems more appropriate to compare the current price of oil with where it has been over the previous year rather than during the previous quarter alone. (Hamilton

⁵See Kilian (2009) for further discussion of this point.

1996, p. 216)

Hamilton (2003) found that the data provide more support for comparison against the prior three years rather than just the prior year, as in Hamilton (1996). Additional evidence in favor of the asymmetric specification was provided in Hamilton (2010). The focus of this line of research has largely been comparing the net oil price increase model against a linear model in the context of predicting GDP growth.

Define the variable

$$rise_s = x_s - \tilde{x}_s$$

where x_s is the natural logarithm of the average price of WTI in month s and \tilde{x}_s is the log of the highest price of oil in the previous 36 months (i.e., $\tilde{x}_s = \max(x_{s-1}, x_{s-2}, \dots, x_{s-36})$).

Then define the dummy variable

$$\delta_s = \begin{cases} 1 & \text{if } rise_s > 0 \\ 0 & \text{otherwise} \end{cases}$$

to capture months that saw net oil price increases. The variable δ_s identifies months for which oil shocks are expected to affect the economy.

We extend the net oil price increase model to the system (1.3).⁶ Oil market shocks ε_o realized during net oil price increase months will cause a change in the price of oil, which in turn affects the economy and stock prices. Oil market shocks in other months would still cause the price of oil to change, but as they would not affect the economy, there would be no response of stock prices ($b = 0$). On the other hand, a stock market shock ε_{mt} should affect both the demand for oil and the price of oil regardless of whether

⁶Hamilton (1996, 2003, 2010) did not explicitly model feedback from the economy to the price of oil.

day t corresponds to a net oil price increase.

The estimation strategy is as follows. We estimate the reduced-form VAR for observations with $\delta_s = 0$ and calculate the reduced form residuals:

$$s_t = \alpha_0 + \alpha_1 s_{t-1} + \alpha_2 w_{t-1} + e_{st} \quad (1.4)$$

$$w_t = \beta_0 + \beta_1 s_{t-1} + \beta_2 w_{t-1} + e_{wt}. \quad (1.5)$$

Replacing the reduced form residuals and imposing $b = 0$,

$$\begin{aligned} s_t &= \alpha_0 + \alpha_1 s_{t-1} + \alpha_2 w_{t-1} + \varepsilon_{mt} \\ w_t &= \beta_0 + \beta_1 s_{t-1} + \beta_2 w_{t-1} + c\varepsilon_{mt} + \varepsilon_{ot}. \end{aligned}$$

We can then follow standard practice in the structural VAR literature (see e.g. Hamilton, 1994) and use the relationship between the reduced form residual covariance matrix and the structural shock covariance matrix to derive a system of three equations in three unknowns:

$$var(e_{st}) = var(\varepsilon_{mt}) \quad (1.6)$$

$$var(e_{wt}) = c^2 var(\varepsilon_{mt}) + var(\varepsilon_{ot}) \quad (1.7)$$

$$cov(e_{st}, e_{wt}) = cov(\varepsilon_{mt}) \quad (1.8)$$

The reduced form VAR estimates provide a consistent estimate of $var(e_{st})$, $var(e_{wt})$, and $cov(e_{st}, e_{wt})$, which can be used to solve for $var(\varepsilon_{mt})$, $var(\varepsilon_{ot})$, and c .

We then re-estimate the reduced-form VAR model for the $\delta_s = 1$ observations. The

structural model for this regime is

$$s_t = \alpha_0 + \alpha_1 s_{t-1} + \alpha_2 w_{t-1} + \varepsilon_{mt} + b\varepsilon_{ot}$$

$$w_t = \beta_0 + \beta_1 s_{t-1} + \beta_2 w_{t-1} + c\varepsilon_{mt} + \varepsilon_{ot}$$

We have the equations

$$\text{var}(e_{st}) = \text{var}(\varepsilon_{mt}) + b\text{var}(\varepsilon_{ot}) \quad (1.9)$$

$$\text{var}(e_{wt}) = c^2\text{var}(\varepsilon_{mt}) + \text{var}(\varepsilon_{ot}) \quad (1.10)$$

$$\text{cov}(e_{st}, e_{wt}) = c\text{var}(\varepsilon_{mt}) + b\text{var}(\varepsilon_{ot}) \quad (1.11)$$

These are three equations in four unknowns, $\text{var}(\varepsilon_{st})$, $\text{var}(\varepsilon_{ot})$, c and b . However, as we have a consistent estimate of c from above, there are only three unknowns. We use the `nleqslv` nonlinear equation solver in R to find b . Using \hat{b} and \hat{c} , we can compute the structural shocks using the following equations:

$$e_{st} = \varepsilon_{mt} + \hat{b}\varepsilon_{ot}$$

$$e_{wt} = \hat{c}\varepsilon_{mt} + \varepsilon_{ot}$$

So we now have

$$\begin{bmatrix} e_{st} \\ e_{wt} \end{bmatrix} = \begin{bmatrix} 1 & \hat{b} \\ \hat{c} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{mt} \\ \varepsilon_{ot} \end{bmatrix}$$

or

$$A^{-1} \begin{bmatrix} e_{st} \\ e_{wt} \end{bmatrix} = \begin{bmatrix} \varepsilon_{mt} \\ \varepsilon_{ot} \end{bmatrix} \text{ where } A = \begin{bmatrix} 1 & \hat{b} \\ \hat{c} & 1 \end{bmatrix}$$

To construct the confidence intervals for the above parameters, we use a bootstrap procedure. The bootstrap is a modified version of the method outlined in Hamilton (1994) and Benkwitz, Lutkepohl, and Neumann (2000). The algorithm is as follows:

1. Resample with replacement 6344 observations on the identified structural shocks. Call these series $\tilde{\varepsilon}_m$ and $\tilde{\varepsilon}_o$.
2. Set the pre-sample values \tilde{s}_0 and \tilde{w}_0 equal to the sample means of s and w , respectively. The pre-sample values of the level of the S&P 500 and price of WTI are equal to their values from January 1983 to December 1985. Set $i = 1$.
3. Generate observations for month i using the estimated model for the NOPI regime as the data generating process⁷:

$$\begin{aligned} \tilde{s}_t &= \hat{\alpha}_0 + \hat{\alpha}_1 \tilde{s}_{t-1} + \hat{\alpha}_2 \tilde{w}_{t-1} + \tilde{\varepsilon}_{mt} + \hat{b} \tilde{\varepsilon}_{ot} \\ \tilde{w}_t &= \hat{\beta}_0 + \hat{\beta}_1 \tilde{s}_{t-1} + \hat{\beta}_2 \tilde{w}_{t-1} + \hat{c} \tilde{\varepsilon}_{mt} + \tilde{\varepsilon}_{ot} \end{aligned}$$

Calculate the mean WTI value for month i . If that value is larger than the maximum of WTI over the previous 36 months, move to step 4. Otherwise, discard the data generated for month i and replace it with data generated imposing $b = 0$:

$$\begin{aligned} \tilde{s}_t &= \hat{\alpha}_0 + \hat{\alpha}_1 \tilde{s}_{t-1} + \hat{\alpha}_2 \tilde{w}_{t-1} + \tilde{\varepsilon}_{mt} \\ \tilde{w}_t &= \hat{\beta}_0 + \hat{\beta}_1 \tilde{s}_{t-1} + \hat{\beta}_2 \tilde{w}_{t-1} + \hat{c} \tilde{\varepsilon}_{mt} + \tilde{\varepsilon}_{ot} \end{aligned}$$

4. Increment i . Repeat step 3 if $i < 37$. The final bootstrapped dataset will have

⁷The data generating process was very time consuming using R, so we generated the data in C++ using the R package Rcpp (Eddelbeuttel and Francois, 2011).

6344 total observations.

5. Estimate all parameters of the model using the bootstrapped dataset.
6. Repeat steps (1) through (5) 5000 times to get 5000 estimates of the parameters of the VAR model. The standard errors of the parameters are the standard deviations of the bootstrap estimates of the parameters.

1.3 Empirical Results

1.3.1 Coefficient Estimates

This section presents the results of estimating the reduced-form VAR model (1.4) and (1.8) as described in section 2.2. Table 1.1 presents these results.

Table 1.1: Reduced Form VAR Estimates

	$\delta_s = 0$		$\delta_s = 1$	
	S&P 500	WTI	S&P 500	WTI
Intercept	0.0004 (2.17)	-0.0001 (-0.39)	0.0000 (0.06)	0.0024 (2.52)
s_{t-1}	-0.0427 (-3.16)	0.0316 (1.10)	-0.0278 (-0.81)	0.1845 (2.00)
w_{t-1}	-0.0081 (-1.27)	-0.0100 (-0.74)	-0.0381 (-2.96)	-0.0309 (-0.90)

Notes:

t – statistics are reported in parentheses below the estimated parameters. Bold-faced coefficients are significant at the 5% level

Using equations (1.6) and (1.8), we find that $var(\varepsilon_{mt}) = var(e_{st}) = 0.000145$, while $cvar(\varepsilon_{mt}) = cov(e_{st}, e_{wt}) = 0.000027$, implying $\hat{c} = \frac{cov(e_{st}, e_{wt})}{var(e_{st})} = 0.182646$. Solving for \hat{b} using equations (1.9) to (1.11), we get $\hat{b} = -0.068236$. Given these coefficient estimates, we can write the structural VAR model as:

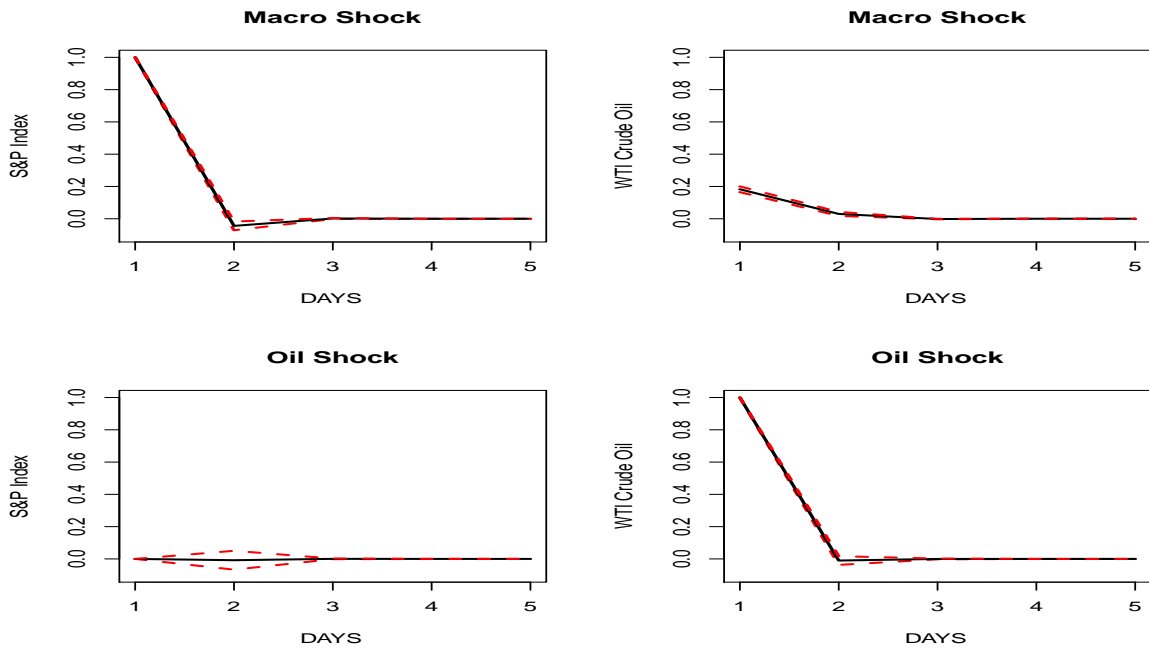
$$\begin{aligned}
s_t &= 0.00036 - 0.04274s_{t-1} - 0.00808w_{t-1} + \varepsilon_{mt} + b\varepsilon_{ot} \\
w_t &= -0.00014 + 0.03163s_{t-1} - 0.01005w_{t-1} + 0.18265\varepsilon_{mt} + \varepsilon_{ot}
\end{aligned}$$

where $\hat{b} = -0.068$ during net oil price increase months and zero otherwise.

One rarely attempts to interpret the coefficient estimates of a reduced form VAR model (Hamilton, 1994). Figures 1.1 and 1.2 are the impulse response functions for each variable following the two shocks, under the two regimes.

When we are not in a net oil price increase month, a stock market shock that causes the S&P 500 to rise by one percentage point will cause the price of oil to rise approximately 0.2 percentage points. While it is difficult to directly compare our estimates with Kilian and Park (2009), as they used monthly data and we are using daily data, it is interesting

Figure 1.1: Response of the S&P 500 Return and WTI Crude Oil Prices to Each Structural Shock: $\delta_s = 0$



to note that the immediate response of the price of oil⁸ is positive and significantly different from zero, and then falls to zero thereafter. This is consistent with the price of oil being an asset price that fully reflects new information immediately. Kilian and Park found a slight (but statistically insignificant) positive response in the month of a shock to aggregate demand, with the impulse response function rising for twelve months, and being significant at a 5% level nine months after the shock. They did not report impulse response functions beyond a twelve month horizon, so it is not known how persistent are the effects of aggregate demand shocks on the price of oil.

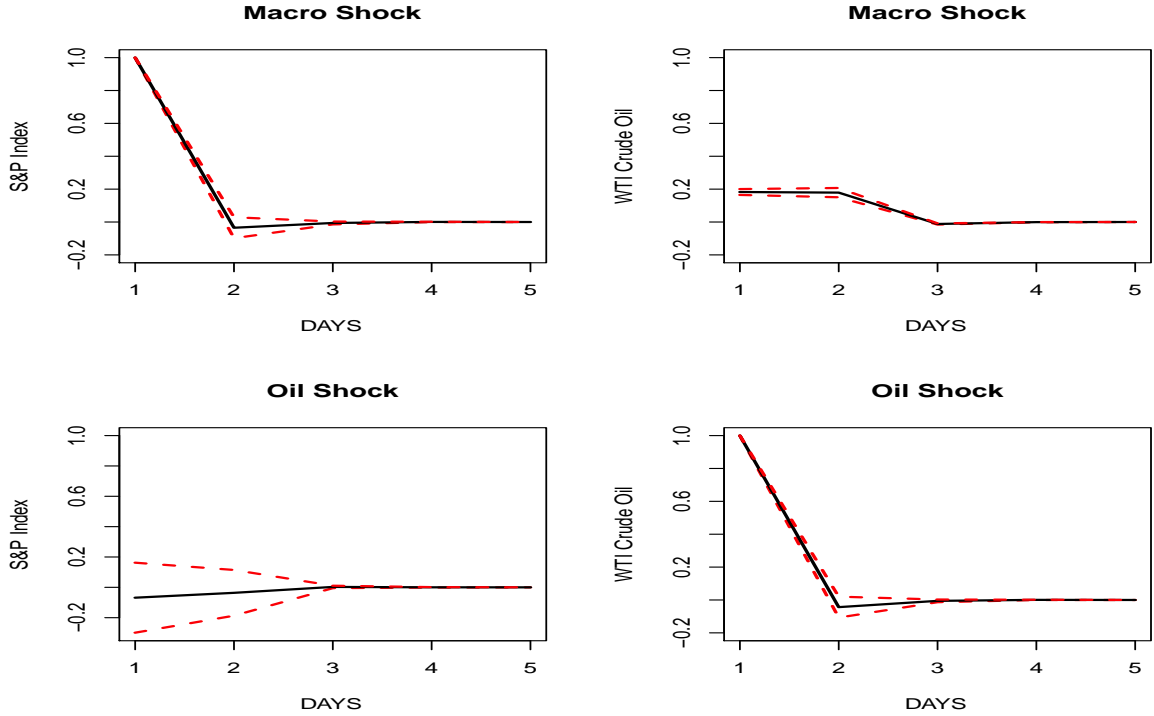
By construction, the S&P 500 does not respond to the oil market shock contemporaneously, so there is nothing to say about the contemporaneous response of the S&P 500 to an oil market shock. There are no lagged responses of either variable to either shock. This result is also in contrast to Kilian and Park (2009), who found that aggregate demand shocks and oil market shocks both have a delayed but prolonged effect on stock returns.

When we are in a net oil price increase month, the initial response of the oil price to a stock market shock is the same (this is by construction). However, we also find a statistically significant positive effect on the day after the stock market shock. This suggests that when the price of oil is hitting new highs, traders initially underestimate the effect of a stock market shock on the demand for oil. The response of the stock market to an oil market shock is a decline of 6.8% when the price of oil doubles⁹.

⁸Kilian and Park studied the real price of oil, but at a daily frequency the distinction between real and nominal oil price changes is likely to be trivial, though we have no way to test that proposition given the lack of daily inflation data.

⁹Note that this requires the price of oil to double after it is already at a higher price than at any point in the previous three years. Also note that the same logic does not apply to oil price decreases.

Figure 1.2: Impulse Response Functions of the S&P 500 Return and WTI Crude Oil Prices to Each Structural Shock: $\delta_s = 1$



How do these results compare with a model that treats the change in the price of oil as an exogenous shock to the oil market? A regression of the stock market return on its own lag, the contemporaneous change in the price of oil, and four lags of the price of oil yields the following results (t - statistics in parentheses):

$$s_t = 0.00 - 0.04 s_{t-1} + 0.03 w_t - 0.01 w_{t-1} - 0.01 w_{t-2} + 0.01 w_{t-3} + 0.00 w_{t-4}$$

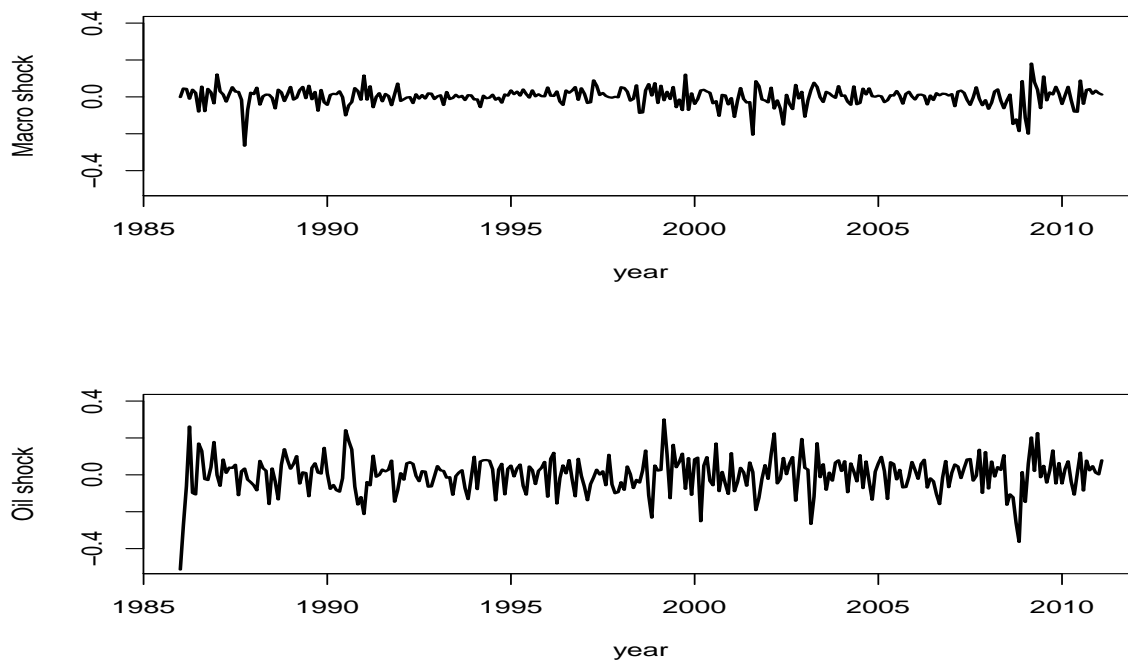
(1.99)
(-3.05)
(5.21)
(-2.06)
(-1.28)
(1.36)
(0.16)

If one does not allow for the possibility that the price of oil responds to changes in output, it will lead to a conclusion that adverse oil shocks have a *positive* effect on the stock market.

1.3.2 Historical Decomposition of the Stock Market and Oil Market Shocks

Figure 1.3 plots the evolution of the identified stock market and oil market shocks. To improve readability, the shocks are aggregated by summing up the shocks in each month. One way to judge the identification is to ask whether specific identified oil shocks reflect information about the oil market. At the beginning of the sample, there was a large negative oil market shock as the OPEC cartel collapsed. There was a large positive oil market shock in August 1990 at the time of the first Gulf War. The identified shocks suggest that the fall of the price of oil in 2008 was partly due to a negative stock market shock, but also partly to a negative oil shock, consistent with the claims in Hamilton (2009) that the price of oil exceeded its “fundamental” value when it rose above \$140 in June 2008.

Figure 1.3: Evolution of Stock and Oil Market Shocks



To assess the contribution of each of these shocks, we regress the stock return and oil price change on the identified shocks separately to see which is more important for each variable. That is, we perform the following regressions:

$$s_t = \alpha_1 + \alpha_2 \varepsilon_{mt} + e_t \quad (1.12)$$

$$s_t = \beta_1 + \beta_2 \varepsilon_{ot} + e_t \quad (1.13)$$

$$w_t = \gamma_1 + \gamma_2 \varepsilon_{mt} + e_t \quad (1.14)$$

$$w_t = \delta_1 + \delta_2 \varepsilon_{mt} + e_t \quad (1.15)$$

Because our interest is only to examine the proportion of the observed variation of these variables that is explained by these shocks, we only report the R^2 for each model. We find (from model 1.12) that about 97.8% of the variation in stock returns is associated with the stock market shock, while only about 0.03% of the variability in stock prices (from equation 1.13) is associated with shocks that drive the oil market. This finding is hardly surprising given that there are days with little or no changes in oil prices. The stock market shock is shown to explain only 0.7% of the variation in oil prices, with a huge proportion of the variation in oil prices (98.03%) explained by the oil-market specific shock, ε_{mt} (equation 1.15).

1.4 Macroeconomic Analysis

The previous section has shown that oil market shocks have a large and statistically significant effect on the S&P 500 when those shocks occur in months where the price of oil is higher than at any time in the previous three years. This section looks at how the identified stock market and oil market shocks interact with important macroeconomic variables. There are two motivations for this exercise. First, our identification restrictions are more credible if the identified shocks affect macroeconomic variables in ways that are consistent with economic theory. Second, our identified stock market and oil market shocks might provide new predictors for macroeconomic variables.

1.4.1 Monetary Policy

It has been argued that monetary policy plays a critical role in the transmission of oil to the economy. Some authors have even gone so far as to conclude that without a response of monetary policy, oil shocks would have little effect on the macroeconomy. Consider the following quote from Bernanke, Gertler and Watson (1997):

Substantively, our results suggest that an important part of the effect of oil price shocks on the economy results not from the change in oil prices, *per se*, but from the resulting tightening in monetary policy (p. 39).

Bernanke, Gertler and Watson argued that oil shocks in the 1970's and 1980's would not have been followed by recessions if the Federal Reserve had not tightened monetary policy to prevent inflation following an increase in the price of oil.

This section examines the response of monetary policy to the identified stock market and oil price shocks. In particular, we are interested in whether the Federal Reserve responds differently to the stock market and oil market shocks. Kilian and Lewis (2011) find no evidence that the Federal Reserve has responded to oil shocks in the past.

Blinder and Reis (2005) even claim that ignoring oil shocks was an “innovation” of the Greenspan Federal Reserve. Nonetheless, a regression of a monetary policy variable on the change in the price of oil is unlikely to provide an informative answer to this question. The price of oil responds to stock market shocks, and it is not controversial to suggest that the Federal Reserve will respond to the stock market shock.

In this section we use the change in interest rates rather than the levels. Stock and Watson (1999), Hangsden (2004), Crowder (2006), and Bec and Basil (2009) have found evidence of a unit root in the federal funds rate. Campbell and Shiller (1987) provided seminal research leading to a literature on testing the term structure of interest rates using cointegration tests, which is built on the premise that interest rates are nonstationary. Shikida and Figueiredo (2010), using $P-ADF$ tests and $M-estimators$ that properly deal with nonlinearities, demonstrate that testing for stationarity of interest rates properly requires a great deal of caution. They concluded that the federal funds rate is nonstationary. Consequently, we use the first difference of interest rates in the analysis that follows.

We analyze the response of market interest rates to oil and stock market shocks using daily data on the change in the federal funds rate, as well as the 30-day, 60-day, and 180-day eurodollar rates from January 1986 to February 28, 2011. The data are collected from the Federal Reserve Statistical Release H.15 of the Federal Reserve Board. As demonstrated by Cochrane and Piazzesi (2002), the 30-day eurodollar rate on the day before an FOMC meeting is an excellent predictor of federal funds rate changes, and viewing the 30-day eurodollar rate as an average of the expected federal funds rate over the next 30 days is a good approximation. We also consider the 60-day and 180-day eurodollar rates to capture delayed Federal Reserve responses to oil shocks.

Denoting the average value of the interest rate series on day t as R_t , we do an OLS regression of each of the interest rates on the identified stock market and oil shocks:

$$R_t = \alpha + \beta R_{t-1} + \lambda \varepsilon_{mt} + \varphi \varepsilon_{ot} + e_t \quad (1.16)$$

Table 1.3 summarizes the results.

Table 1.2: Response of Interest Rates to Oil and Stock Market Shocks

		30-Day	60-Day	180-Day
	<u>FFR</u>	<u>Eurodollar</u>	<u>Eurodollar</u>	<u>Eurodollar</u>
α	-0.0016 (-0.48)	-0.0013 (-1.18)	-0.0011 (-1.38)	-0.0011 (-1.38)
β	-0.1902 (-15.3)	-0.0133 (-1.06)	-0.0875 (-7.00)	-0.0438 (-3.48)
λ	-0.6574 (-2.37)	-0.2709 (2.91)	-0.1004 (-1.56)	-0.1596 (-2.51)
φ	0.1991 (1.57)	0.0636 (1.50)	0.0458 (1.56)	0.0373 (1.29)
R^2	0.037	0.002	0.009	0.003

Notes:

t - statistics are reported in parentheses below the estimated parameters. Bold-faced coefficients are significant at the 5% level

The coefficient on the oil shock is positive but statistically insignificant in all four regressions. This is consistent with the findings of Kilian and Lewis (2011) that the Federal Reserve does not react to oil shocks. Something that is interesting is that in three of the four regressions, the coefficient on the identified stock market shock is negative and statistically significant. This suggests that the stock market shock is capturing news about monetary policy (i.e., a lower than expected federal funds rate target is a positive stock market shock) and news about productivity (which pushes down expectations of inflation and implies a more expansionary path of monetary policy in the future). The clear conclusion from these results is that short-term interest rates do not respond to oil shocks, but that it would be very misleading to simply regress the interest rate series on the change in the price of oil. Those results are presented in

Table 1.3, where the regression equation is

$$R_t = \alpha + \beta R_{t-1} + \lambda w_t + e_t$$

Table 1.3: Response of Interest Rates to Oil Price Change

	FFR	30-Day Eurodollar	60-Day Eurodollar	180-Day Eurodollar
α	0.0016 (0.48)	-0.0012 (1.09)	-0.0011 (-1.38)	-0.0011 (-1.38)
β	-0.1897 (15.3)	-0.0153 (-1.21)	0.1013 (8.10)	0.0463 (3.67)
λ	0.1504 (1.20)	-0.0156 (-0.17)	0.0595 (2.06)	0.0362 (1.26)
R^2	0.036	0.000	0.011	0.002

Notes:

t – statistics are reported in parentheses below the estimated parameters. Bold-faced coefficients are significant at the 5% level

1.4.2 Forecasts of Macroeconomic Variables

Having examined the (nonlinear) effects of stock market and oil price shocks on the stock market, as well as the response of monetary policy to these shocks, we turn to the question of whether the information from our decomposed stock market and oil shocks improve the forecasts of macroeconomic variables. We examine both the in-sample fit and the historical out-of-sample forecast performance of our model relative to a benchmark model that includes only the change in the price of oil as a predictor.

The data used for the in-sample predictability tests are monthly data covering the period February 1986 to February 2011. We consider a range of macroeconomic series measuring inflation, output, interest rates, and money growth. The data on oil prices were downloaded from the website of the Energy Information Administration. All other variables were downloaded from the FRED database at the Federal Reserve Bank of

Saint Louis. The variables used are listed in Table 1.5.

Table 1.4: Variables, Adjustments and Transformations

Variable	Seasonal Adjustment	Transformation
Core CPI	Seasonally Adjusted	log-difference
CPI (All Goods)	Seasonally Adjusted	log-difference
Federal Funds Rates	Not Applicable	Level
Federal Funds Rate (Δ)	Not Applicable	First Difference
Industrial Production	Seasonally Adjusted	log-difference
M2 Money Stock	Seasonally Adjusted	log-difference
PPI (All Commodities)	Seasonally Adjusted	log-difference
Core PPI	Seasonally Adjusted	log-difference
3-month T-bill Rate	Not Applicable	Level
Unemployment Rate	Seasonally Adjusted	Level

1.4.2.1 In-Sample Analysis

Denote by y_t , the macroeconomic variable we want to predict, w_t , the percentage change in WTI oil price, ε_{mt} , the structural stock market shock, and ε_{ot} , the structural oil shock. We consider two non-nested forecasting models for each of the macroeconomic variables:

$$y_t = \alpha + \sum_{i=1}^k \beta_i y_{t-i} + \sum_{i=0}^k \gamma_i w_{t-i} + e_t \quad (1.17)$$

$$y_t = \alpha + \sum_{i=1}^k \beta_i y_{t-i} + \sum_{i=0}^k \lambda_i \varepsilon_{mt-i} + \sum_{i=0}^k \varphi_i \varepsilon_{ot-i} + e_t \quad (1.18)$$

where the lag-length, k , is chosen by the AIC, and e_t is the error term. Since these are non-nested models, we cannot simply use the standard F -tests and t -tests to determine which model fits better. Davidson and MacKinnon (1981) proposed the J -test as an alternative way of comparing non-nested hypotheses. If (1.17) is the true model, then the fitted values from (1.18) should not have significant coefficients when added to equation (1.17). We test the null hypothesis that (1.17) is the best prediction

model for the given macroeconomic variable as follows:

1. Estimate equation (1.18) by OLS and obtain the fitted values \hat{y}_t .
2. Estimate the model

$$y_t = \alpha + \sum_{i=1}^k \beta_i y_{t-i} + \sum_{i=0}^k \gamma_i w_{t-i} + \sum_{i=0}^k \eta_i \hat{y}_{t-i} + e_t \quad (1.19)$$

A significant F -statistic on $\eta_0 = \eta_1 = \dots = \eta_k = 0$ implies a rejection of the null hypothesis that (1.17) is the best prediction model for y_t .

3. Estimate equation (1.18) by OLS and obtain the fitted values \check{y}_t .
4. Estimate the model

$$y_t = \alpha + \sum_{i=1}^k \beta_i y_{t-i} + \sum_{i=0}^k \lambda_i \varepsilon_{mt-i} + \sum_{i=0}^k \varphi_i \varepsilon_{ot-i} + \sum_{i=0}^k \eta_i \check{y}_{t-i} + e_t \quad (1.20)$$

A significant F -statistic on $\eta_0 = \eta_1 = \dots = \eta_k = 0$ implies a rejection of the null hypothesis that (1.18) is the best prediction model for y .

Table 1.5 shows the J -test results. $t_{\Delta oil}^{jtest}$ is the J -test statistic for (1.19) and t_{nopi}^{jtest} is the J -test statistic for (1.20). Only for the federal funds rate do we fail to reject the null hypothesis that model (1.17) provides the best predictions at a 5% significance level. Interestingly, the J -test in some cases also rejects that (1.18) is the proper specification.

1.4.2.2 Out-of-Sample Analysis

Significant in-sample predictability does not necessarily translate into good out-of-sample forecasts. Structural breaks and the difficulty of obtaining precise parameter estimates are reasons why a model might fit well in the full sample but fail to provide improvements over simpler models in an out-of-sample forecasting experiment.

Our forecasting models are based on models (1.17) and (1.18). The recursive one-step ahead forecasting models are then given by:

$$\hat{y}_{t+1|t} = \hat{\alpha} + \sum_{i=1}^k \hat{\beta}_i \hat{y}_{t+1-i|t} + \sum_{i=0}^k \hat{\gamma}_i \hat{w}_{t+1-i|t} \quad (1.21)$$

$$\hat{y}_{t+1|t} = \hat{\alpha} + \sum_{i=1}^k \hat{\beta}_i \hat{y}_{t+1-i|t} + \sum_{i=0}^k \hat{\lambda}_i \hat{\varepsilon}_{mt+1-i|t} + \sum_{i=0}^k \hat{\varphi}_i \hat{\varepsilon}_{ot+1-i|t} \quad (1.22)$$

where the lag length, k , is chosen using the AIC.

We measure forecast performance using a mean squared forecast error (MSE) loss function. Because the models (1.21) and (1.22) are non-nested, we compute Diebold and Mariano (DM, 1995) test statistics to compare the forecasts of each macroeconomic variable produced by the two models. The null hypothesis is that the forecasts of the two models have equal MSE. That is

$$\xi [(e_{i,t+1|t})^2] = \xi [(e_{j,t+1|t})^2]$$

where ξ is the expectations operator, $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$, and i and j index the models. The DM test is carried out as follows. Define the loss associated with the use of model k as

$$L(e_{k,t+1|t}) = (e_{k,t})^2$$

where the form of $L(\cdot)$ follows from the assumption of MSE loss.

We want to test the null hypothesis

$$H_0 : \xi [L(e_{t+1|t})^i] = \xi [L(e_{t+1|t})^j]$$

Table 1.5: In-Sample J -Test Results: Our Model Vs Change in Oil Price Model

Variable	$t_{\Delta oil}^{jtest}$	t_{nopi}^{jtest}	Lags
Core CPI	2.0903	1.5674	6
CPI (All goods)	4.8163	5.3466	12
Federal Funds Rate	1.8579	3.3677	6
Federal Funds Rate (Δ)	2.4482	1.8915	3
Industrial Production	6.3452	0.0720	4
M2 Money Supply	4.7545	1.9526	5
PPI (All commodities)	3.5061	6.3345	7
Core PPI	2.9307	2.8989	6
3-month T-bill	3.7885	2.6677	4
Unemployment	4.7398	0.5227	9

Notes:

t_{nopi}^{jtest} is the J -test t -statistic of model (1.20)- our model and $t_{\Delta oil}^{jtest}$ is the J -test t -statistic of model (1.19) - the change in oil price model

Boldface numbers represent significant t-statistics. Significant t-statistics are a rejection of the null hypothesis that the model in question provides better in-sample fit.

against the two-sided alternative

$$H1 : \xi [L(e_{t+1|t})^i] \neq \xi [L(e_{t+1|t})^j]$$

Define the loss differential series as¹⁰

$$d_t = L(e_{i,t+1|t}) - L(e_{j,t+1|t})$$

Diebold and Mariano (1995) and West (1996) showed that with non-nested models and MSE loss, d is distributed normally with a zero mean. The null hypothesis can be tested by regressing d_t on a constant

$$d_t = \alpha + e_t$$

¹⁰In other words, simply take the difference in squared forecast errors at each date that a forecast is made.

and (using the Newey-West HAC corrected t -statistic) testing whether α is significantly different from zero.

If $DM > 1.96$, then the average difference in forecast loss is different from zero and hence model 2 forecasts better. If $DM < -1.96$, we conclude that model 1 forecasts better. If $-1.96 < DM < 1.96$, we cannot reject the hypothesis that both models forecast equally well.

Table 1.6 reports the results of the forecast comparisons for one-step ahead forecasts of each of the macroeconomic variables. In only two cases, for the core CPI and the PPI, does the change in the price of oil provide better forecasts of the macroeconomic variables, and in both of those cases the difference in MSE is small and insignificant. In five cases there is a statistically significant improvement in the forecasts of the macroeconomic variables when we first decompose oil price movements into the underlying stock market and oil market structural shocks.

Table 1.6: One-step-ahead Forecasts

Variable	<i>MSE</i> Δ_{oil}	<i>MSE</i> <i>NOPI</i>	<i>DM</i> $\Delta_{oil}/NOPI$
Core CPI	<i>0.0080</i>	0.0082	-0.6085
CPI (All goods)	0.0679	<i>0.0435</i>	2.3718
Federal Funds Rate	0.0159	<i>0.0143</i>	0.5536
Federal Funds Rate (Δ)	0.0146	<i>0.0112</i>	1.9368
Industrial Production	0.5129	<i>0.4929</i>	0.5724
M2 Money Supply	0.0718	<i>0.0449</i>	3.5552
PPI (All commodities)	<i>0.3969</i>	0.4107	-0.3532
Core PPI	0.0675	<i>0.0543</i>	3.6614
3-month T-bill	0.0157	<i>0.0025</i>	5.7590
Unemployment Rate	0.0050	<i>0.0005</i>	4.7049

Notes:

Numbers in italics indicate that the given model has a lower MSE

Boldface DM test statistics indicates the model with the underlying structural shocks forecast better than the change in oil price model

To ensure that the one-step ahead forecasts are not driven by outliers, we plot the forecasts, as well as the forecasts errors. Figures 1.4 to 1.13 show no significant outliers

that may possibly be driving our results. We conclude therefore that in many instances, our new model forecasts better than the change in oil price model.

Figure 1.4: One-Step Ahead Forecasts and Forecast Errors - Core CPI Inflation

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

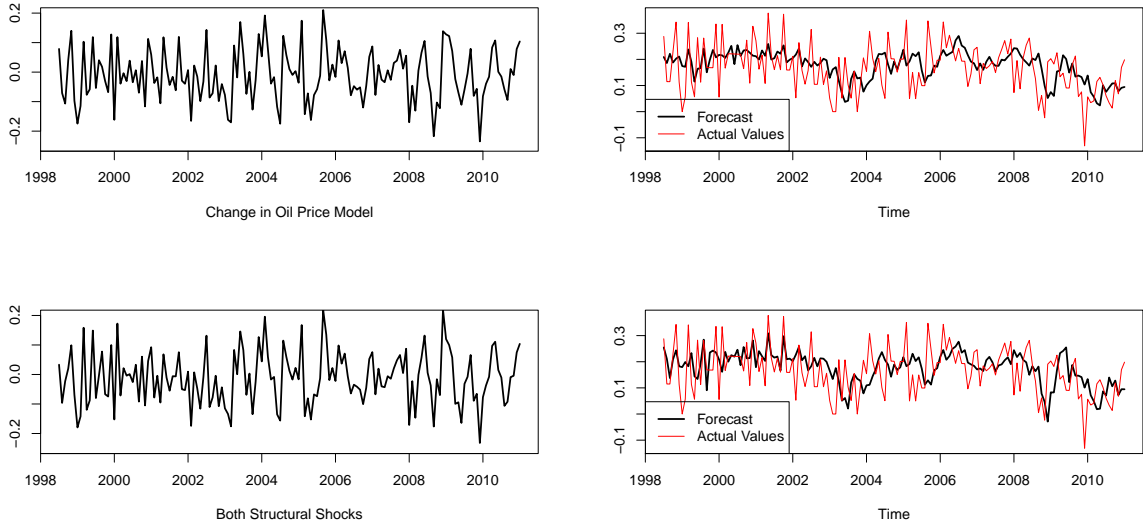


Figure 1.5: One-Step Ahead Forecasts and Forecast Errors - CPI Inflation

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

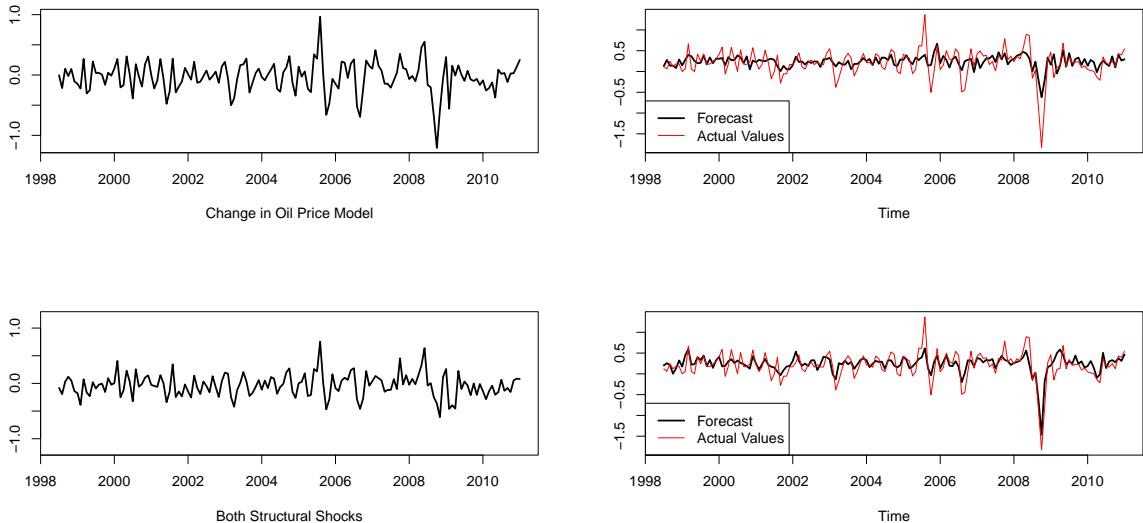


Figure 1.6: One-Step Ahead Forecasts and Forecast Errors - Federal Funds Rate: 1998:7-2011:2

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

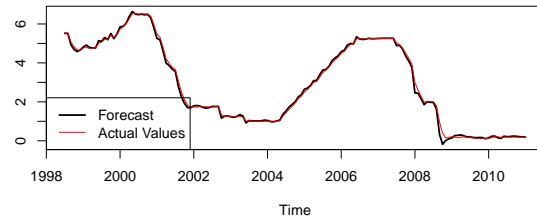
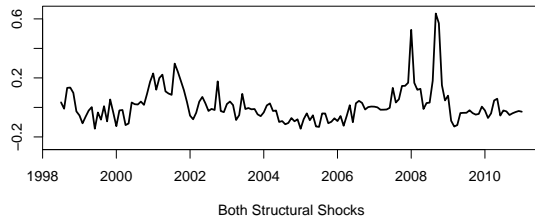
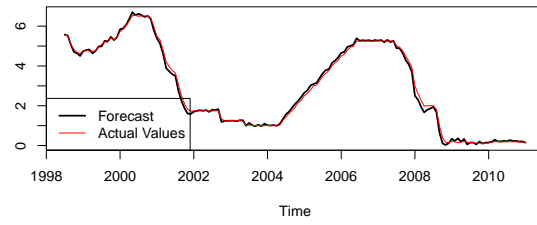
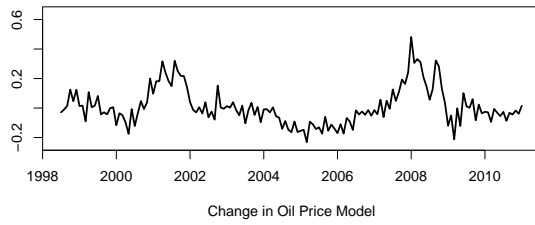


Figure 1.7: One-Step Ahead Forecasts and Forecast Errors - Change in Federal Funds Rate: 1998:7-2011:2

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

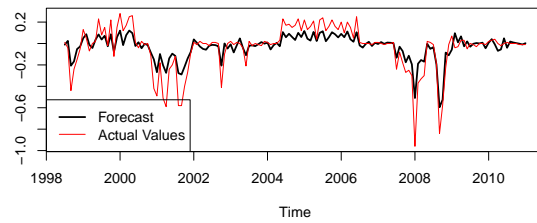
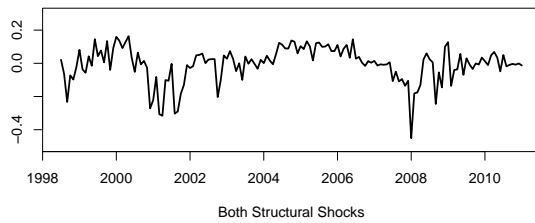
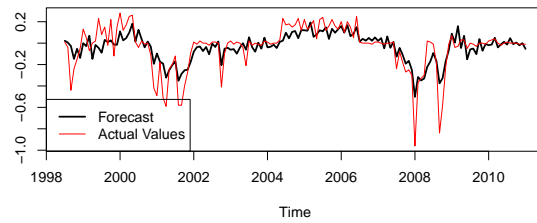
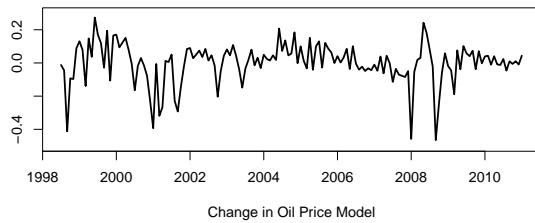


Figure 1.8: One-Step Ahead Forecasts and Forecast Errors - Industrial Production: 1998:7-2011:2

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

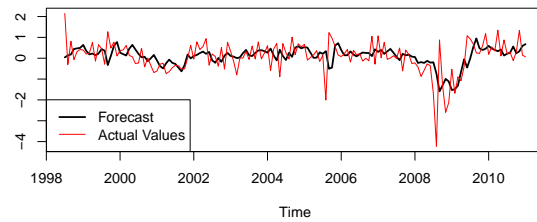
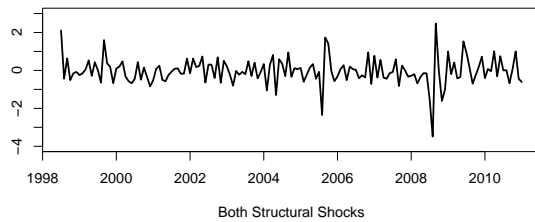
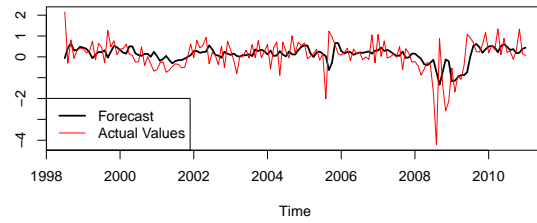
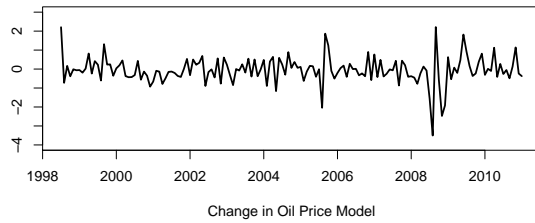


Figure 1.9: One-Step Ahead Forecasts and Forecast Errors - M2 Money Supply: 1998:7-2011:2

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

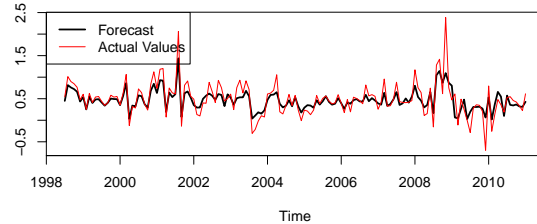
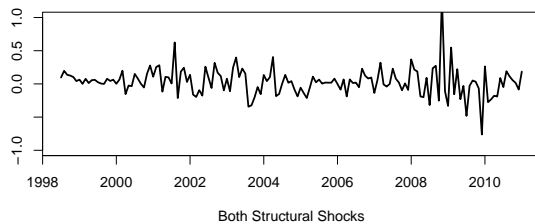
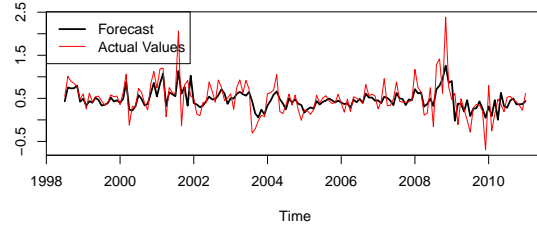
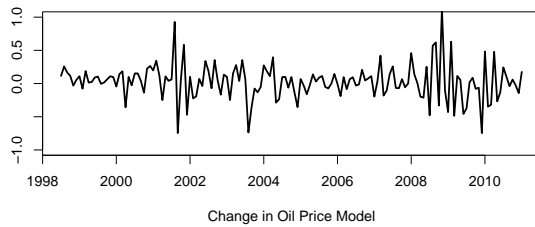


Figure 1.10: One-Step Ahead Forecasts and Forecast Errors - PPI Inflation (All Commodities): 1998:7-2011:2

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

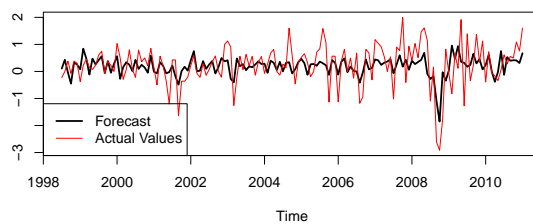
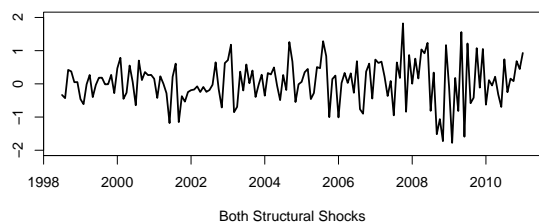
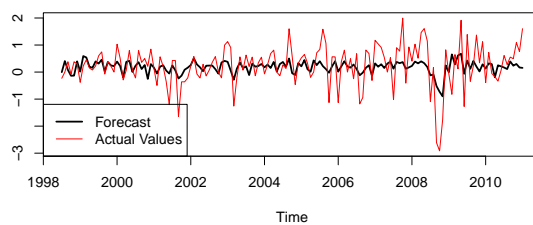
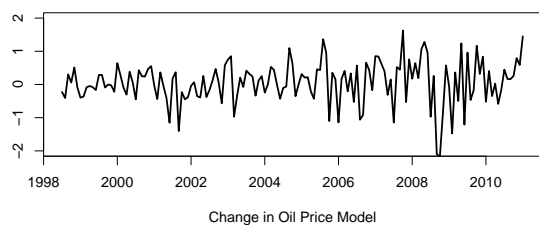


Figure 1.11: One-Step Ahead Forecasts and Forecast Errors - Core PPI: 1998:7-2011:2

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

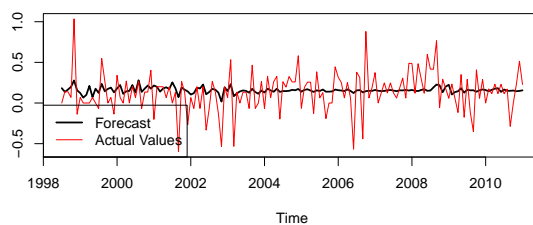
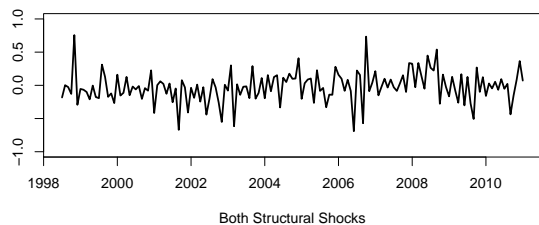
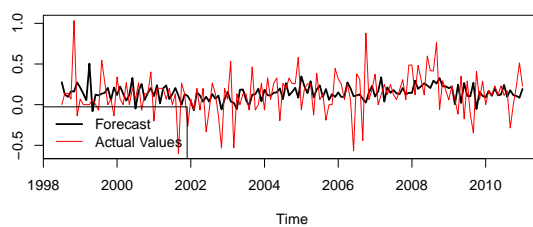
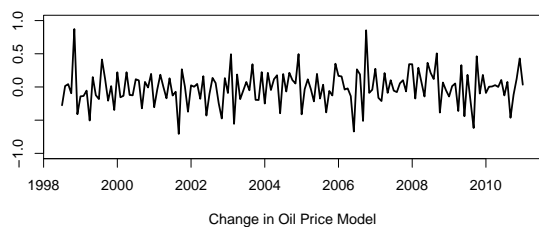


Figure 1.12: One-Step Ahead Forecasts and Forecast Errors - 3-Month T-Bill Rate: 1998:7-2011:2

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

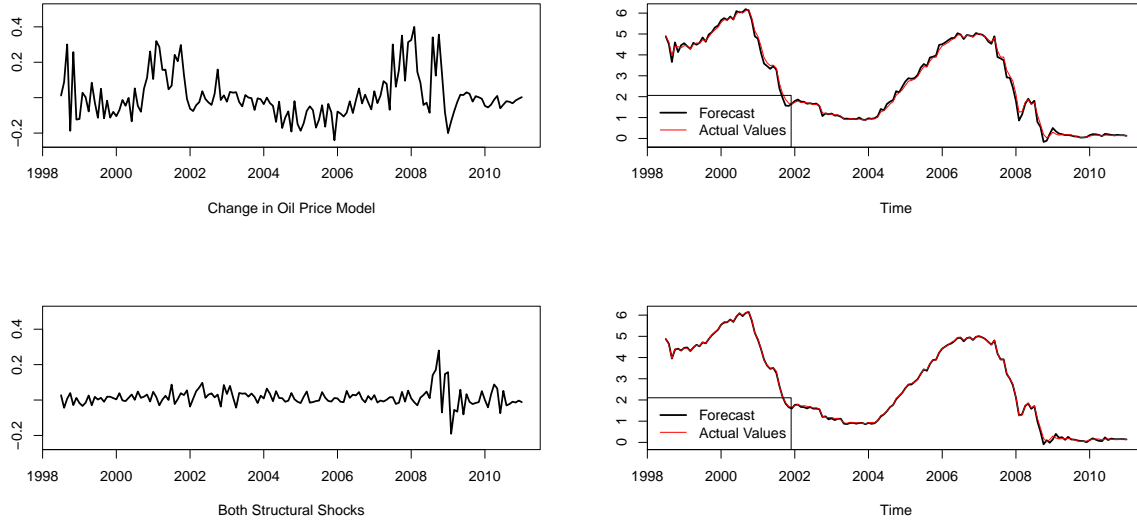


Figure 1.13: One-Step Ahead Forecasts and Forecast Errors - Unemployment Rate: 1998:7-2011:2

a. Forecast Errors for Each Model

b. Actual Values Vs Forecasts

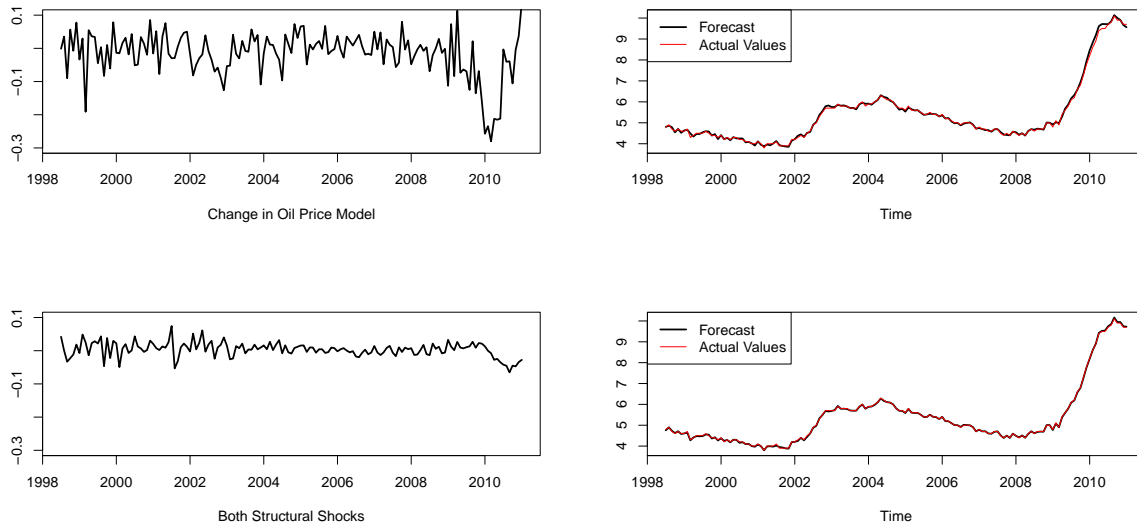


Table 1.7 presents DM test statistics for the comparison of the same two models

but at longer forecast horizons. To this end, we compute the *MSE* and *DM* test for each of these variables for 3-, 6-, 12-, and 24-step ahead forecasts. For this exercise, we use a direct estimation approach where we estimate a horizon-specific model and form direct h -step ahead predictions of y_t . The h -step ahead forecasts for the models we estimate are:

$$\hat{y}_{t+h|t} = \hat{\alpha} + \sum_{i=1}^k \hat{\beta}_i \hat{y}_{t-i|t} + \sum_{i=1}^k \hat{\gamma}_i \hat{w}_{t-i|t} \quad (1.23)$$

$$\hat{y}_{t+h|t} = \hat{\alpha} + \sum_{i=1}^k \hat{\beta}_i \hat{y}_{t-i|t} + \sum_{i=1}^k \hat{\lambda}_i \hat{\varepsilon}_{mt-i|t} + \sum_{i=1}^k \hat{\varphi}_i \hat{\varepsilon}_{ot-i|t} \quad (1.24)$$

The general pattern is the same at longer horizons as for the one-step ahead forecasts. Using the change in the price of oil leads to inferior forecasts in almost all cases. In many cases the difference in MSE is significant, so that we can conclude that the model with decomposed oil shocks produces forecasts with a statistically significantly lower MSE.

Table 1.7: Multi-Step Ahead Forecasts

Variable	$h = 3$	$h = 6$	$h = 12$	$h = 24$
Core CPI	-1.5938	-1.4154	-0.9195	-2.8265
CPI (All goods)	0.8298	0.5863	2.2315	-0.9233
Federal Funds Rate	0.5954	1.6435	1.4515	0.0013
Federal Funds Rate (Δ)	1.2032	2.3185	3.3973	2.4055
Industrial Production	1.1422	1.1958	1.0603	0.8037
M2 Money Supply	2.5946	1.9227	2.0282	1.6832
PPI (All commodities)	0.6961	-1.7363	1.5787	-1.2691
Core PPI	2.7733	3.0816	1.5161	2.2484
3-month T-bill	2.7820	2.9544	2.9680	0.4695
Unemployment Rate	3.5026	1.6606	2.8111	2.3971

NOTES:

Boldface numbers indicate that the model with the underlying structural shocks forecasts better than the change in the oil price model.

1.5 Conclusion

Using daily data on the S&P 500 returns and the West Texas Intermediate (WTI) crude oil prices for the period 1986-2011, and making use of Hamilton's (1996, 2003, 2010) net oil price increase model, we identified stock market and oil market shocks that affect the stock and oil markets. We show that the response of the stock market differs depending on whether the price of oil has risen due to a stock market shock raising the demand for oil, or due to an oil-market shock decreasing the supply of oil. The stock market shock is shown to be associated with about 97.8% of the variation in stock returns, while the oil-market shock accounts for over 98% of the variation in oil prices.

We find that when we are not in a net oil price increase month, a stock market shock that causes the S&P 500 to rise by one percentage point causes the price of oil to rise approximately 0.2 percentage points. This finding is consistent with the price of oil being an asset price that fully reflects new information immediately. We also find a statistically significant positive effect on the day after the stock market shock. The implication is that every time the price of oil hits a new high, traders initially underestimate the effect of a stock market shock on the demand for oil. On the other hand, the response of the stock market to an oil market shock is a decline of 6.8% when the price of oil doubles.

We then analyzed the response of monetary policy to the identified stock market and oil price shocks using daily data on the change in the federal funds rate and the 30, 60, and 180-day eurodollar rates. We found that short-term interest rates do not respond to oil shocks, but that it would be very misleading to simply regress the interest rate series on the change in the price of oil. We also considered whether the information from the decomposed stock market and oil shocks improve the in-sample fit and the historical out-of-sample forecast performance of our model relative to a benchmark model that

includes only the change in the price of oil as a predictor. We concluded that the model with the decomposed oil shocks produces better forecasts both in-sample and out-of-sample.

Chapter 2

Revisiting the Statistical Specification of Near-Multicollinearity in the Logistic Regression Model

2.1 Introduction

In econometrics, statistics and related disciplines, multiple linear regression models are widely used to assess the relationship between two or more variables. The interpretation and validity of these models is dependent upon the validity of individual regression coefficient estimates. When the regressors in the model are not linearly related, they are said to be orthogonal or independent. Traditionally, when the regressors are not orthogonal and become almost perfectly related, estimates of the individual regression coefficients may become unstable and the inferences based on the model may tend to be misleading. This condition is known as multicollinearity (Mason, Webster and Gunst, 1975). Multicollinearity occurs when variables in the model are correlated to an extent that individual regression coefficient estimates become unreliable. When the

variables have an exact linear relationship, they are said to be perfectly collinear. When the relationship between the predictor variables is almost linear (but not exact), this results in the phenomenon known as **near-multicollinearity**, the problem specifically addressed in this paper.

The traditional account of near-multicollinearity in the linear regression model relies upon a *ceteris paribus* clause which allows for the examination of the impacts on the standard errors ($se(\beta)$) and asymptotic t – *statistics* ($\tau(\beta)$) (or other test statistics) as the correlation (ρ) between the regressors increases ($|\rho| \rightarrow 1$), while holding the parameter estimates β , σ^2 and R^2 at their current values (i.e constant). However, β , σ^2 and R^2 , as well as their associated variances and t – *ratios* are all functions of ρ , changing as $|\rho| \rightarrow 1$ (Spanos and McGuirk, 2002). Therefore, the *ceteris paribus* clause is unattainable, making it necessary to re-evaluate the specification of near-multicollinearity in statistical models (Spanos and McGuirk, 2002). Making use of the notion of *statistical reparametization*, Spanos and McGuirk (2002) revisit the problem of near-multicollinearity in the context of the linear regression model and conclude that their:

revised account of the changes in the statistics ($V\hat{ar}(\hat{\beta});\tau(\hat{\beta}); R^2$) induced by $|\hat{\rho}| \rightarrow 1$, shows that the traditional account needs to be thoroughly amended; *neither of the statistics ($V\hat{ar}(\hat{\beta});\tau(\hat{\beta})$) varies monotonically with $\hat{\rho}$, and there is no conflict between the relevant t -ratios and the R^2* (p. 392).

From the traditional perspective, Hosmer and Lemeshow (1989) note that the problems caused by the presence of near-multicollinearity in the logistic regression model are similar to those in the linear regression model. However, because the conventional discussion about near-multicollinearity in the linear regression model needs to be revisited as noted by Spanos and McGuirk (2002), it follows that the traditional discussion of near-multicollinearity in the logistic regression model needs to be revisited as well.

The purpose of this paper is to re-examine the statistical specification of near multicollinearity in the logistic regression model. Making use of the *Probabilistic Reduction* (PR) approach to model specification, we show that the logistic regression model parameters are functions of ρ . Like the linear regression model, the *ceteris paribus* clause fails to hold for the logistic regression model, changing the traditional account of near-multicollinearity for this discrete choice model, as well. Hence the specification of near-multicollinearity for the logistic regression needs to be amended. A simple two covariate case is simulated to examine what happens to the parameter estimates and other statistics (e.g. asymptotic t-ratios, marginal effects, variance inflation factors) as $|\rho| \rightarrow 1$. In addition, some extensions for examining logistic regression models with nonlinear index functions and binary covariates is presented as extensions to provide a base for future research.

The paper proceeds as follows. As a prelude to the discussion of near-multicollinearity in the logistic regression model, the next section reviews the traditional and revised accounts of the specification of near-multicollinearity for the linear regression model. Section 2.3 presents the traditional discussion of the logistic regression model, and using the probabilistic approach, section 2.4 presents a revised account of near-multicollinearity in the context of the logistic regression model with continuous covariates. Section 2.5 presents the results from a simulation study of the revised account presented in section 2.4. In section 2.6, we extend the analysis to specifications of the logistic regression model with binary covariates and nonlinear index/predictor functions. Section 2.7 concludes.

2.2 Multicollinearity in the Linear Regression Model

In this section, we take a closer look at near-multicollinearity in the linear regression model. In section 2.2.1, we review the usual account of near-multicollinearity, discuss its consequences, and present conventional methods for detecting it. In section 2.3, we present the revised account of near-multicollinearity as presented by Spanos and McGuirk (2002).

2.2.1 Multicollinearity in the Linear Regression Model: The Traditional Story

Consider the linear regression model

$$y = \mathbf{X}\beta + e$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{k,1} \\ x_{1,2} & \cdots & x_{k,2} \\ \vdots & \cdots & \vdots \\ x_{1,T} & \cdots & x_{k,T} \end{bmatrix}, e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix} \sim N_T(0, \sigma^2 I_n)$$

Perfect multicollinearity occurs when at least one of the columns of \mathbf{X} is a linear transformation of the others (Bierens, (2007)). This situation can occur when the correlation (ρ) between the regressors is equal to one in absolute value. Perfect multicollinearity results in parameter identification problems since the $(\mathbf{X}'\mathbf{X})$ matrix is singular (see Greene, 2000, p. 256). Thus, $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}y$, $\hat{s}^2 = \frac{\hat{e}'\hat{e}}{T-K}$, and $\widehat{Cov}(\hat{\beta}) = \hat{s}^2(\mathbf{X}'\mathbf{X})^{-1}$ are indeterminate.

On the other hand, if the smallest eigenvalue of the $(\mathbf{X}'\mathbf{X})$ matrix is very small

such that $\det(\mathbf{X}'\mathbf{X}) \approx 0$, but $(\mathbf{X}'\mathbf{X})$ is still nondegenerate, then we have the associated problem of near-multicollinearity. Simply put, near-multicollinearity occurs when $\rho \approx 1$ or very close to 1. When the regressors are highly correlated in this way, the data matrix is not well-conditioned, but the OLS estimators still exist. However, high correlation among the regressors may lead to numerical instability in the precision and significance of the estimates of β .

2.2.1.1 Effects of Near-Multicollinearity

Near-multicollinearity has significant consequences on the least squares estimates of the linear regression coefficients. Greene (2000) summarizes the symptoms of near-multicollinearity as follows:

- Small changes in the data produce wide swings in the parameter estimates.
- Coefficients may have very high standard errors and low significance levels even though they are jointly significant and the R^2 for the regression is quite high.
- Coefficients may have the “wrong” sign or implausible magnitudes (p. 256).

To examine the above discussion of near-multicollinearity, let ϑ be a unit vector of length k . Following Bierens (2007), the $Var(\hat{\beta}) = \sigma^2 \vartheta' (\mathbf{X}'\mathbf{X})^{-1} \vartheta$. Now let Λ represent $diag\{eig(\mathbf{X}'\mathbf{X})\}$, and U the orthogonal matrix of the corresponding eigenvectors. Then one can rewrite the $\mathbf{X}'\mathbf{X}$ matrix as $U\Lambda U'$. Denote by u_m , the m^{th} row of the matrix U , and λ_m , the m^{th} diagonal element of the eigenvector Λ . Then the variance of $\hat{\beta}$ can be rewritten as (Bierens, 2007):

$$Var(\hat{\beta}) = \sigma^2 \vartheta' U \Lambda^{-1} U' \vartheta = \sigma^2 \sum_{m=1}^k \frac{u_m^2}{\lambda_m} \quad (2.1)$$

Now suppose $\lambda_1 < \lambda_m$ for $m = 2, 3, \dots, k$ and $u_1 \neq 0$, where λ_1 is the smallest eigenvalue. Then from equation (2.1), as $\lambda_1 \rightarrow 0$, the $Var(\hat{\beta}) \rightarrow \infty$. This implies

that when near-multicollinearity is present, the variances and standard errors of the coefficients will become inflated. Furthermore, confidence intervals for coefficients will become wide and t -statistics will be smaller.

2.2.1.2 Detecting Near-Multicollinearity

How can one distinguish true parameter instability that results from near-multicollinearity from other forms of model misspecification? Some researchers have suggested different “rules of thumb” for detecting near-multicollinearity such as: (i) the Variance Inflation Factor (VIF) (and Tolerance Index), and (ii) the Condition Number .

The *VIF* indicates how much the variance of a regression coefficient increases over and above what it would otherwise be if the R^2 of the regression were equal to zero. It is therefore an index that quantifies the degree of multicollinearity present in a model. To see this, recall that the $Var(\hat{\beta}_i)$ can be written as $Var(\hat{\beta}_i) = \frac{\sigma^2}{\sigma_{ii(1-R_i^2)}}$ (Spanos and McGuirk, 2002), where $\sigma_{ii} = \frac{1}{T} \sum_{t=1}^T (x_{it} - \bar{x}_i)^2$ and R_i^2 is the unadjusted R^2 when x_i is regressed on the other explanatory variables. The Tolerance Index for x_i , is defined as 1 minus the fraction of the variance between x_i and the other regressors ($1 - R_i^2$). The Variance Inflation Factor (VIF) is the inverse of the tolerance. The $VIF(\hat{\beta})$, and its corresponding Tolerance Index can be written as (see O’Brien, 2007):

$$VIF(\hat{\beta}) = \frac{\frac{\sigma^2}{\sigma_{ii}}}{\frac{\sigma^2}{\sigma_{ii(1-R_i^2)}}} = \frac{1}{(1 - R_i^2)} \quad \text{and} \quad Tolerance(\hat{\beta}) = \frac{1}{VIF(\hat{\beta})} = 1 - R_i^2.$$

If x_i is orthogonal with the other predictor variables, then $R_i^2 = 0$, hence $VIF(\hat{\beta}) = 1$. On the other hand, if there is near-multicollinearity between x_i and the other regressors, $R_i^2 \rightarrow 1$ and $VIF(\hat{\beta}) \rightarrow \infty$. Thus when $Var(\hat{\beta})$ is high, the $VIF(\hat{\beta})$ is high and $Tolerance(\hat{\beta})$ is low. As a rule of thumb, a $VIF > 10$ is indicative of a significant

multicollinearity problem (Marquardt, 1970). Despite its wide use and appeal, it must be noted that the *VIF* is not without its deficiencies. A significant drawback of the *VIF* is that it only detects overall near-multicollinearity problems that do not include the constant term. The statistic is neither able to detect multiple near-singularities nor determine the source of the singularities (Rawlings, Pantula and Dickey, 1998)

As examined earlier, an eigensystem analysis of the data matrix $\mathbf{X}'\mathbf{X}$ can also be used to detect the presence of near-multicollinearity. If there exists one or more almost linear relationships between the explanatory variables, then one or more of the eigenvalues will be small. Let $\lambda_1, \dots, \lambda_k$ be the eigenvalues of the matrix $\mathbf{X}'\mathbf{X}$. Then the condition number of $\mathbf{X}'\mathbf{X}$ is defined as (Belsey, Kuh and Welsch, 1980)

$$\kappa = \left[\frac{\lambda_{max}}{\lambda_{min}} \right]^{\frac{1}{2}}$$

where λ_{max} is the maximum eigenvalue of the matrix $\mathbf{X}'\mathbf{X}$, and λ_{min} is the smallest eigenvalue of the $\mathbf{X}'\mathbf{X}$ matrix. When the condition number of $\mathbf{X}'\mathbf{X}$ is low, $\mathbf{X}'\mathbf{X}$ is said to be well-conditioned. On the other hand, if $\mathbf{X}'\mathbf{X}$ has a high condition number, it is said to be ill-conditioned.

Condition numbers can be extended to condition indexes for each principal component. The condition index of $\mathbf{X}'\mathbf{X}$ for the i^{th} principal component, denoted κ_i is defined as:

$$\kappa_i = \frac{\lambda_{max}}{\lambda_i} \quad i = 1, 2, \dots, k$$

A large number of condition indices that are high is indicative of strong linear dependencies in the data matrix $\mathbf{X}'\mathbf{X}$. An informal rule of thumb is that if (Belsey, Kuh and Welsch, 1980):

- $\kappa < 100$ or $\kappa_j < 10$ multicollinearity is not a major concern.

- $100 < \kappa < 1000$ or $30 < \kappa_j < 100$ multicollinearity is of moderate concern.
- $\kappa > 1000$ or $\kappa_j > 100$ multicollinearity is a serious concern.

Snee and Marquardt (1984) note that there is no difference between the rules of thumb for serious near-multicollinearity when the $VIF > 10$, and when the condition number (condition index): $100 < \kappa < 1000$ or $30 < \kappa_j < 100$. It must be noted that both the VIF (and *Tolerance*) and the condition number only provide rules of thumb to guide the modeler in detecting the presence of near-multicollinearity as there exists no definitive thresholds for the detection of these problems in the literature.

2.2.2 Revisiting the Problem of Near-multicollinearity in the Linear Regression Model

Spanos and McGuirk (2002) explore in detail the problem of near-multicollinearity in the context of the linear regression model. In their paper, they argue that the traditional discussion of near-multicollinearity reduces to two distinct issues: a structural issue associated with high correlation among regressors leading to *systematic volatility*; and a numerical issue associated with the data matrix $\mathbf{X}'\mathbf{X}$ being ill-conditioned leading to *erratic volatility* (see Spanos and McGuirk (2002) for a detailed discussion). They argue that even though the traditional account attempts to model erratic volatility, it inadvertently leads to a discussion of systematic volatility (high correlation among the regressors).

They consider the following linear model

$$y = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + e \quad e \sim N(0, \sigma^2 I_N)$$

The OLS estimators of β are:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2, \quad \hat{\beta}_1 = \frac{\hat{\sigma}_{33} \hat{\sigma}_{12} - \hat{\rho} \hat{\sigma}_{13} \sqrt{\hat{\sigma}_{22} \hat{\sigma}_{33}}}{(1 - \hat{\rho}^2) \hat{\sigma}_{22} \hat{\sigma}_{33}}, \quad \hat{\beta}_2 = \frac{\hat{\sigma}_{22} \hat{\sigma}_{13} - \hat{\rho} \hat{\sigma}_{12} \sqrt{\hat{\sigma}_{22} \hat{\sigma}_{33}}}{(1 - \hat{\rho}^2) \hat{\sigma}_{22} \hat{\sigma}_{33}}$$

where $\hat{\sigma}^2 = \hat{\sigma}_{11} - \frac{\hat{\sigma}_{33} \hat{\sigma}_{12} - 2 \hat{\rho} \hat{\sigma}_{12} \hat{\sigma}_{13} \sqrt{\hat{\sigma}_{22} \hat{\sigma}_{33}} + \hat{\sigma}_{22} \hat{\sigma}_{13}}{(1 - \hat{\rho}^2) \hat{\sigma}_{22} \hat{\sigma}_{33}}$, $R^2 = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_{11}^2}$, $\hat{\rho}^2 = \frac{\hat{\sigma}_{23}^2}{\hat{\sigma}_{22} \hat{\sigma}_{33}}$, and

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T (z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j), \quad i, j = 1, 2, 3; \quad z_{1t} = y_t; \quad z_{2t} = x_{1t}; \quad z_{3t} = x_{2t} \quad (\text{Spanos and McGuirk, 2002})$$

Spanos and McGuirk (2002) show that the estimated variances of the above estimators take the form:

$$Var(\hat{\beta}_1) = \frac{T s^2}{(1 - \hat{\rho}^2) \hat{\sigma}_{22}}, \quad Var(\hat{\beta}_2) = \frac{T s^2}{(1 - \hat{\rho}^2) \hat{\sigma}_{33}}, \quad s^2 = \frac{\sum_{t=1}^T \hat{e}_t^2}{T - 3};$$

and the corresponding t -statistics, $\tau(\beta_i)$ are:

$$\tau(\hat{\beta}_1) = \frac{\hat{\beta}_1}{s \sqrt{T}} \sqrt{(1 - \hat{\rho}^2) \hat{\sigma}_{22}}, \quad \tau(\hat{\beta}_2) = \frac{\hat{\beta}_2}{s \sqrt{T}} \sqrt{(1 - \hat{\rho}^2) \hat{\sigma}_{33}}$$

The estimators and their associated variances and t -ratios (both numerators and denominators) are all functions of $\hat{\rho}$. Note that $s^2 \xrightarrow{p} \sigma^2$, which is a function of ρ .

The conventional account of near-multicollinearity is associated with the effects of $|\hat{\rho}| \rightarrow 1$, **ceteris paribus**. The **ceteris paribus** clause assumes that $\hat{\beta}$, $\hat{\sigma}^2$, and R^2 are held constant, when analysing the impact as $|\hat{\rho}| \rightarrow 1$ on the statistics $\tau(\hat{\beta}_1)$, $\tau(\hat{\beta}_2)$, $Var(\hat{\beta}_1)$, and $Var(\hat{\beta}_2)$. However, because $\hat{\beta}$, $\hat{\sigma}^2$, and R^2 as well as their associated variances and t -ratios as shown above are all functions of $\hat{\rho}$, the changes in these statistics are rather different than the traditional account implies. Therefore the **ceteris paribus** assumption fails to hold. Making use of the notion of *statistical reparametization*, Spanos and McGuirk (2002) thoroughly revisit this issue and conclude that their

revised account of the changes in the statistics ($Var(\hat{\beta}); \tau(\hat{\beta}); R^2$) induced

by $|\hat{\rho}| \rightarrow 1$, shows that the traditional account needs to be thoroughly amended; *neither of the statistics $(\text{Var}(\hat{\beta}); \tau(\hat{\beta}))$ varies monotonically with $\hat{\rho}$, and there is no conflict between the relevant t -ratios and the R^2* (p. 392).

Spanos and McGuirk (2002) goes on to show that the problem of near-multicollinearity in the linear regression model may be one of ill-conditioning (erratic volatility), and not necessarily always one of high correlation among regressors (systematic volatility). These problems are likely to exist in nonlinear models, as well. In this paper, the focus primarily is on the story concerning systematic volatility in the context of the logistic regression model. We start by reviewing the statistical properties of the logistic regression model.

2.3 Multicollinearity in the Logistic Regression Model

So far, we have discussed the traditional and revised accounts of near-multicollinearity in the context of the linear regression model. In this section, we extend the discussion to strong dependencies among the regressors in the logistic regression model. Section 2.3.1 reviews the traditional specification of the logistic regression model, while section 2.3.2 discusses the traditional approach to detecting and dealing with near-multicollinearity in the logistic model.

2.3.1 The Logistic Regression Model

The logistic regression model describes the relationship that exists between a binary response variable and other categorical or continuous independent variables. The binary response variable y_i is binomial with mean parameter p , (i.e $y_i \sim \text{Bin}(p)$), and \mathbf{X}_i is a $k \times 1$ vector of predictors or explanatory variables. The conditional mean of the logistic

regression model is $E(y_i|\mathbf{X}_i) = Prob(y_i = 1|\mathbf{X}_i)$, such that

$$\begin{aligned} Prob(y_i = 1|\mathbf{X}_i; \beta) &= p = F(\mathbf{X}_i, \beta) = \frac{\exp(\beta' \mathbf{X}_i)}{1 + \exp(\beta' \mathbf{X}_i)} = \frac{1}{1 + \exp(-\beta' \mathbf{X}_i)}, \\ Prob(y_i = 0|\mathbf{X}_i; \beta) &= 1 - p = 1 - F(\mathbf{X}_i, \beta) = \frac{1}{1 + \exp(\beta' \mathbf{X}_i)} \end{aligned}$$

The logistic regression or logit model can be thought of as the canonical link for the binomial distribution within the rather broad class of generalized linear models (GLM), because it is the link function of $E(y_i|\mathbf{X})$ for which the predictor is linear, i.e.:

$$h(\mathbf{X}_i, \beta) = \ln \frac{p}{1-p} = \beta' \mathbf{X}_i \quad (2.2)$$

Unlike ordinary least squares regression, estimation of the parameters of the logistic regression model is done using the method of maximum likelihood (Peng, Lee and Ingersoll, 2002), or equivalently by iterative weighted least squares (IWLS) (Schaefer, 1985). The log likelihood function is derived from the joint probability distribution of the dependent and explanatory variables, and can be written as:

$$\begin{aligned} \text{Log - Likelihood } l(\beta|y; \mathbf{X}) &= \sum_{i=1}^n f(y_i|\mathbf{X}_i; \beta) \\ &= \sum_{i=1}^n \left[y_i \log \left(\frac{\exp(\beta' \mathbf{X}_i)}{1 + \exp(\beta' \mathbf{X}_i)} \right) + (1 - y_i) \log \left(1 - \frac{\exp(\beta' \mathbf{X}_i)}{1 + \exp(\beta' \mathbf{X}_i)} \right) \right] \end{aligned}$$

The objective during estimation is to find β that maximizes the above log-likelihood function. Taking the derivative of the log-likelihood function and setting it to 0 results in the score/gradient function. The parameter estimates $\hat{\beta}$ are obtained by solving the score function. Because the score function is highly nonlinear, and cannot be solved analytically, a numerical iterative procedure, such as the Newton-Raphson algorithm or the method of scoring must be used for estimation.

The $\hat{\beta}$ s in the logistic regression model can be interpreted as the marginal effect on

the log of the odds ratio for a unit change in the i^{th} regressor. In the case of a binary regressor, say gender, $exp(\hat{\beta}_i)$ is the the odds that the outcome is, say males compared to females. The reliability of these odds, as well as their interpretation heavily depends on the extent to which the regressors are collinear. Another useful measure for substantive inference are the marginal effects or the change in probability of $Y_i = 1$ given a one unit change in $X_{k,i}$. That is, the marginal effect is $\frac{\partial \mathbf{P}(Y_i = 1 | \mathbf{X}_i)}{\partial X_{k,i}}$. Marginal effects can be computed at the means of the explanatory variables or as the mean across the marginal effects calculated for each observation (i.e. average marginal effects) (Greene, 2000).

2.3.2 Multicollinearity in the Logistic Regression Model: The Traditional Story

From the traditional perspective, the problems caused by near-multicollinearity in the logistic regression model are similar to those in the linear regression model (Hosmer and Lemeshow, 1989). This follows from equation (2.2) which expresses the logit as a linear function of the predictors. Therefore, similar to the linear regression model, the presence of near-multicollinearity leads to near-singularity among the columns of the matrix of explanatory variables. This in turn induces numerical instability in the maximum likelihood estimates, $\hat{\beta}$, with severe consequences for their precision, evidenced by inflated variances and deflated asymptotic *t-values* (Schaeffer, 1985).

Following Schaeffer (1986), consider the $(l + 1)st$ iteration using the method of maximum likelihood or IWLS:

$$\hat{\beta}_{l+1} = \hat{\beta}_l + (\mathbf{X}'\hat{V}_l\mathbf{X})^{-1}\mathbf{X}'(y - \hat{\pi}_l),$$

where y is a $n \times 1$ vector of the binary outcome variable, \mathbf{X} is the $n \times k$ dimensional matrix of regressors, $\hat{\pi}_l$ is the vector of fitted probabilities using $\hat{\beta}_l$, $\hat{V}_l = \text{diag} \{ \hat{\pi}_{li}(1 - \hat{\pi}_{li}) \}$,

and $\hat{\pi}_{li}$ is the i^{th} element of $\hat{\pi}_l$. In the linear regression model (see equation 2.1) $\mathbf{X}'\mathbf{X} = U\Lambda U'$; $Var(\hat{\beta}) = \sigma^2 \vartheta' U \Lambda^{-1} U' \vartheta = \sigma^2 \sum_{m=1}^k \frac{u_m^2}{\lambda_m}$, so that for $\lambda_1 < \lambda_m$ for $m = 2, 3, \dots, k$ and $u_1 \neq 0$, a smaller λ_1 , leads to a higher variance of β . Intuitively, near-multicollinearity in the logistic regression model should lead to an analogous problem, since the $Var(\hat{\beta})$ in the logistic regression is approximated by $(\mathbf{X}'\hat{V}\mathbf{X})^{-1}$, which is comparable to $(\mathbf{X}'\mathbf{X})^{-1}$ in the linear case. Define:

1. μ_m^* for $m = 1, \dots, k$ to be the ordered $\mu_1^* \leq \mu_2^* \leq \dots \leq \mu_k^*$ eigenvalues of the $(\mathbf{X}'V\mathbf{X})$ matrix;
2. γ_m^* for $m = 1, \dots, k$ to be the the eigenvectors of the $(\mathbf{X}'V\mathbf{X})$ matrix associated with μ_m^* ; and
3. γ_{mj}^* to be the j^{th} element of γ_m^*

Schaefer (1986) argues that assuming that there exists only one collinear relationship, such that μ_1^* is close to zero (i.e. high near-multicollinearity), then without loss of generality, $\mu_1^* \leq \frac{1}{4}\mu_1$, where μ_1 is the minimum eigenvalue of the $(\mathbf{X}'\mathbf{X})$ matrix. It can further be shown that (Schafer, 1986):

$$\begin{aligned} Var(\beta_m) &= (\mathbf{X}'\hat{V}\mathbf{X})^{mm} \\ &= \sum_{l=1}^k \mu_l^{*-1} \gamma_{lm}^{*2} \geq \mu_1^{*-1} \gamma_{1m}^{*2} \geq 4\mu_1^{-1} \gamma_{1m}^{*2} \end{aligned}$$

For some m , $\gamma_{1m}^* \neq 0$ and $Var(\beta_m)$ will be large as μ_1^* approaches 0, resulting in imprecise estimates (Schaefer, 1986). Higher variances of the estimated coefficients will “deflate” asymptotic t-ratios or similar test statistics, resulting in low significance levels, as well.

Overall, due to the similarity between the logistic regression model and the multiple linear regression model, the problems caused by near-multicollinearity in the logistic regression model are implied to be similar to those in the multiple linear regression

model (see Menard, 2010). Therefore just like linear regression, the traditional account of near-multicollinearity in logistic regression must also be revisited, since the estimated βs are still functions of ρ . In other words, the **ceteris paribus** clause implicitly assumed by the traditional account of near-multicollinearity fails to hold for the logistic regression model, as well. However, the analysis is not directly applicable to the logistic regression model since the variance-covariance structure of the estimated parameters for the logistic regression model is rather different than that of the linear regression model. A convenient approach to examine near multicollinearity in logistic regression models is using the log-odds specification given by equation (2.2) (McCullagh and Nelder 1983 p. 19). This approach is best motivated using the Probabilistic Reduction approach to model specification.

2.4 The Probabilistic Reduction Approach and the Logistic Regression Model

The probabilistic reduction (PR) approach provides a flexible approach to model specification that fully accommodates empirical modeling using both experimental and non-experimental data. It makes use of the De Finetti representation theorem to formally reduce the **Haavelmo distribution** of the observable random covariates into a product of univariate marginal and conditional distributions by imposing certain probabilistic assumptions (Spanos, 1999). This approach to model specification provides a framework for learning about the underlying sample data generating process (DGP), thus improving upon the reliability of inferences, as well as their their accuracy. Spanos and McGuirk (2001) conclude that the PR approach to model specification offers a methodical framework for learning and modeling the actual DGP by viewing statistical models as a reduction from the Haavelmo (joint) distribution of all the observable

vector stochastic processes, hence presenting a consistent approach for specification, misspecification and respecification of statistical models.

As applied to the logistic regression model, the PR approach emphasizes the central role that the inverse conditional distribution of the explanatory variables conditioned on the nominal outcome variable of the logistic regression plays in model specification. Bergtold, Spanos and Onukwugha (2010) emphasize the key role that the inverse conditional distribution plays in providing relevant statistical information for specifying the logistic regression model. Unlike previous studies that have recognized the importance of the inverse conditional distribution in specifying and modeling the logistic regression model by assuming linearity in the variables for the predictor functions (See Andersen, 1972 and McFadden, 1976), the PR approach reparametizes the inverse conditional distributions for the logistic regression model by allowing for nonlinearity in the parameters, variables or both (Bergtold, Spanos and Onukwugha, 2010). This reparametization allows for the re-examination of multicollinearity in the logistic regression model by providing a mechanism to specify the parameters and related statistics of the logistic regression model in terms of the correlation between the explanatory variables .

2.4.1 A Re-examination of Near-Multicollinearity in Logistic Regression using the PR Approach

Following Bergtold, Spanos and Onukwugha (2010), for $i = 1, \dots, N$, let $Y_i \sim Bin(1, p)$ be a finite stochastic outcome process defined over the probability space $S, \mathcal{F}, P(\cdot)$ where $E(Y_i) = p$ and $Var(Y_i) = p(1-p)$. In addition, for $i = 1, \dots, N$, let $\mathbf{X}_i = (X_{1i}, \dots, X_{ki})'$, denote the vector of K stochastic observed covariates for the i^{th} individual defined over the same probability space. Assume that the probability density function of \mathbf{X}_i is given by $f_{\mathbf{X}}(\mathbf{X}_i; \theta)$, where θ is an appropriate set of parameters. S is called the support of

\mathbf{X}_i and Y_i , and \mathcal{F} is the borel field generated by (Y_i, \mathbf{X}_i) . The joint probability mass function can now be expressed as:

$$f(Y_1, \dots, Y_N, \mathbf{X}_1, \dots, \mathbf{X}_N; \phi) \quad (2.3)$$

where ϕ is an appropriate set of parameters.

As noted by Spanos (1999), equation (2.3) can reduced using a set of testable probabilistic assumptions from the categories: Distribution (D), Dependence (M), and Heterogeneity (H). Assume that (Y_i, \mathbf{X}_i) is Independent and Identically Distributed (*IID*). Imposing the probabilistic reduction assumptions mentioned above on equation (2.3) results in:

$$f(Y_1, \dots, Y_N, \mathbf{X}_1, \dots, \mathbf{X}_N; \phi) \stackrel{I}{=} \prod_{i=1}^N f_i(Y_i, \mathbf{X}_i; \varphi_i) \stackrel{ID}{=} \prod_{i=1}^N f(Y_i, \mathbf{X}_i; \varphi) = \prod_{i=1}^N f_{Y|X}(Y_i|\mathbf{X}_i; \beta) \cdot f_X(\mathbf{X}_i; \theta) \quad (2.4)$$

where $\varphi_i, \varphi, \beta$ and θ are appropriate sets of parameters and $f_{Y|X}(Y_i|\mathbf{X}_i; \beta)$ is the conditional distribution of Y_i given \mathbf{X}_i . Now let $f_{X|Y}(\mathbf{X}_i; \theta_j)$ for $Y_i = j$, be the inverse conditional distribution of \mathbf{X}_i given Y_i with appropriate parameter vector θ_j . The reduction in equation (2.4) provides an adequate operational paradigm for defining logistic regression models, where the adequacy of (2.4) is contingent upon the compatibility of $f_{Y|X}(Y_i|\mathbf{X}_i; \beta)$ and its inverse conditional. A sufficient condition for the compatibility of $f_{Y|X}(Y_i|\mathbf{X}_i; \beta)$ and $f_{X|Y}(\mathbf{X}_i; \theta_j)$, shown by Bergtold, Spanos and Onukwugha (2010) is:

$$\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i; \theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i; \theta_0)} \cdot \frac{f_Y(Y_i = 1; p)}{f_Y(Y_i = 0; p)} = \frac{f_{Y|X}(Y_i = 1|\mathbf{X}_i; \beta)}{f_{Y|X}(Y_i = 0|\mathbf{X}_i; \beta)} \quad (2.5)$$

Bergtold, Spanos and Onukwugha (2010) show that if $f_{Y|X}(Y_i|\mathbf{X}_i; \beta)$ is conditionally

distributed Bernoulli of the form:

$$f_{Y|X}(Y_i|\mathbf{X}_i; \beta) = h(\mathbf{X}_i; \beta)^{Y_i}[1 - h(\mathbf{X}_i; \beta)]^{1-Y_i} \quad (2.6)$$

where $h(\mathbf{X}_i; \beta) : \Theta_\beta \rightarrow [0, 1]$, and Θ_β is the parameter space associated with β , then equations (2.5) and (2.6) together yield

$$h(\mathbf{X}_i; \beta) = \frac{p \cdot f_{X|Y=1}(\mathbf{X}_i; \theta_1)}{p \cdot f_{X|Y=1}(\mathbf{X}_i; \theta_1) + (1-p) \cdot f_{X|Y=0}(\mathbf{X}_i; \theta_0)} \quad (2.7)$$

Using the identity that $f(\cdot) \equiv \exp\{\ln(f(\cdot))\}$ it can be shown that

$$h(\mathbf{X}_i; \beta) = \frac{\exp\{\eta(\mathbf{X}_i; \beta)\}}{1 + \exp\{\eta(\mathbf{X}_i; \beta)\}} = [1 + \exp\{-\eta(\mathbf{X}_i; \beta)\}]^{-1} \quad (2.8)$$

where $\eta(\mathbf{X}_i; \beta) = \ln\left(\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i; \theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i; \theta_0)}\right) + \kappa$, and $\kappa = \ln\left(\frac{p}{1-p}\right)$. The function $\eta(\cdot)$ is known as the predictor or index function. Equation (2.8) is usually reparametrized such that $\beta = \beta(\theta_j, j = 0, 1)$. That is, the parameters β are reparameterizations of the parameters θ_0 and θ_1 of the inverse conditional distribution.

Using the logistic density function $h(\mathbf{X}_i; \beta) = [1 + \exp\{-\eta(\mathbf{X}_i; \beta)\}]^{-1}$, the logistic regression function for the conditional stochastic process $Y_i|\mathbf{X}_i = \mathbf{x}_i, i = 1, \dots, N$ can be written as:

$$Y_i = h(\mathbf{X}_i; \beta) + u_i = [1 + \exp\{-\eta(\mathbf{X}_i; \beta)\}]^{-1} + u_i \quad (2.9)$$

To date, the literature examining multicollinearity in the the logistic regression model has not captured the fact that $\beta = \beta(\rho)$, where ρ is the correlation between the explanatory variables when $f_{\mathbf{X}|Y}(\mathbf{X}_i; \theta_j)$ is multivariate normal. From this vantage point, one can show that the traditional account of near-multicollinearity and the invocation of the *ceteris paribus* clause are unattainable. To see this, consider the simple

variant of (2.8) written as:

$$\eta(\mathbf{X}_i; \beta) = \ln \left(\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i; \theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i; \theta_0)} \right) + \ln \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u_i \quad (2.10)$$

As noted earlier, the index/predictor function can take many forms. That is, the logistic regression model does not make any assumptions regarding the distribution of the predictor or explanatory variables. Thus, the explanatory variables are not constrained to be normally distributed nor assumed to be linearly related. Furthermore, the inverse conditional distribution can have homogeneous or heterogeneous variances within each group, giving rise to nonlinear relationships between the explanatory variables (see Bergtold, Spanos and Onukwugha, 2010). To re-assess the impact of near-multicollinearity here, we consider the more traditional case here with a logistic regression model with explanatory variables that are normally distributed and have homogenous variances. In equation (2.10), the predictor is linear, resulting from the distributional assumption concerning the inverse conditional distribution

2.4.2 The Logistic Regression Model with Two Correlated Normal Covariates with Homogeneous Covariance Matrix

Assume that $\mathbf{X} = (x_1, x_2)'$ and

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} | Y_i = j \sim N \left[\begin{pmatrix} \mu_{1,j} \\ \mu_{2,j} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right]$$

and $\rho = \text{Corr}(x_1, x_2)$. The inverse conditional distribution function can be expressed as (see Spanos, 1986: p 119-121 and Bergtold et al., 2010):

$$\begin{aligned}
f_{X|Y=j}(x_1, x_2; \theta_j) &= \frac{(1-\rho^2)^{-\frac{1}{2}}}{2\pi\sigma_1\sigma_2} \cdot \exp\left\{-\frac{1}{2} \frac{1}{1-\rho^2} \left[\left(\frac{x_1-\mu_{1,j}}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_{2,j}}{\sigma_2}\right)^2 \right. \right. \\
&\quad \left. \left. - 2\rho \left(\frac{x_1-\mu_{1,j}}{\sigma_1}\right) \left(\frac{x_2-\mu_{2,j}}{\sigma_2}\right) \right] \right\} \\
&= \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{x_1^2-2x_1\mu_{1,j}+\mu_{1,j}^2}{\sigma_1^2} + \frac{x_2^2-2x_2\mu_{2,j}+\mu_{2,j}^2}{\sigma_2^2} \right. \right. \\
&\quad \left. \left. - 2\rho \frac{x_1x_2-x_1\mu_{2,j}-x_2\mu_{1,j}+\mu_{1,j}\mu_{2,j}}{\sigma_1\sigma_2} \right] \right\}
\end{aligned} \tag{2.11}$$

Define $\Delta_j = -\frac{1}{2(1-\rho^2)} \left[\frac{x_1^2-2x_1\mu_{1,j}+\mu_{1,j}^2}{\sigma_1^2} + \frac{x_2^2-2x_2\mu_{2,j}+\mu_{2,j}^2}{\sigma_2^2} - 2\rho \frac{x_1x_2-x_1\mu_{2,j}-x_2\mu_{1,j}+\mu_{1,j}\mu_{2,j}}{\sigma_1\sigma_2} \right]$, so that (2.11) becomes $f(x_1, x_2; \theta_j) = \exp\{\Delta_j\}$. Then:

$$\ln \frac{f_{\mathbf{X}|Y=1}(x_1, x_2; \theta_1)}{f_{\mathbf{X}|Y=0}(x_1, x_2; \theta_0)} = \ln \frac{\exp\{\Delta_1\}}{\exp\{\Delta_0\}} = \Delta_1 - \Delta_0 \tag{2.12}$$

Combining (2.11) and (2.12) for $j = 0, 1$, results in

$$\begin{aligned}
\Delta_1 - \Delta_0 &= -\frac{1}{2(1-\rho^2)} \left[\left(\frac{\mu_{1,1}^2}{\sigma_1^2} - \frac{2\rho\mu_{1,1}\mu_{2,1}}{\sigma_1\sigma_2} + \frac{\mu_{2,1}^2}{\sigma_2^2} \right) + \left(-\frac{2\mu_{1,1}}{\sigma_1^2} + \frac{2\rho\mu_{2,1}}{\sigma_1\sigma_2} \right) x_1 \right. \\
&\quad + \left(\frac{2\rho\mu_{1,1}}{\sigma_1\sigma_2} - \frac{2\mu_{2,1}}{\sigma_2^2} \right) x_2 - \left(\frac{\mu_{1,0}^2}{\sigma_1^2} - \frac{2\rho\mu_{1,0}\mu_{2,0}}{\sigma_1\sigma_2} + \frac{\mu_{2,0}^2}{\sigma_2^2} \right) + \left(-\frac{2\mu_{1,0}}{\sigma_1^2} + \frac{2\rho\mu_{2,0}}{\sigma_1\sigma_2} \right) x_1 \\
&\quad \left. - \left(\frac{2\rho\mu_{1,0}}{\sigma_1\sigma_2} - \frac{2\mu_{2,0}}{\sigma_2^2} \right) x_2 \right] \\
&= -\frac{1}{2(1-\rho^2)} \left[\left(\frac{\mu_{1,1}^2-\mu_{1,0}^2}{\sigma_1^2} - \frac{2\rho(\mu_{1,1}\mu_{2,1}-\mu_{1,0}\mu_{2,0})}{\sigma_1\sigma_2} + \frac{\mu_{2,1}^2-\mu_{2,0}^2}{\sigma_2^2} \right) \right. \\
&\quad \left. + \left(\frac{2\rho(\mu_{2,1}-\mu_{2,0})}{\sigma_1\sigma_2} - \frac{2(\mu_{1,1}-\mu_{1,0})}{\sigma_1^2} \right) x_1 + \left(\frac{2\rho(\mu_{1,1}-\mu_{1,0})}{\sigma_1\sigma_2} - \frac{2(\mu_{2,1}-\mu_{2,0})}{\sigma_2^2} \right) x_2 \right]
\end{aligned} \tag{2.13}$$

Using the relationships in equations (2.10) and (2.13), the corresponding logistic regression model can be written as :

$$Y_i = [1 + \exp\{-\beta_0 - \beta_1 x_1 - \beta_2 x_2\}]^{-1} + u_i \quad (2.14)$$

where

$$\beta_0 = \ln\left(\frac{p}{1-p}\right) - \frac{1}{2(1-\rho^2)} \left[\left(\frac{\mu_{1,1}^2 - \mu_{1,0}^2}{\sigma_1^2} - \frac{2\rho(\mu_{1,1}\mu_{2,1} - \mu_{1,0}\mu_{2,0})}{\sigma_1\sigma_2} + \frac{\mu_{2,1}^2 - \mu_{2,0}^2}{\sigma_2^2} \right) \right], \quad (2.15)$$

$$\beta_1 = \frac{1}{2(1-\rho^2)} \left[\frac{2(\mu_{2,1} - \mu_{2,0})}{\sigma_2^2} - \left(\frac{2\rho(\mu_{1,1} - \mu_{1,0})}{\sigma_1\sigma_2} \right) \right], \text{ and} \quad (2.16)$$

$$\beta_2 = \frac{1}{2(1-\rho^2)} \left[\frac{2(\mu_{2,1} - \mu_{2,0})}{\sigma_2^2} - \left(\frac{2\rho(\mu_{1,1} - \mu_{1,0})}{\sigma_1\sigma_2} \right) \right] \quad (2.17)$$

Clearly the parameters $\{\beta_0, \beta_1, \beta_2\}$ are all functions of ρ and thus change as $|\rho| \rightarrow 1$. It is difficult to mathematically show the magnitude and direction of change for $\beta_k, k = 0, 1, 2$ as $|\rho| \rightarrow 1$, because ρ appears in both the numerator and the denominator of β_i . For such an exercise, one would have to know which, the numerator or the denominator, grows faster, which is somewhat dependent on the other parameters of the inverse conditional distribution, as well. The behavior of β_0, β_1 and β_2 is therefore examined by simulating the above model and analyzing the dynamics of β_0, β_1 and β_2 as $|\rho| \rightarrow 1$.

2.4.3 The Variance and Standard Error of β

Schaefer, Roi and Wolfe (1984) and Gourieroux (2000) propose methods for deriving the variance of the parameter estimates for the logistic regression model. These sources derive the $Var(\beta_i)$ using a method reminiscent of the generalized least squares estimator.

Let

$$P(y_i = 1|\mathbf{X}_i; \beta) = p_i = F(\beta' \mathbf{X}_i) = F_i$$

Gourieroux (2000) shows that the asymptotic variance of β can be written as¹

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \left\{ \sum_{i=1}^n \frac{[f(\hat{\beta}' \mathbf{X}_i)]^2}{F(\hat{\beta}' \mathbf{X}_i) [1 - F(\hat{\beta}' \mathbf{X}_i)]} \mathbf{X}_i' \mathbf{X}_i \right\}^{-1} \\ &= (\mathbf{X}' \hat{\mathbf{V}} \mathbf{X})^{-1} \end{aligned} \quad (2.18)$$

where $\hat{\mathbf{V}} = \text{diag}[F_i(1 - F_i)]_{i=1}^n$, $f = F(1 - F)$, and \mathbf{X} is the $n \times k$ dimensional matrix of predictors, and \mathbf{X}_i is the i^{th} row of \mathbf{X} . Schaefer, Roi and Wolfe (1984) comment on the effect of multicollinearity on $\text{Var}(\beta_i)$ that “in practice, we have found that multicollinearity does indeed inflate the estimated variance and hence can cause precision problems when identifying the effects of the independent variables”.

However, what has not been adequately pointed out by the traditional account of near-multicollinearity is that as shown above, $\text{Var}(\hat{\beta}_i)$ is a function of $F(x_i \beta)$ and thus ρ . That is, since β is a function of ρ (see equations (14) to (17)), as $|\rho| \rightarrow 1$, β changes, and so too does the $\text{Var}(\hat{\beta}_i)$. As $|\rho| \rightarrow 1$, if $|\beta| \rightarrow \infty$, then F will approach 0 or 1, depending on the sign of β . This implies that $F(1 - F) \rightarrow 0$ as $|\beta| \rightarrow \infty$. Examining equation (18), this suggests that the $\text{Var}(\hat{\beta}_i)$ will likely follow the shifts in β as $|\rho| \rightarrow 1$, since $f_i^2/[F_i(1 - F_i)] = f_i$. Simulations in the next section provide evidence of this occurring, but it is not uniformly the case. This therefore calls into question the traditional account that argues that the presence of near-multicollinearity inflates the variances and standard errors of the estimators.

¹See Gourieroux (2000), page 16 for complete derivation of this variance.

2.5 Simulation and Graphics

This section of the paper presents simulation methods and results examining the simple case presented in section 2.4.2 of a logistic regression model specified using a bivariate normal inverse conditional distribution with homogenous covariance matrix. Section 2.5.1 presents some model simplifications that make the simulations smoother without losing generality. Section 2.5.2 examines simulation methods. Section 2.5.3 presents the results of the simulations using graphs of the estimated parameters and statistics.

2.5.1 Simplifications

The simplifications below substantially reduce the dimensionality of the problem being simulated without necessarily limiting the generalizability of the analysis in any vital way (e.g it may affect the scaling of the results).

2.5.1.1 Admissible Parameter Values

In light of the statistical reparametizations using the PR approach shown above, it is important to determine the range of admissible values for which the model exists. The importance of this is emphasized by Spanos and McGuirk (2002). Let Σ denote the the variance-covariance matrix of the inverse conditional distribution, then $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. Following this parametrization, the above parameters $\{\beta_0, \beta_1, \beta_2\}$ are valid when (Spanos, 1986):

$$\det(\Sigma) = \sigma_1^2\sigma_2^2(1 - \rho^2) > 0 \tag{2.19}$$

Condition (2.19) states that the inverse conditional distribution given by equation (11) will not exist when $\Sigma \not\prec 0$. Thus, **during simulations if $\det(\Sigma) = 0$ then zeros**

are assigned for all the parameter values and associated statistics for that simulation run.

2.5.1.2 Simplification 1

Standardize the variables x_1 and x_2 by dividing them by σ_k^2 for $k = 1, 2$, so that $\sigma_1^2 = \sigma_2^2 = 1$. It follows that equation (2.19) becomes $\det(\Sigma) = (1 - \rho^2) > 0$. Thus, the admissible parameter range for ρ is $\rho \in (-1, 1)$. Following this simplification, the parameters $\{\beta_0, \beta_1, \beta_2\}$ from equation (2.19) reduce to:

$$\begin{aligned}\beta_0 &= \ln\left(\frac{p}{1-p}\right) - \frac{1}{2(1-\rho^2)} [\mu_{1,1}^2 - \mu_{1,0}^2 - 2\rho(\mu_{1,1}\mu_{2,1} - \mu_{1,0}\mu_{2,0}) + \mu_{2,1}^2 - \mu_{2,0}^2] \\ \beta_1 &= \frac{1}{2(1-\rho^2)} [2(\mu_{1,1} - \mu_{1,0}) - 2\rho(\mu_{2,1} - \mu_{2,0})] \\ \beta_2 &= \frac{1}{2(1-\rho^2)} [2(\mu_{2,1} - \mu_{2,0}) - 2\rho(\mu_{1,1} - \mu_{1,0})]\end{aligned}\tag{2.20}$$

2.5.1.3 Simplification 2

Further standardize the variables x_1 and x_2 by taking the following mean deviation form: $\bar{x}_1 = x_1 - \mu_{1,0}$ and $\bar{x}_2 = x_2 - \mu_{2,0}$, where $\mu_{k,0} = E(x_k/Y = 0)$. Then $\bar{\mu}_{k,0} = E(\bar{x}_k/Y = 0) = 0$ for $k = 1, 2$. Now let $\bar{\mu}_{k,1} = E(\bar{x}_k/Y = 1) = \alpha_k$. Then $\bar{\mu}_{k,1} = \alpha_k = \mu_{k,1} - \mu_{k,0}$ for $k = 1, 2$. This further reduces the dimensionality of the parameters $\{\beta_0, \beta_1, \beta_2\}$ from equation(2.20) to:

$$\begin{aligned}\beta_0 &= \ln\left(\frac{p}{1-p}\right) - \frac{1}{2(1-\rho^2)} [\bar{\mu}_{1,1}^2 - 2\rho(\bar{\mu}_{1,1}\bar{\mu}_{2,1}) + \bar{\mu}_{2,1}^2] \\ \beta_1 &= \frac{1}{2(1-\rho^2)} [2\bar{\mu}_{1,1} - 2\rho\bar{\mu}_{2,1}] \\ \beta_2 &= \frac{1}{2(1-\rho^2)} [2\bar{\mu}_{2,1} - 2\rho\bar{\mu}_{1,1}]\end{aligned}\tag{2.21}$$

2.5.2 Simulation Design

We use the PR approach described above to simulate binary choice data using the relation specified in equation (2.10). Equation (2.10) expresses the log-odds as a linear combination of the explanatory variables \mathbf{X}_i . As shown in equations (2.15) to (2.17), the coefficients, β_k depend on the value of the means, $\mu_{k,j}$ and variances, $\sigma_{k,j}^2$ of the conditional distribution $f_{X/Y}(\mathbf{X}_i; \theta_j)$. *Simplifications* 1 and 2 reduce the dimensionality of β_k even further, such that β s now only depend on $\bar{\mu}_{k,j}$, p and ρ . The trajectories of β_0, β_1 and β_2 as $|\rho| \rightarrow 1$ can be easily obtained by plotting their surfaces from equation (2.21) as $|\rho| \rightarrow 1$ for chosen values of μ_{11} and μ_{21} ; however, we chose to simulate for two important reasons:

1. Firstly, while it is possible to derive and plot the dynamics of β_i , $i = 0, 1, 2$ as $|\rho| \rightarrow 1$, no closed form solutions exist for their corresponding standard errors, t – ratios and related statistics. Most of these statistics are functions of the observed data and therefore cannot be readily derived in terms of the parameters of the inverse conditional distribution.
2. Simulating using the PR approach to generate binary choice data is advantageous because the PR approach is purely statistical, and hence requires no *a priori* theoretical assumptions. It therefore allows for the parsimonious description, specification, and simulation of models with near-multicollinearity and nonlinearities (Bergtold, Spanos and Onukwugha, 2010).

The data generation process takes place in two stages. **First**, using a binomial random number generator, we generate realizations of the vector stochastic process $\{Y_i, i = 1, \dots, N\}$. **Second**, using Y_i as the conditioning variable, the vector stochastic process of predictors $\{\mathbf{X}_i, i = 1, \dots, N\}$ is generated from the inverse conditional distributions $f_{X/Y}(\mathbf{X}_i; \theta_j)$ using appropriate random number generators. Simulations follow

the methods described in Bergtold, Spanos and Onukwugha (2010) and Scrucca and Weisberg (2004).

Given simplifications 1 and 2, Monte Carlo simulations are conducted for different mean pairs $(\bar{\mu}_{11}, \bar{\mu}_{21}) = (0.5, 0.5), (2, 2), (0.5, 1), (-0.5, -0.5), (-2, -2), (0.5, -2), (1, -1), (1, -2),$ and $(2, -2)$, as well as for $p = 0.6$ and 0.95 , where $p = P(y_i = 1 | \mathbf{X}_i; \beta)$. Different values of p allow for the examination of the impacts, if any, of changing the likelihood of “success”. The Monte Carlo simulations involved 400 runs, each run replicated 200 times for all mean pair and p combinations. For each run, ρ was varied between -1 and 1 by increments of 0.005 for each run. The max and min values for ρ were 0.995 and -0.995, respectively. We calculate the mean of $\beta_0, \beta_1, \beta_2$, their associated standard errors (denoted $se(\beta_0), se(\beta_1), se(\beta_2)$), t – ratios ($\tau(\beta_0), \tau(\beta_1), \tau(\beta_2)$) and McFadden Pseudo- R^2 across replications for each run. We also estimate the mean of near-multicollinearity diagnostic statistics, including variance inflation factors ($VIF(x_1), VIF(x_2)$) and condition number (κ) across replications for each run. Marginal effects and their associated p – values when severe near-multicollinearity is present in the logistic regression model has not received much attention in the literature. Thus these statistics are estimated, and we calculate the means of these statistics across replications for each run. Graphs of mean parameter estimates and statistics across replications are developed for each mean pair and p combination as functions of ρ . All simulations and graphics were done in MATLAB. The results are shown in Figures 2.2 through 2.31 in section 2.5.3.

2.5.3 Graphics

2.5.3.1 The β_0 surface with varying ρ and different mean vectors of $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$

In view of the fact that following simplifications 1 and 2, β_i s are functions of ρ , $\bar{\mu}_{1,1}$ and $\bar{\mu}_{2,1}$, the β_i surfaces in equation (2.19) represent three-dimensional surfaces, and any attempt to visualize them requires four dimensions. To make use of three-dimensional graphs, one must keep one of the arguments constant. For our purposes, we keep $\bar{\mu}_{1,1}$ and p constant (at $\bar{\mu}_{1,1} = 2$ and $p = 0.6$) and vary $\bar{\mu}_{2,1}$ between -10 to 10. Again, it is important to stress that these only serve to reduce dimensionality without necessarily affecting the analysis in any crucial way. Figure 2.1 shows the β_0 surface.

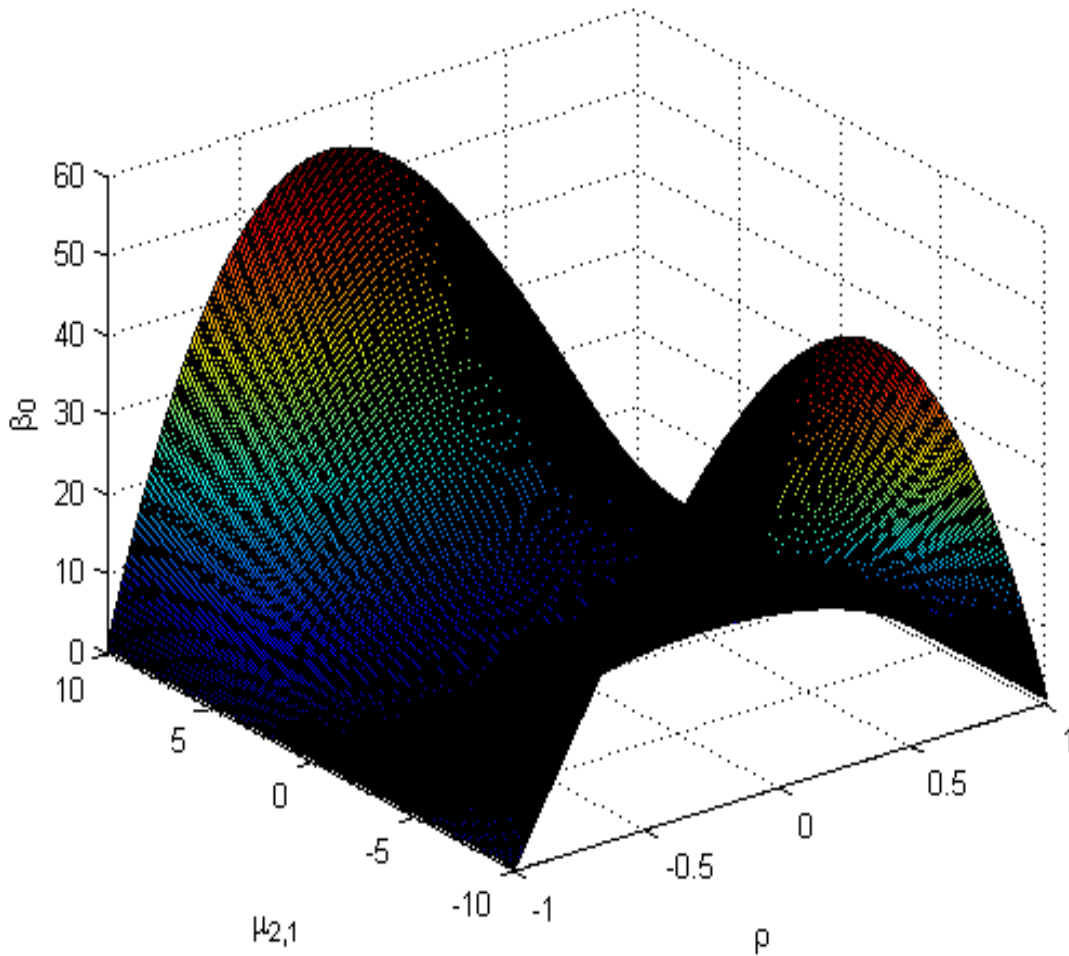
A glimpse of Figure 2.1 shows that the β_0 surface consists of two wide swings as $|\bar{\mu}_{2,1}| \rightarrow 10$. It is important to note that the choice of $\bar{\mu}_{1,1} = 2$ and $p = 0.6$ is arbitrary. Several values of $\bar{\mu}_{1,1}$ and p were considered but did not affect the general shape of β_0 in any significant way. Since our goal is to assess the impact of near-multicollinearity on β_0 , we consider in more detail the dynamics of the β_0 surface as $|\rho| \rightarrow 1$. These are shown in Figures 2.2 and 2.3.

In order to utilize two-dimensional graphs, without loss of generality, we further hold $\bar{\mu}_{1,1}$ and $\bar{\mu}_{2,1}$ constant. Figures 2.2 and 2.3 show the dynamics of β_0 for $p = 0.6$ and $p = 0.95$ respectively, estimated for nine different mean pairs of $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$. From these figures, it is clear that the choice of $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$ greatly affects the shape and dynamics of β_0 as $|\rho| \rightarrow 1$. When $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$ is positive, say $(0.5, 0.5)$ or $(2, 2)$, β_0 is in general a downward-sloping function of ρ . When $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$ is negative, say $(-0.5, -0.5)$, or $(-2, -2)$, β_0 in general slopes upward as ρ increases from -1 to 1.

When $\bar{\mu}_{1,1}$ and $\bar{\mu}_{2,1}$ have different signs, such as in panels *f*, *g*, *h*, and *i* in Figures 2.2 and 2.3, there is no distinct pattern of the dynamics of β_0 as ρ changes. Equally important to note are the (long) vertical swings that characterize β_0 at some points

where $|\rho|$ gets very close to 1. A possible explanation for this is that these are points

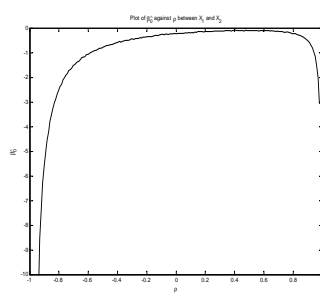
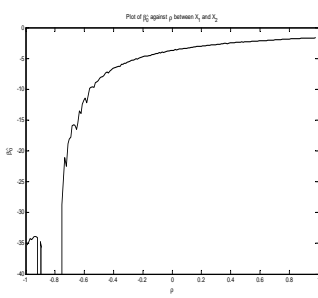
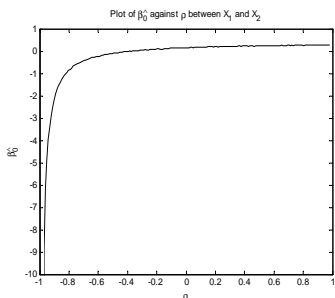
Figure 2.1: The β_0 surface with varying ρ , and varying $\bar{\mu}_{2,1}(\bar{\mu}_{1,1} = 2)$



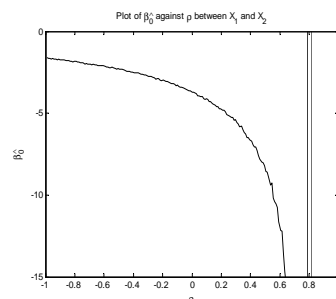
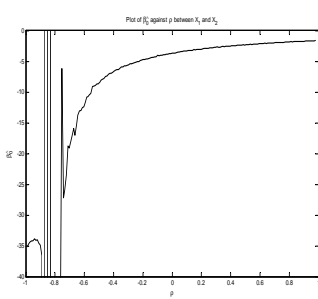
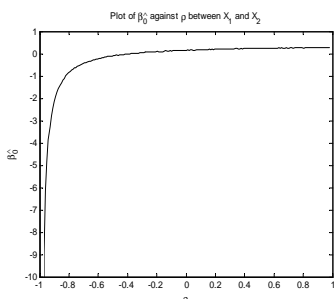
of numerical instability in the data (i.e. erratic volatility), not necessarily near-multicollinearity that arises from systematic volatility. The more important message from Figures 2.2 and 2.3 is that in contrast to the traditional account of near-multicollinearity that invokes the **ceteris paribus** clause - that β_0 remains constant as $|\rho| \rightarrow 1$ - our account suggests that β_0 fluctuates in a non-monotonic way as $|\rho| \rightarrow 1$.

Figure 2.2: β_0 with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$

- (a) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 0.5)$ (b) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, 2)$ (c) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 1)$



- (d) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-0.5, -0.5)$ (e) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-2, -2)$ (f) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, -2)$



- (g) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -1)$ (h) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -2)$ (i) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, -2)$

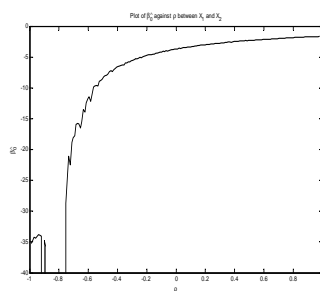
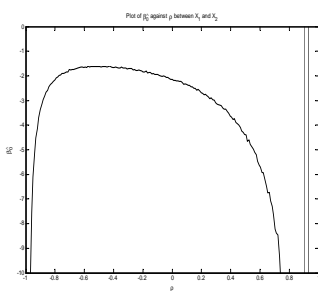
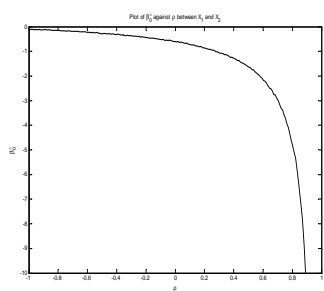
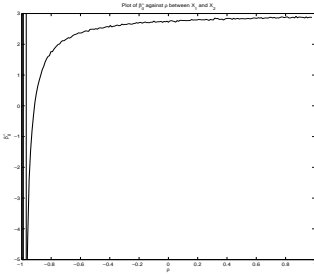
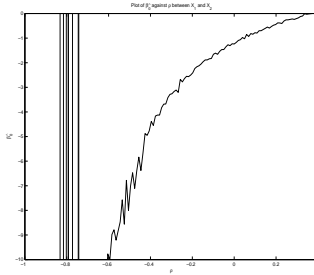


Figure 2.3: β_0 with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$

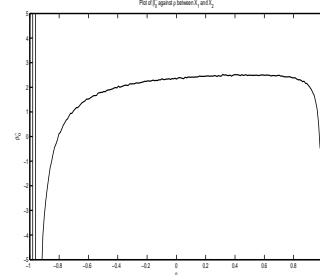
(a) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 0.5)$



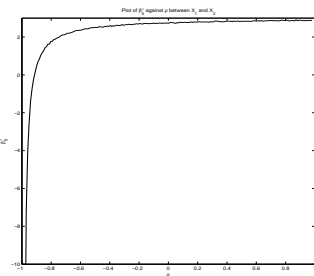
(b) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, 2)$



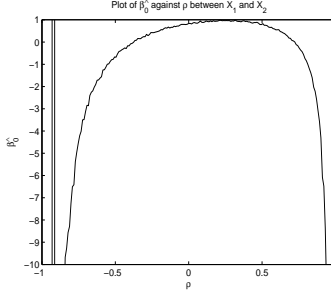
(c) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 1)$



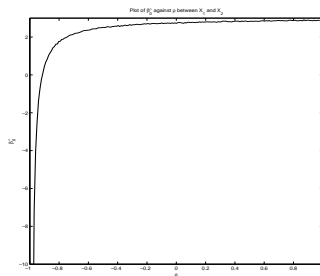
(d) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-0.5, -0.5)$



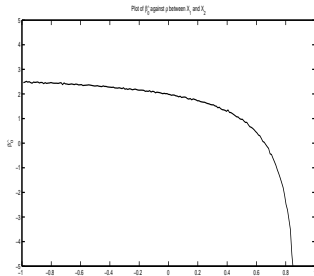
(e) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-2, -2)$



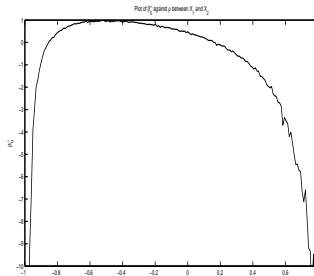
(f) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, -2)$



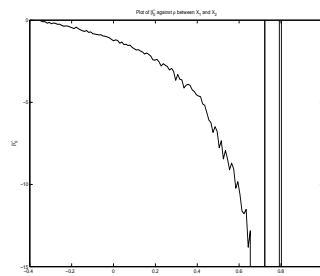
(g) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -1)$



(h) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -2)$



(i) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, -2)$



2.5.3.2 The β_1 Surface with varying ρ and different mean pairs (μ_{11}, μ_{21})

In Figures 2.4, 2.5 and 2.6 we can see the dynamics of the surface of β_1 , where

$$\beta_1 = \frac{1}{2(1 - \rho^2)} [2\bar{\mu}_{1,1} - 2\rho\bar{\mu}_{2,1}], \quad \rho \in (-1, 1), \quad p = 0.60, 0.95$$

Like the β_0 surface, points of systematic volatility for the β_1 surface are shown in Figure 2.4. There is no clear pattern of β_1 as $|\bar{\mu}_{2,1}| \rightarrow 10$. It turns out, as was the case with the β_0 surface that the choice of $\bar{\mu}_{2,1}$ affects the value of the β_1 as well as the idiosyncrasy of its surface. Despite this, it must be noted again that fixing $\bar{\mu}_{1,1}$ at 2 does not impact the general shape of the β_1 surface. In Figures 2.5 and 2.6, we examine the dynamics of β_1 as $|\rho| \rightarrow 1$.

Figure 2.4: The β_1 Surface with varying ρ and varying $\mu_{2,1}$ ($\mu_{1,1} = 2$), $p = 0.6$

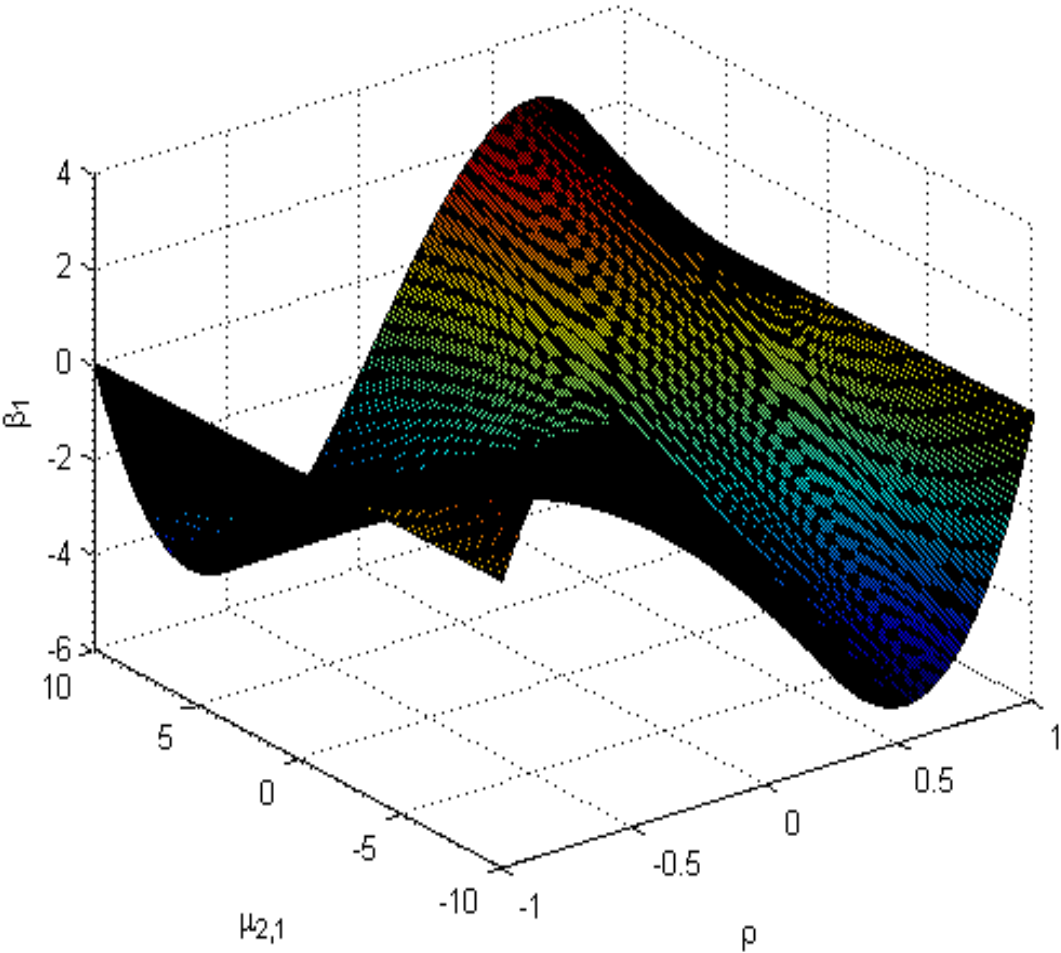
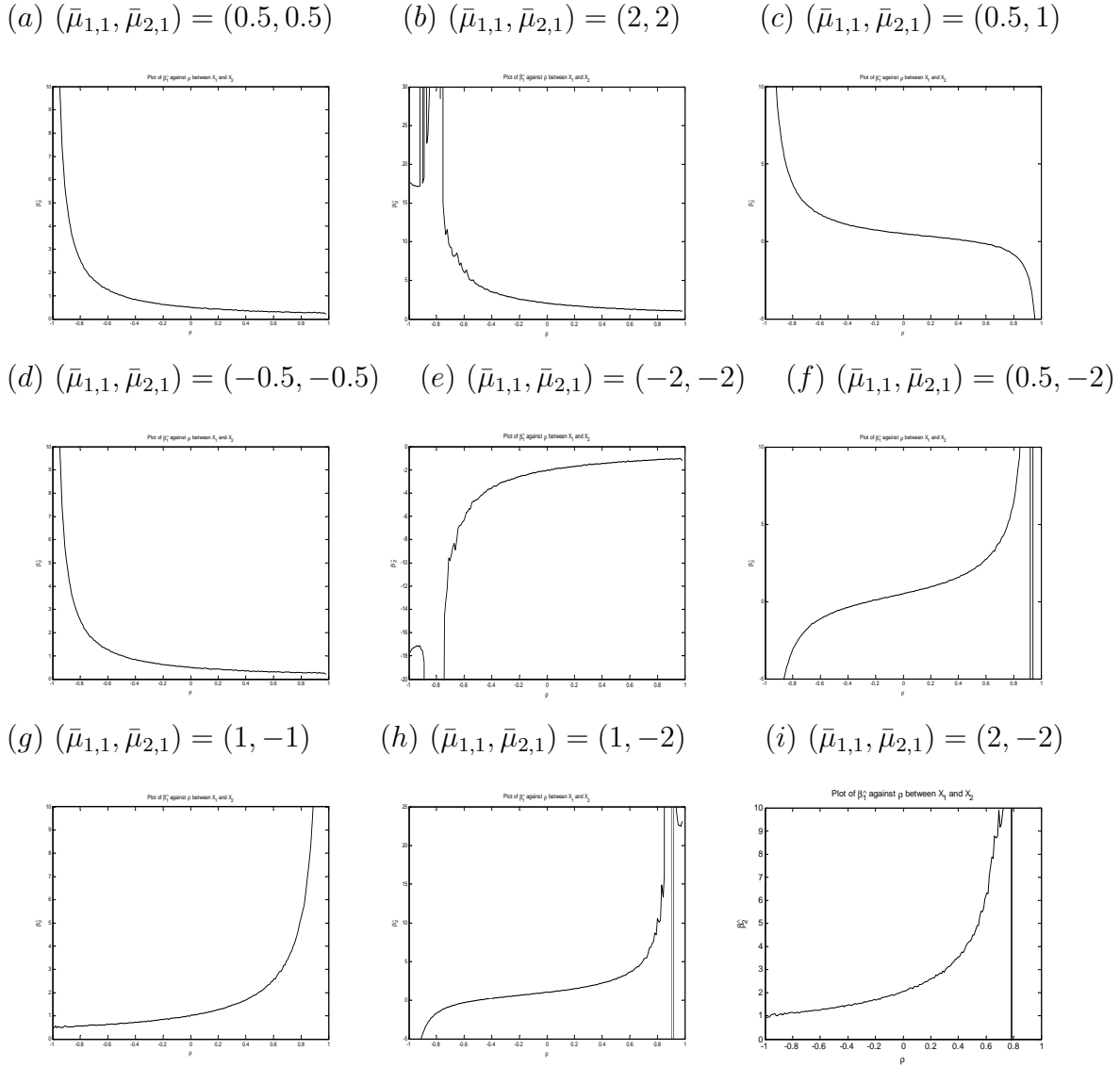
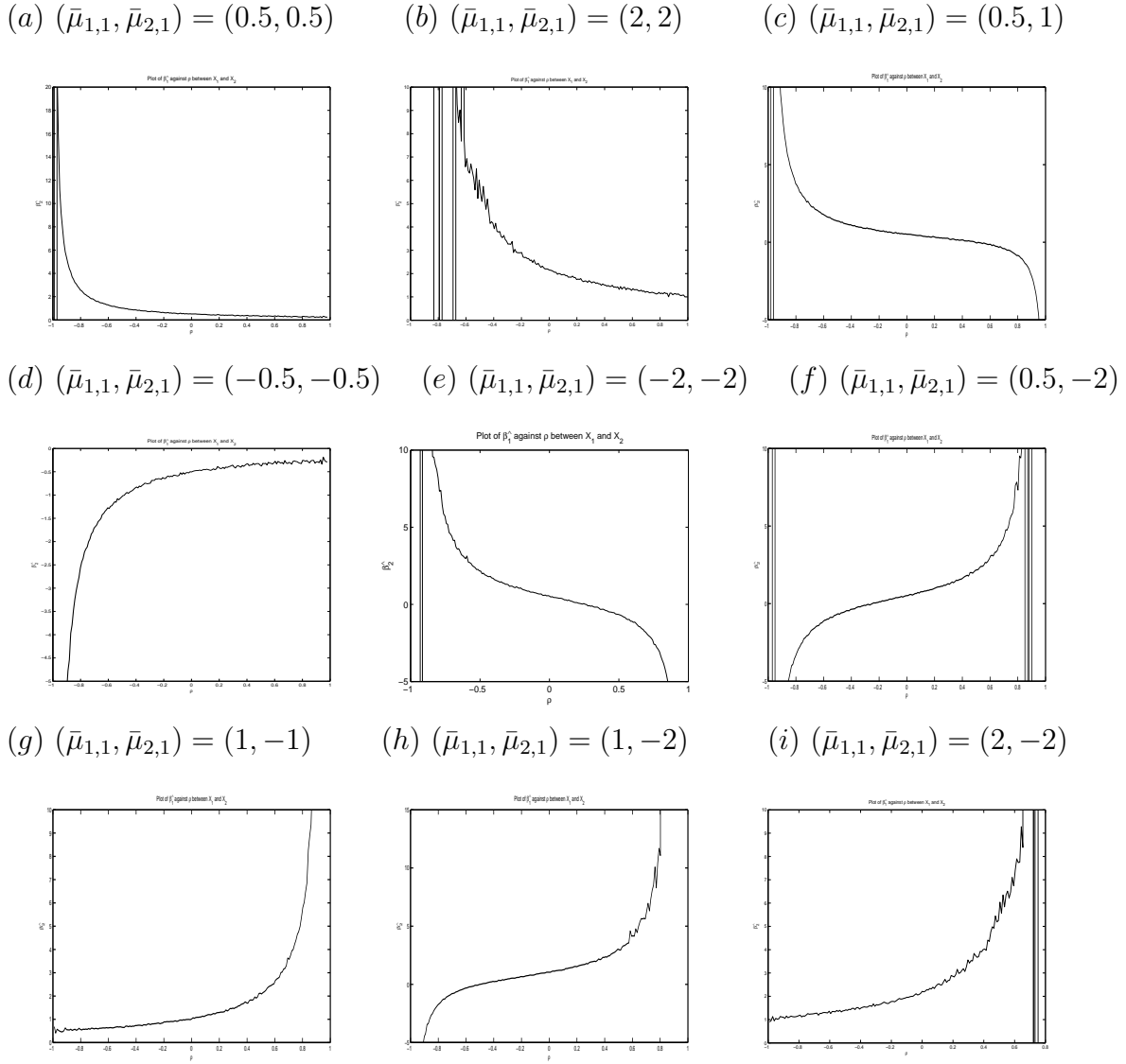


Figure 2.5: β_1 with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$



Similar to Figures 2.2 and 2.3, we see that the β_1 surface varies non-monotonically as ρ changes. The ceteris paribus clause is unattainable since β_1 varies with $|\rho| \rightarrow 1$. In addition, even though we are modelling a structural issue - high correlation among the regressors - which results in *systematic volatility*, we still find evidence of numerical instability, evident in long vertical swings at points where $|\rho|$ approaches 1 at the limit.

Figure 2.6: β_1 with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$



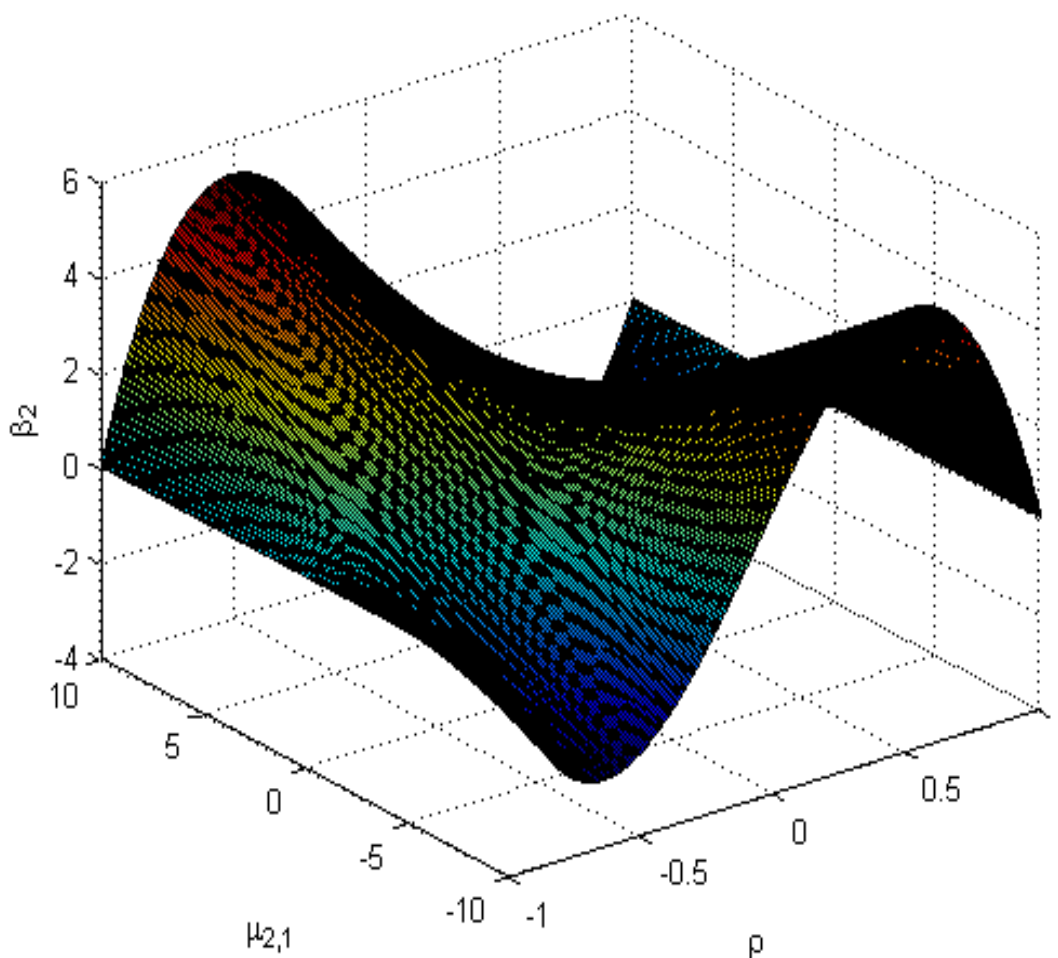
2.5.3.3 The β_2 Surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$

Similar to the β_0 and β_1 surfaces, points of systematic volatility in the β_2 surface:

$$\beta_2 = \frac{1}{2(1 - \rho^2)} [2\bar{\mu}_{2,1} - 2\rho\bar{\mu}_{1,1}] \quad \rho \in (-1, 1), \quad p = 0.6, 0.95$$

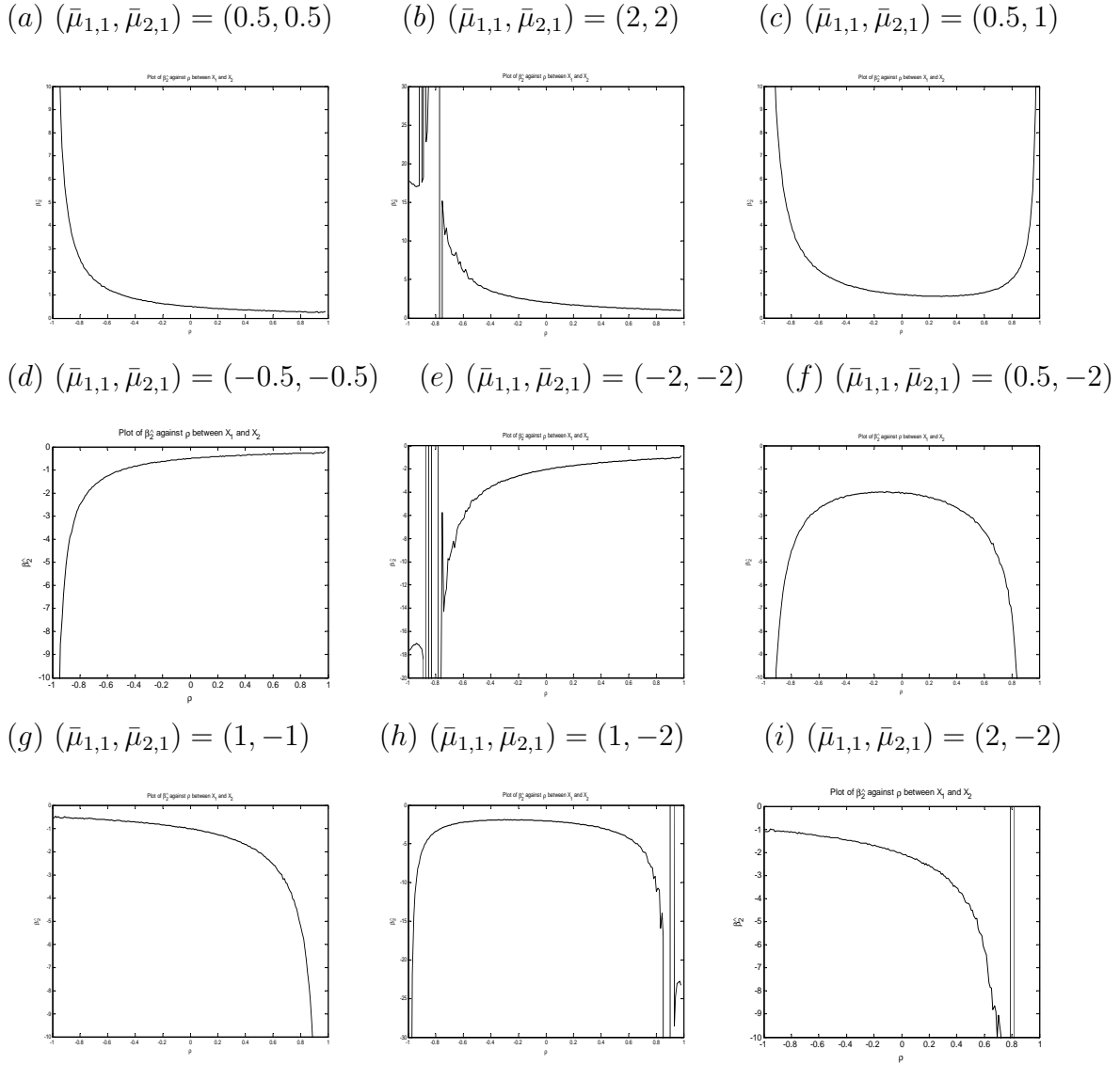
are shown in Figures 2.7, 2.8 and 2.9.

Figure 2.7: The β_2 surface with varying ρ and varying $\bar{\mu}_{2,1}$ ($\bar{\mu}_{1,1} = 2$)



The same information derived from the surfaces of β_0 and β_1 also apply here. β_2 varies non-monotonically as $|\rho| \rightarrow 1$, lending support to the view that the ceteris paribus clause to near-multicollinearity is unattainable. Like in Figures 2.2 through 2.6 we still see evidence of numerical instability in the data shown by the vertical swings at high values of ρ in Figures 2.8 and 2.9. It must be stressed that even though these swings are evidence of numerical instability in the data, it is not necessarily predictable. These swings however do not show up in Figures 2.1, 2.4 and 2.7. This is because the data used for these figures are not simulated.

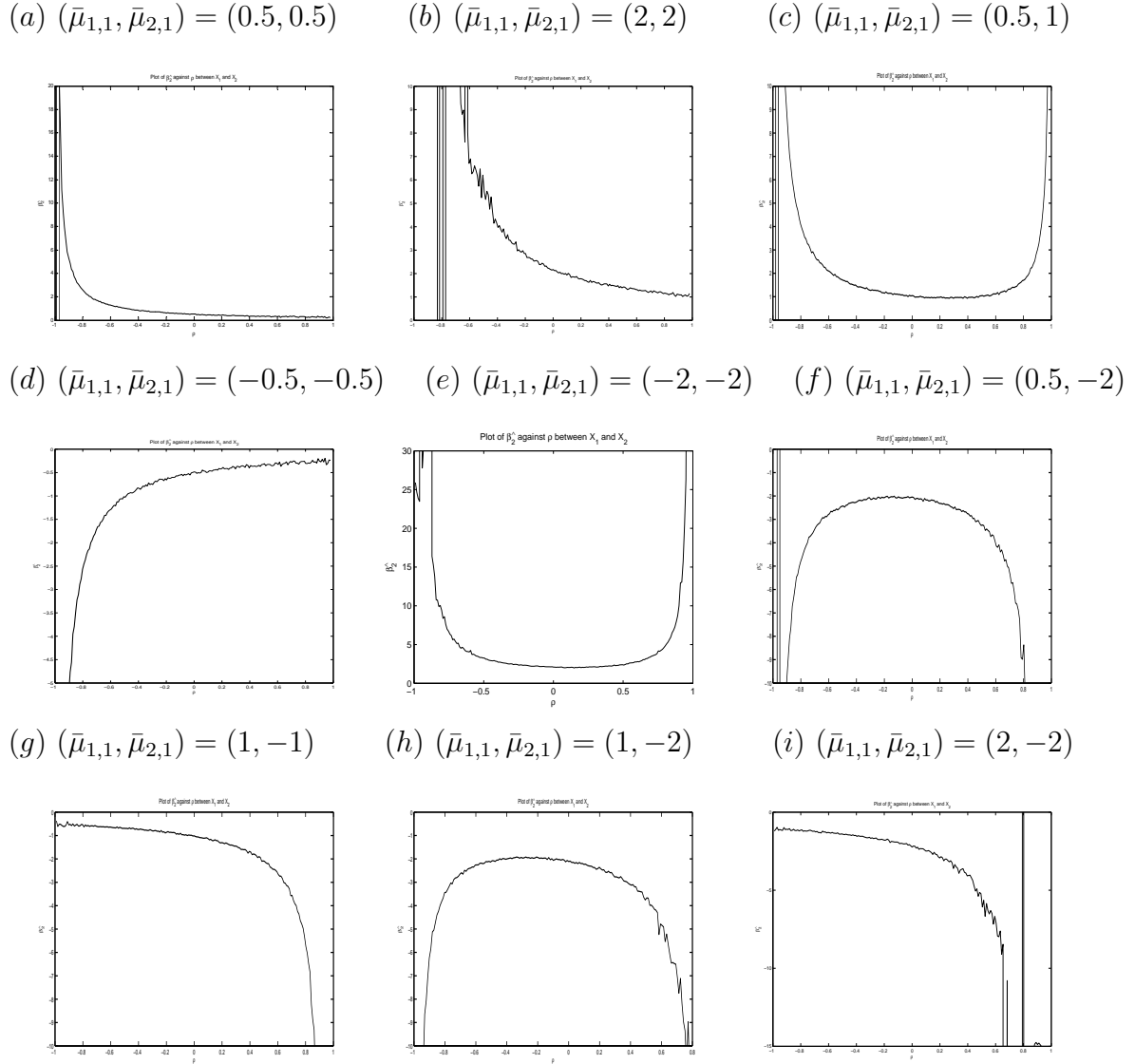
Figure 2.8: β_2 with Varying ρ and different mean pairs (μ_{1i}, μ_{2i}) : $p = 0.6$



Several interesting observations can be made from the slices of the β_2 surface shown in Figures 2.8 and 2.9. Firstly, for $(\mu_{11}, \mu_{21}) = (-0.5, -0.5)$, $(-2, -2)$, $(0.5, -2)$, $(1, -1)$, $(1, -2)$, and $(2, -2)$, $\beta_2 < 0$. In other words, when one or both of the means are negative, β_2 is always negative. Also, notice the hump-shaped graphs in Figures 2.8 and 2.9 (panels *f* and *h*) and the U-shape curves of panel *c*. This implies that near-multicollinearity may cause the coefficients to collapse as $|\rho| \rightarrow 1$ (as in Figures 2.8 and

2.9; f and g), or inflate (2.8 and 2.9 c).

Figure 2.9: β_2 varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$

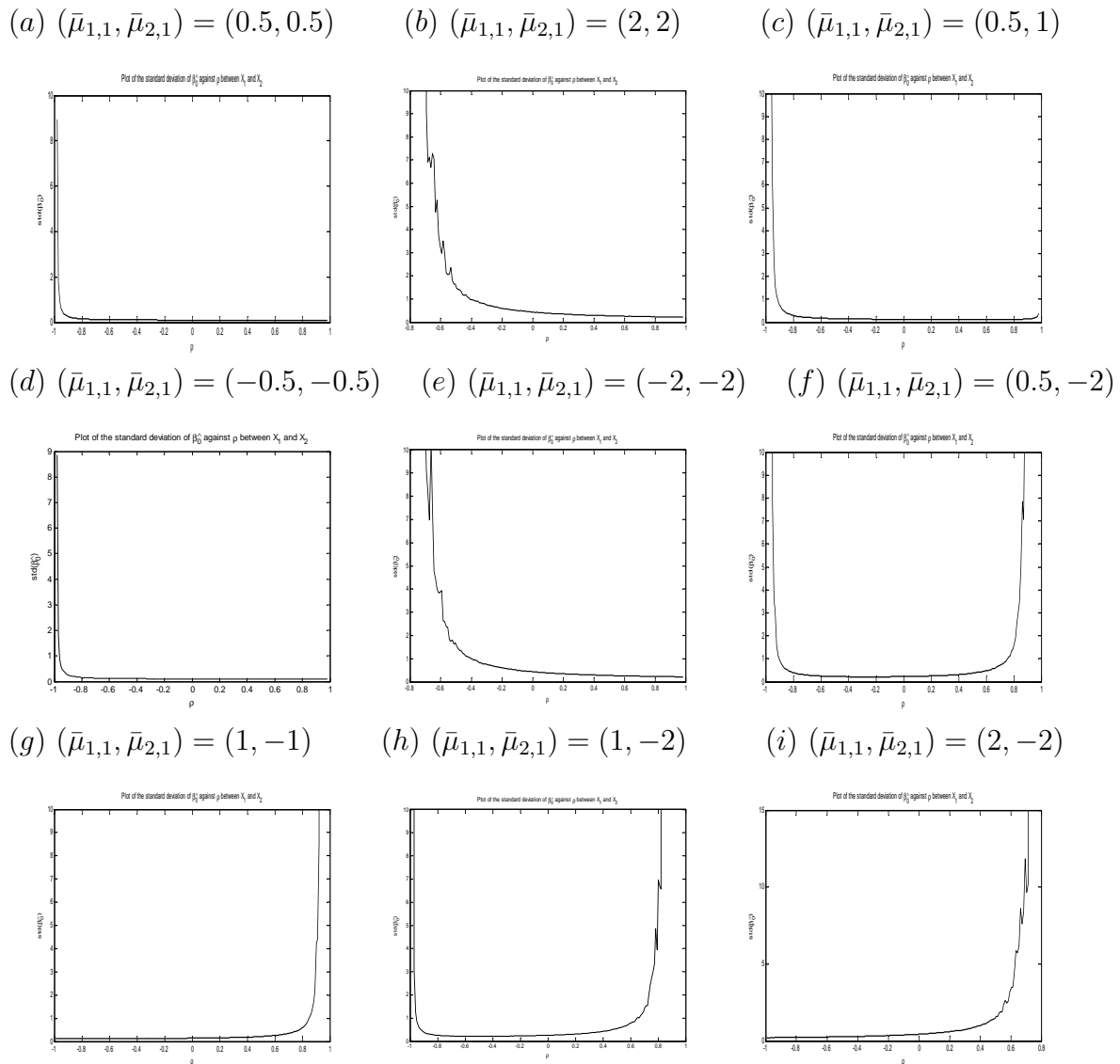


The overall finding derived from Figures 2.1 through 2.9 is that β_0 , β_1 and β_2 are all functions of ρ and thus change as ρ changes. Consequently, assuming constancy of β_0 , β_1 and β_2 as is the case in the traditional account of near-multicollinearity is misleading, hence the traditional account needs to be ammended.

2.5.3.4 The standard error of β_0 surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$

Figures 2.10 and 2.11 show how the standard error of β_0 changes as $|\rho| \rightarrow 1$. The conventional story about near-multicollinearity suggests that when severe near-multicollinearity occurs, the standard errors tend to be very large. This implies one should expect a

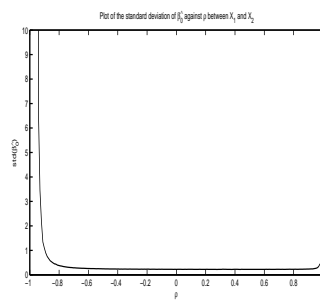
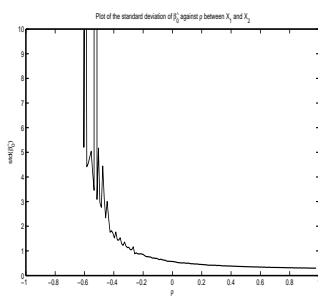
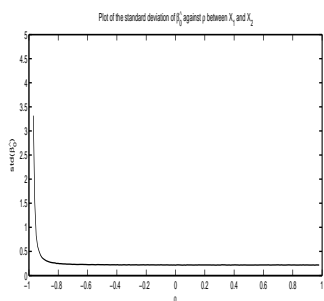
Figure 2.10: The $se(\beta_0)$ with varying ρ and different pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$



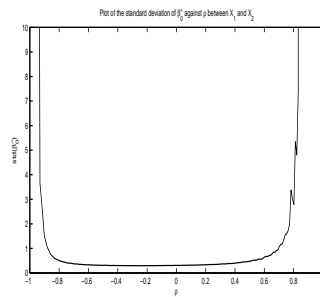
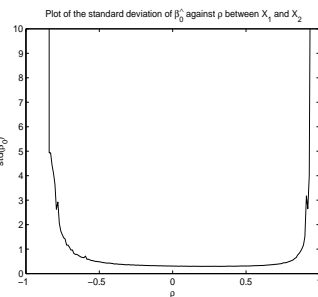
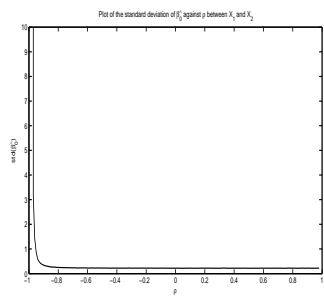
U-shaped graph centered around $\rho = 0$. However, the figures suggest that the standard errors of β_0 , $se(\beta_0)$ do not always follow the conventional account. Our results for the most part are in agreement with the discussion in section 2.4.3: we see that as β_0 increases, the $se(\beta_0)$ increases as well, even though this pattern is by no means uniform.

Figure 2.11: The $se(\beta_0)$ with varying ρ and different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$

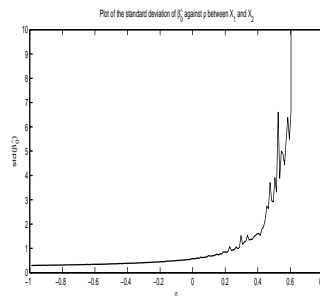
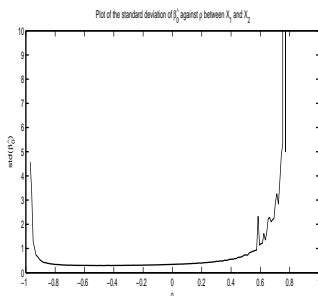
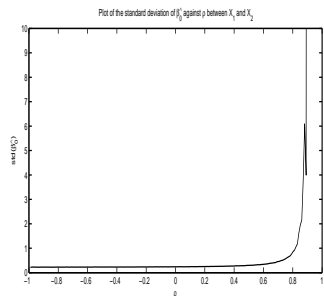
- (a) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 0.5)$ (b) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, 2)$ (c) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 1)$



- (d) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-0.5, -0.5)$ (e) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-2, -2)$ (f) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, -2)$



- (g) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -1)$ (h) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -2)$ (i) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, -2)$



2.5.3.5 The standard error of the β_1 Surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$

The standard errors of β_1 can be seen in Figures 2.12 and 2.13. Analogous to Figures 2.10 and 2.11, these standard errors look nothing like what the usual account suggests. Only panels *f* and *h* of both figures have the typical U shape as $|\rho| \rightarrow 1$ that the usual account suggests.

Figure 2.12: The $se(\beta_1)$ with varying ρ for different (μ_{1i}, μ_{2i}) : $p = 0.6$

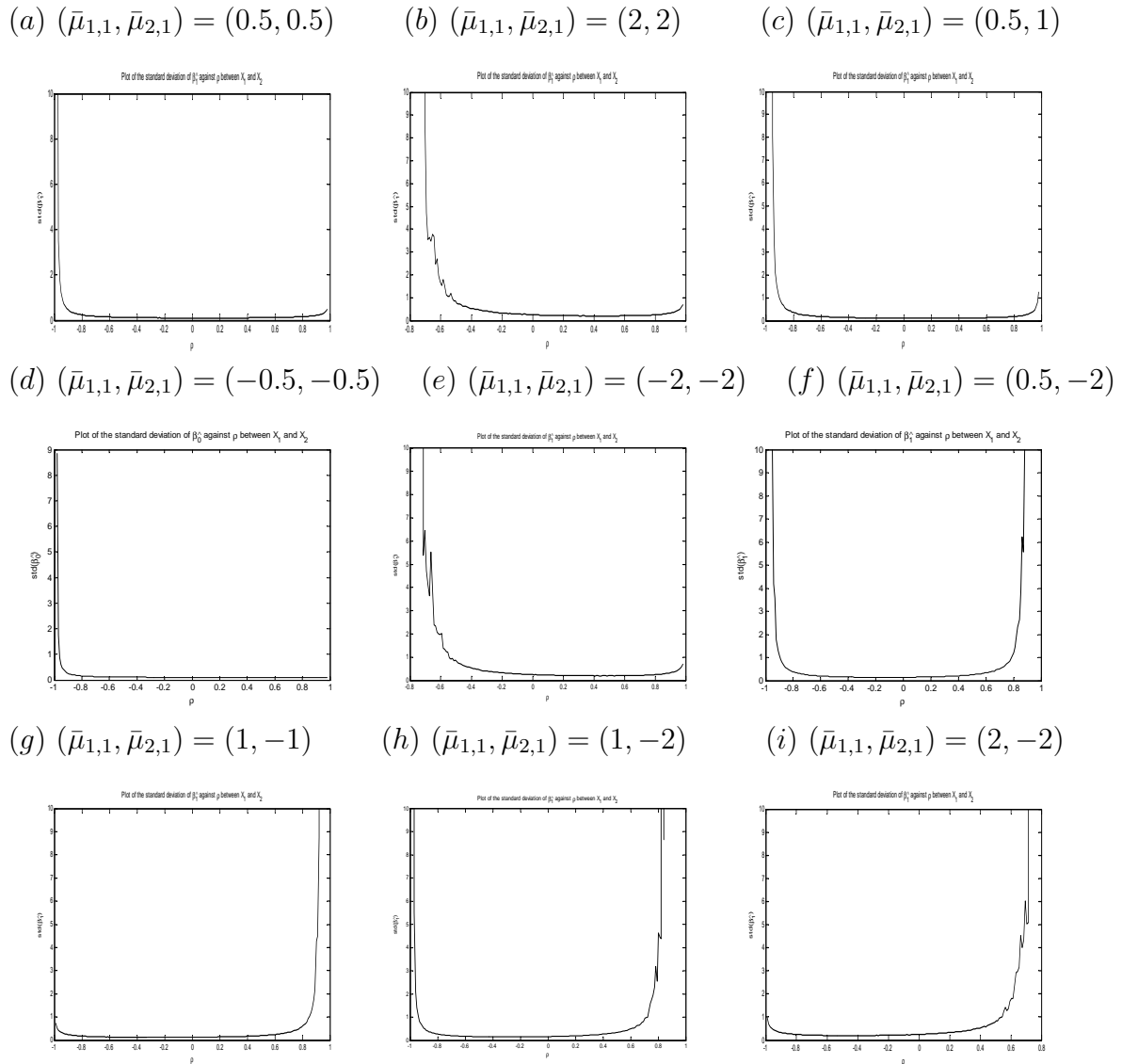
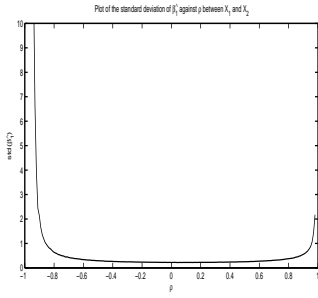
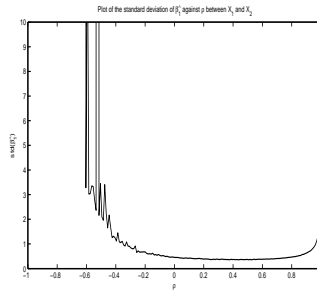
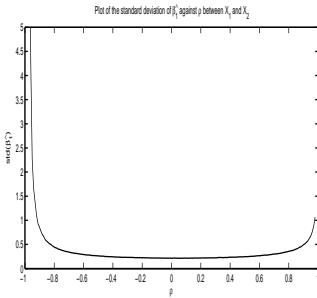


Figure 2.13: The $se(\beta_1)$ surface with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$

(a) $(\mu_{11}, \mu_{21}) = (0.5, 0.5)$

(b) $(\mu_{11}, \mu_{21}) = (2, 2)$

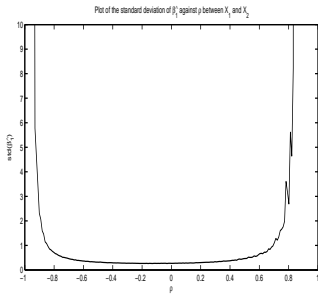
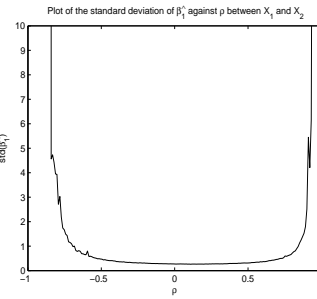
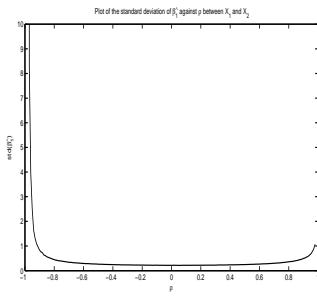
(c) $(\mu_{11}, \mu_{21}) = (0.5, 1)$



(d) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-0.5, -0.5)$

(e) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-2, -2)$

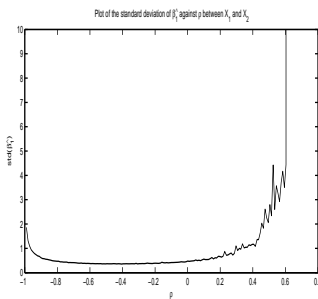
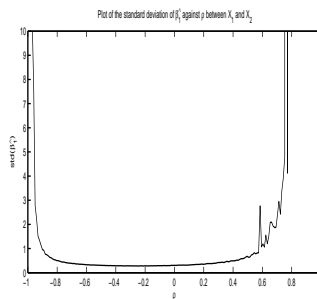
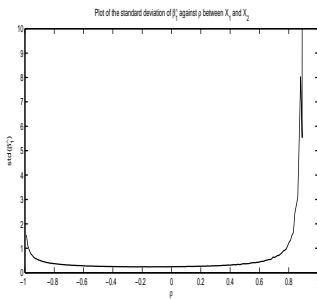
(f) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, -2)$



(g) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -1)$

(h) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -2)$

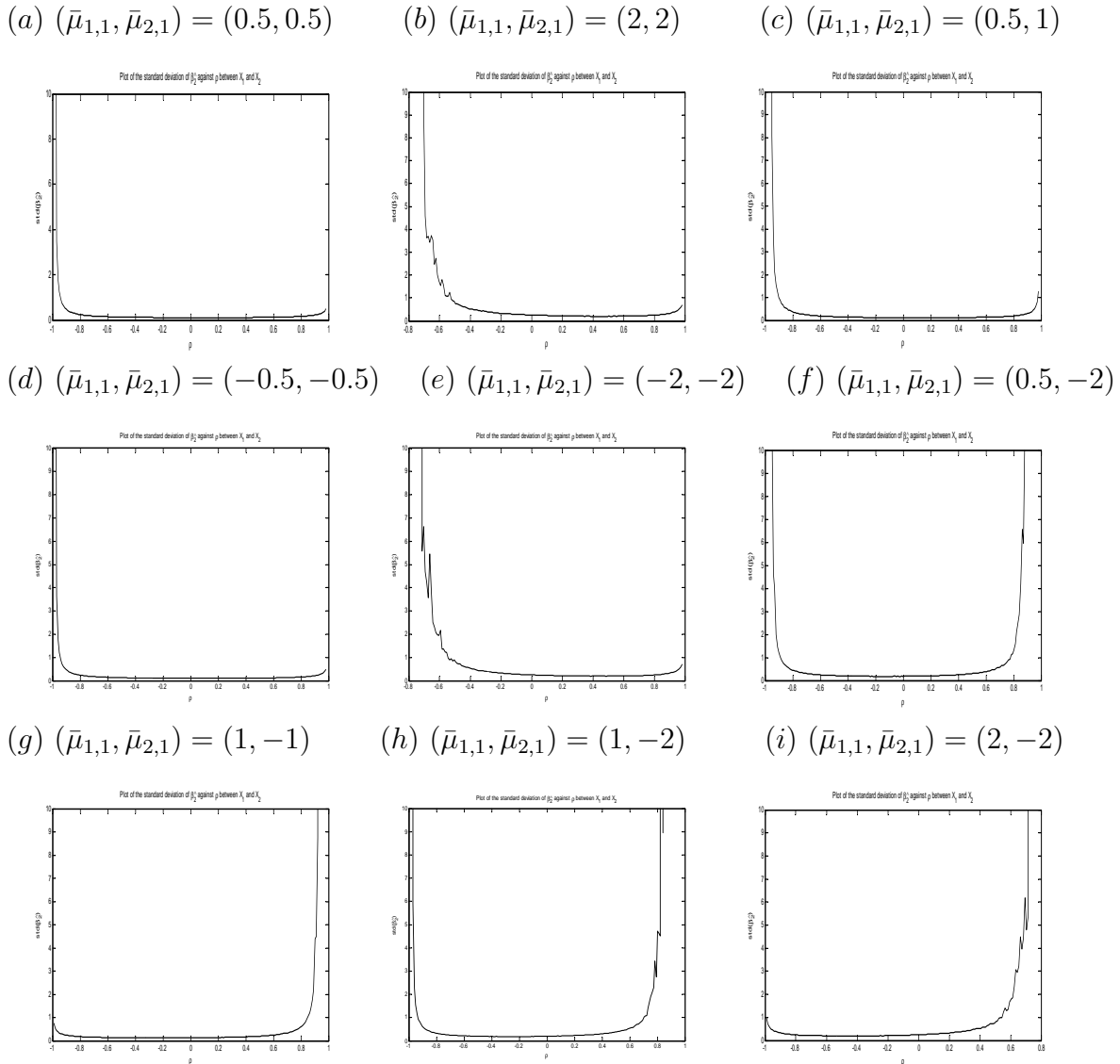
(i) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, -2)$



2.5.3.6 The standard error of the β_2 surface with varying ρ and different mean pairs (μ_{1i}, μ_{2i})

The standard errors of β_2 , $se(\beta_2)$, shown in Figures 2.14 and 2.15 for various $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$ combinations, and $p = 0.6$, $p = 0.95$ look quantitatively and qualitatively similar to $se(\beta_1)$ and $se(\beta_0)$. The conventional account that as $|\rho| \rightarrow 1$, the $se(\beta_2)$ increases

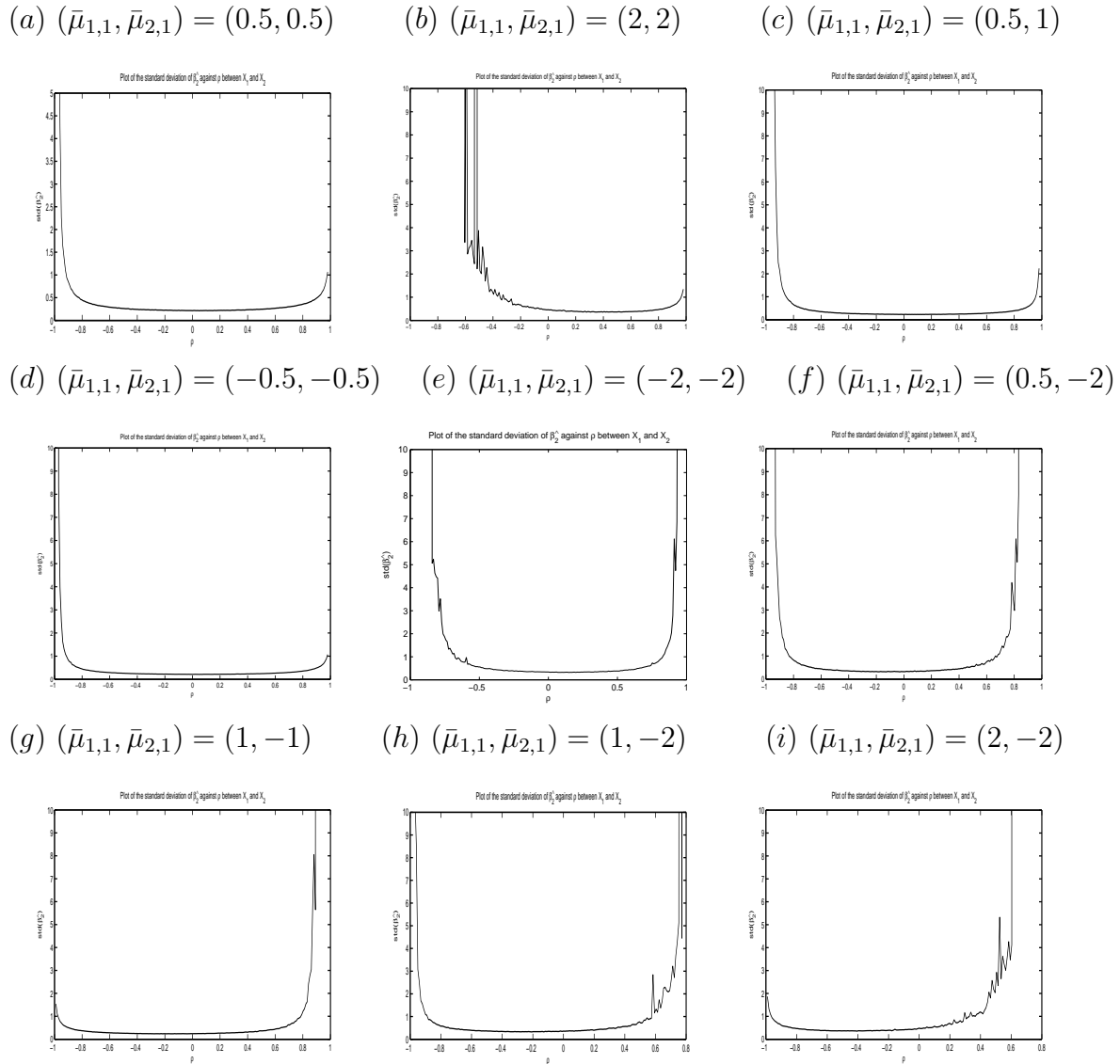
Figure 2.14: The $se(\beta_2)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$



monotonically is again called into question. Similar to Figures 2.10 through 2.13, the only noticeable U-shaped curves are in panels *f* and *h*.

An important conclusion derived from Figures 2.10 through 2.15 is that they call into question the traditional account of the effects of near-multicollinearity on the standard errors. When the variance inflates, it is not consistent either. Schaefer, Roi and Wolfe's (1984) conclusion, that "... we have found that multicollinearity does indeed inflate the

Figure 2.15: The $se(\beta_2)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$



estimated variance and hence can cause precision problems when identifying the effects of the independent variables” is not entirely confirmed by the analysis here.

2.5.3.7 The $\tau(\beta_0)$ surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$

The traditional discussion about near-multicollinearity usually associates near-multicollinearity with low t – statistics. As mentioned above, the traditional argument assumes that β_i s

are constant, while the standard errors are inflated. Since the standard t – *statistic* is calculated as $\beta_i/se(\beta_i)$ (under the $H_0 : \hat{\beta}_i = 0$) it follows then that near-multicollinearity should result in low $\tau(\beta_i)$. However, in light of the above discussion of the dynamics of β_i and $se(\beta_i)$, it should come as no surprise that the traditional argument is not supported by the figures of the asymptotic t – *statistics* here.

Figure 2.16: The $\tau(\beta_0)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$

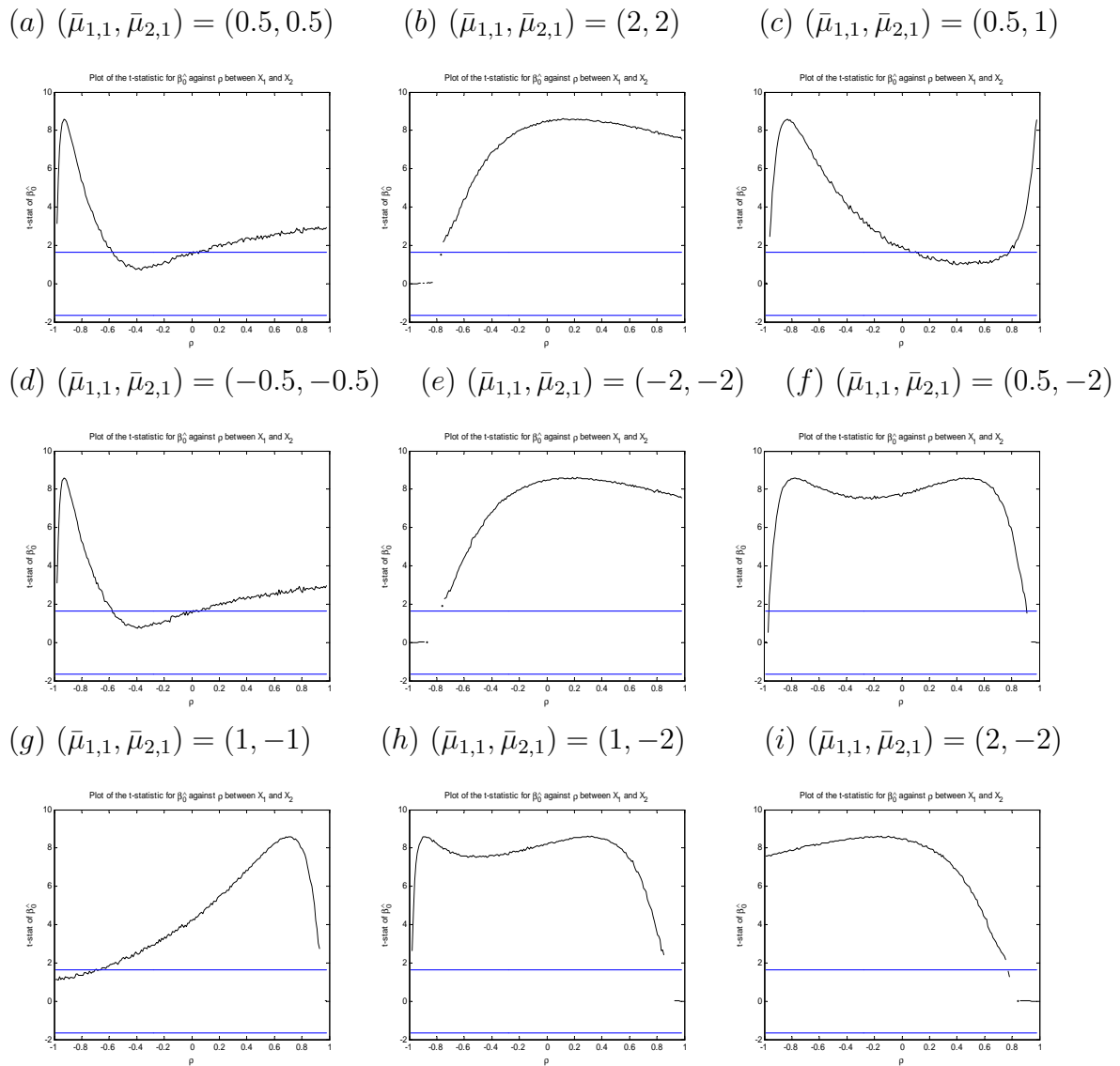
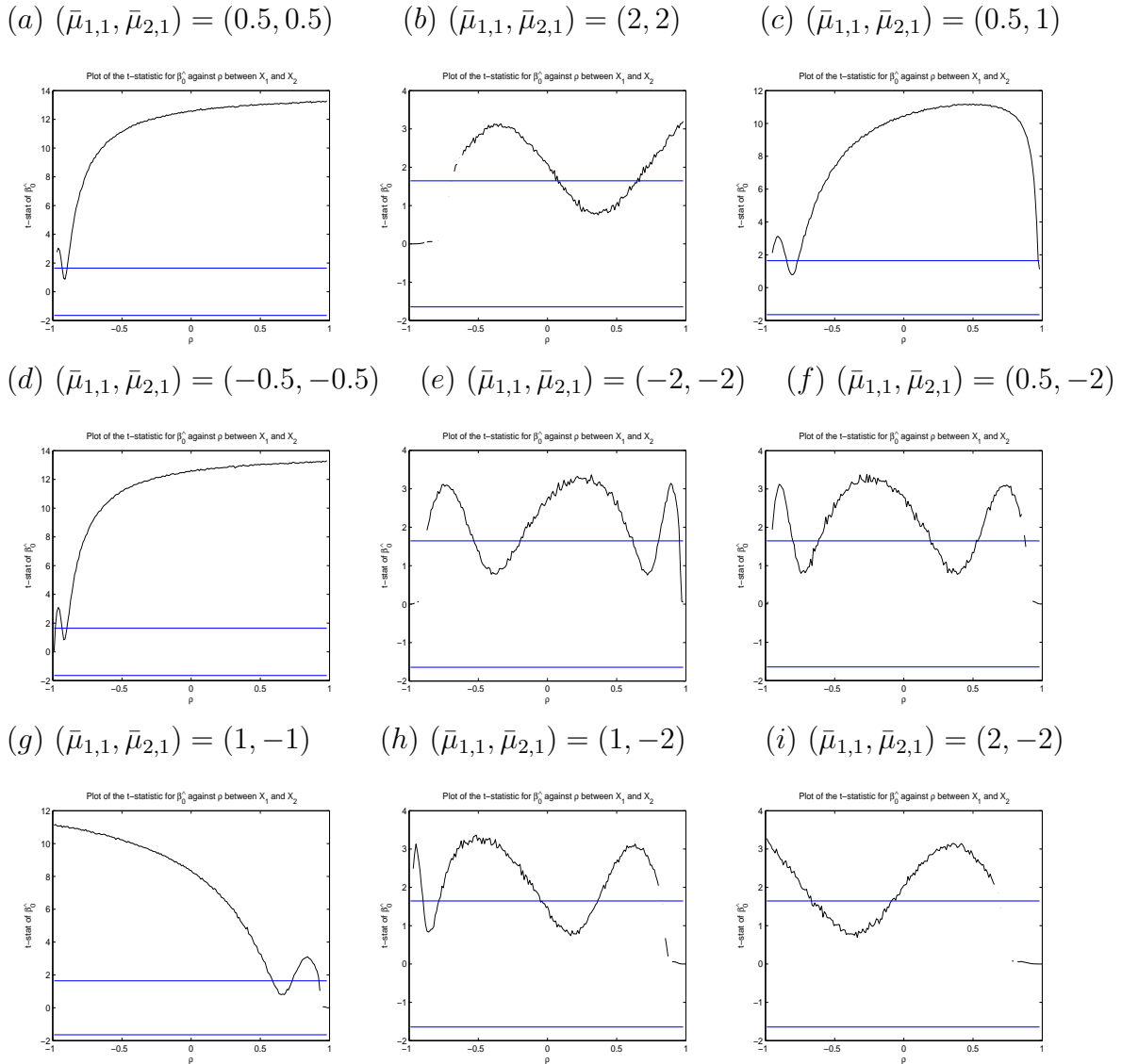


Figure 2.17: The $\tau(\beta_0)$ with varying ρ and different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$

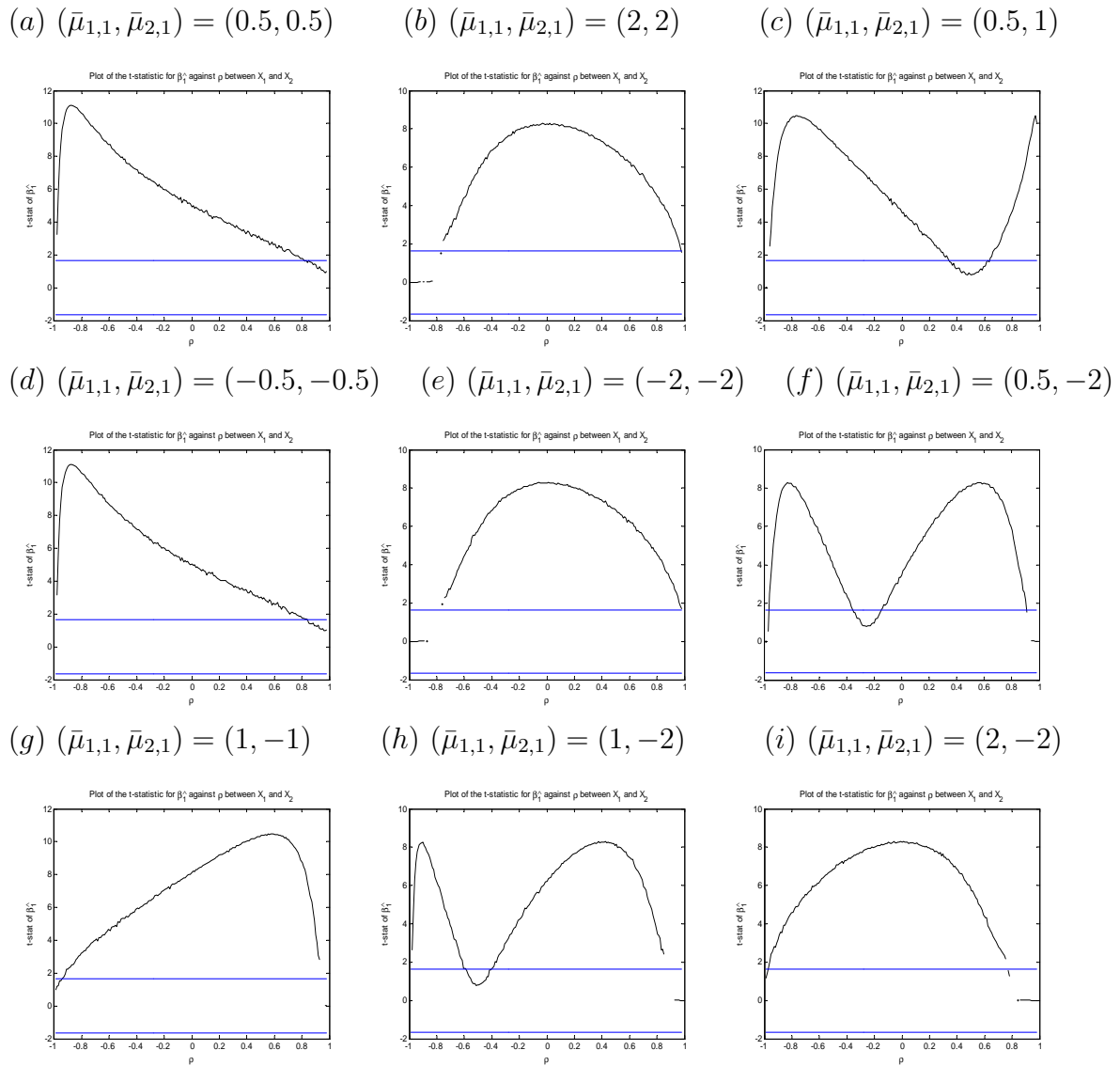


Figures 2.16 and 2.17 show how $\tau(\beta_0)$ changes as ρ changes for different p and different mean pairs (μ_{11}, μ_{21}) . In general, one observes no distinct pattern from these figures. While one should expect an inverse U pattern for $\tau(\beta_0)$ as $|\rho| \rightarrow 1$ under the traditional account, the figures suggest no monotonic relationship with ρ .

2.5.3.8 The $\tau(\beta_1)$ surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$

In Figures 2.18 and 2.19, we present the dynamics of $\tau(\beta_1)$. Similar to Figures 2.16 and 2.17, instead of $\tau(\beta_1)$ monotonically decreasing and symmetric around $\rho = 0$ as conventionally expected, there is no clear pattern for changes in $\tau(\beta_1)$ as $|\rho| \rightarrow 1$.

Figure 2.18: The Surface of the β_1 t -statistics With Varying ρ for Different (μ_{1i}, μ_{2i}) : $p = 0.6$

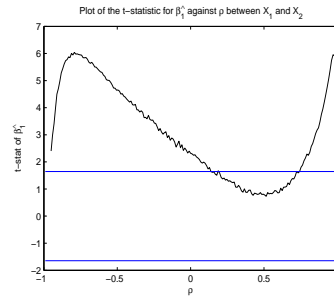
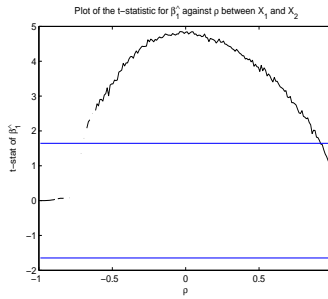
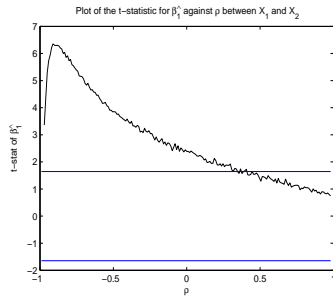


This result is expected because as noted earlier, β_1 and $se(\beta_1)$ do not follow the

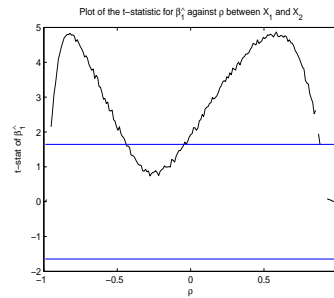
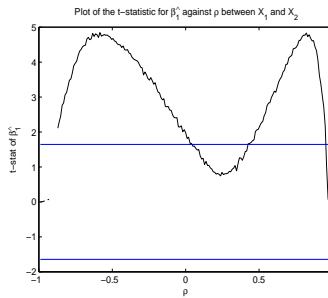
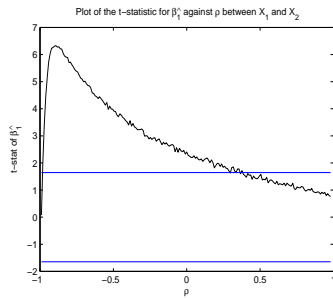
traditional account. There is therefore no reason to expect that $\tau(\beta_1)$ will.

Figure 2.19: The $\tau(\beta_1)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$

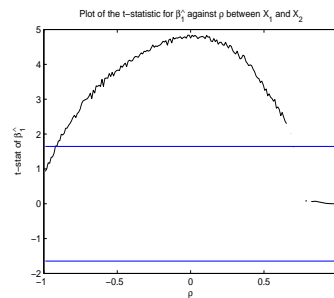
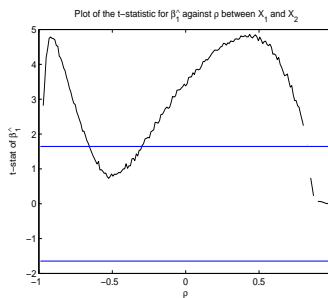
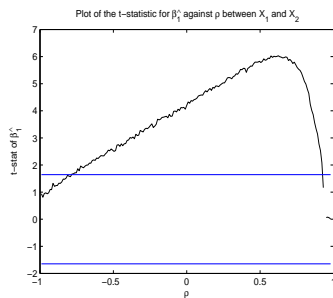
- (a) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 0.5)$ (b) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, 2)$ (c) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 1)$



- (d) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-0.5, -0.5)$ (e) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-2, -2)$ (f) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, -2)$

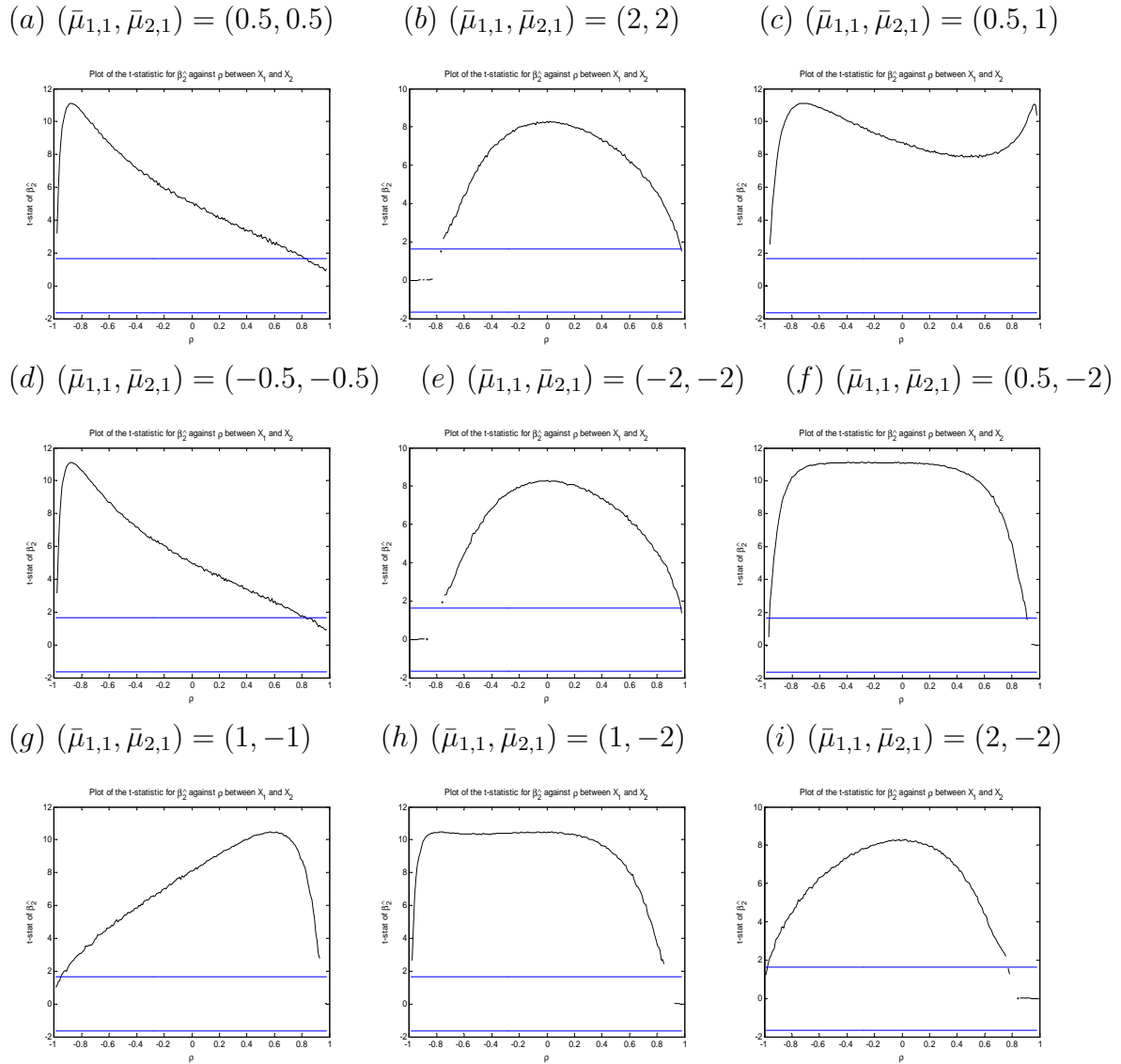


- (g) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -1)$ (h) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (1, -2)$ (i) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, -2)$



2.5.3.9 The $\tau(\beta_2)$ surface with varying ρ and different mean pairs $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$

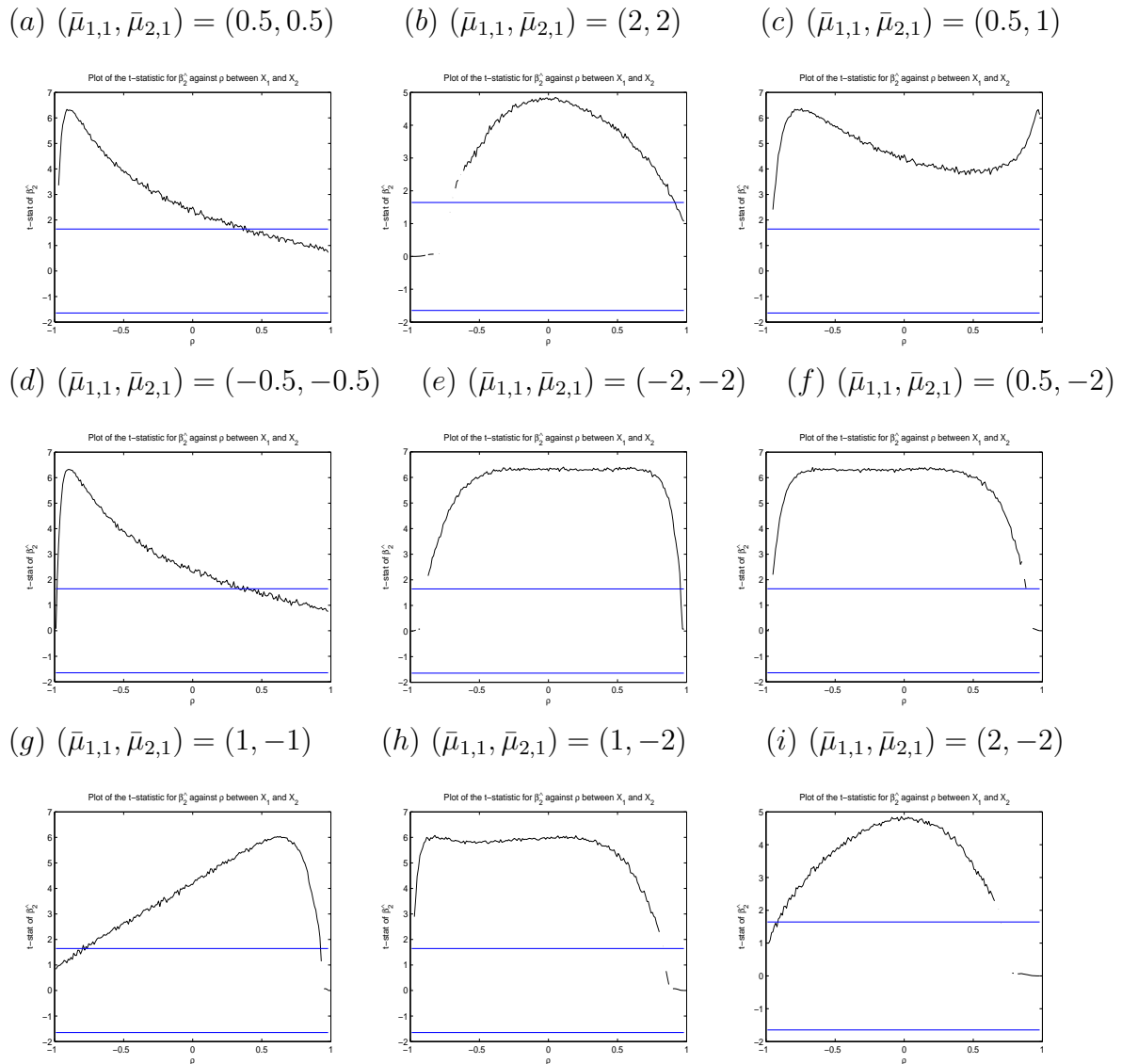
Figure 2.20: The $\tau(\beta_2)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$



Looking at Figures 2.20 and 2.21 which show the dynamics of the $\tau(\beta_2)$ surface with varying ρ for $p = 0.6$ and $p = 0.95$ respectively, we conclude that as far as the logistic regression model is concerned, the much-emphasized monotonically decreasing asymptotic t – ratios reported in the literature are unattainable for the logistic regression

model where near-multicollinearity is present in this case. This results because the **ceteris paribus** clause is unattainable - β_i are all functions of ρ and thus change (in a non-monotonic way) as ρ increases. Consequently, the t - ratios all change non-monotonically as $|\rho| \rightarrow 1$, as well.

Figure 2.21: $\tau(\beta_2)$ with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$



2.5.3.10 The R^2 , Condition Number, and VIF surfaces with varying ρ and different Mean Pairs (μ_{1i}, μ_{2i})

We only show the surfaces of the R^2 , Condition Number, and VIF for $\{(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})\} = \{(-0.5, -0.5), (2, 2), (0.5, 1), (2, -2)\}$ because the surfaces of the other mean pairs are similar to these. The traditional account of near-multicollinearity in the logit model suggests that as $|\rho| \rightarrow 1$, the measure of the model's goodness-of-fit, R^2 should not change. This implies that we expect R^2 to be constant as $\rho \rightarrow 1$.

Figures (2.22) and (2.23) show the dynamics of the McFadden pseudo R^2 (hereinafter referred to as R^2) as $|\rho| \rightarrow 1$. By the conventional argument, we expect the R^2 to be non-changing as $|\rho| \rightarrow 1$, but a glance at Figures (2.22) and (2.23) indicates that this is clearly not the case. The R^2 is at its minimum at some point ρ^* , where $\rho^* \neq 0$.

Figure 2.22: The McFadden pseudo R^2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$

(a) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, 2)$ (b) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 1)$ (c) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-0.5, -0.5)$ (d) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, -2)$

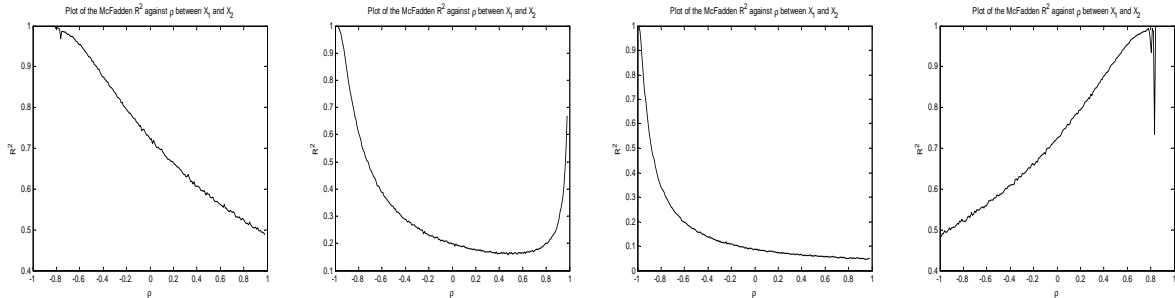
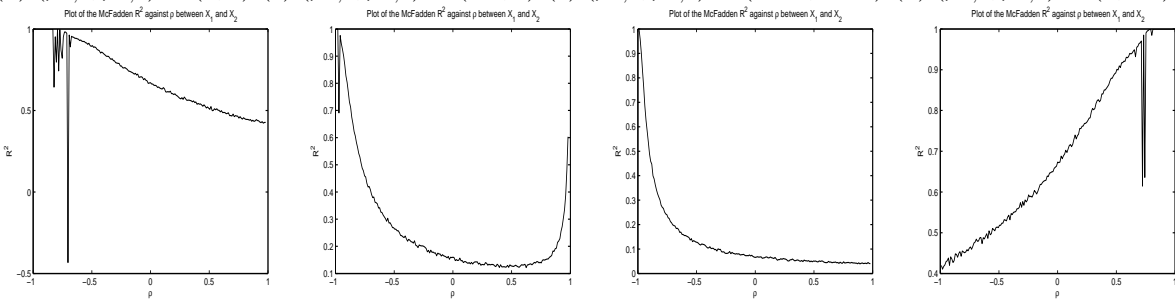


Figure 2.23: The McFadden pseudo R^2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$

(a) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, 2)$ (b) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (0.5, 1)$ (c) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (-0.5, -0.5)$ (d) $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1}) = (2, -2)$



Figures 2.24 and 2.25 show how the condition number, κ changes as $|\rho| \rightarrow 1$ for $p = 0.6$ and $p = 0.95$ respectively. As noted by Menard (2002), much of the methods for detecting near-multicollinearity, for example the *VIF* and κ can be obtained from the multiple linear regression model using the same outcome and predictor variables that one uses in the logistic regression model. Menard (2002) states that “because the concern is with the relationship among the independent variables, the functional form of the model for the dependent variable is irrelevant to the estimation of collinearity” (page 76). Therefore, one can estimate the linear model, ignore most of the results but still use these statistics that assist in diagnosing near-multicollinearity. However, Spanos and McGuirk (2002) conclude that:

one should clearly distinguish between *systematic* and *erratic* volatility. If the problem of concern is *erratic volatility*, matrix norm bounds based on $(\mathbf{X}^T \mathbf{X})$ not \mathbf{R} (where $\mathbf{R} := [\rho_{i,j}]_{i,j=1}^m$) should be used to quantify the potential erratic volatility associated with the particular \mathbf{X} . Diagnostics based on the matrices \mathbf{R} and \mathbf{R}_c (where \mathbf{R}_c is the centered correlation matrix) such as a condition number, can be misleading for quantifying erratic volatility, and are completely ineffective for detecting points of *systematic volatility*.
(page 392)

Based on Spanos and McGuirk’s (2002) conclusion, it follows that Menard’s (2002) account (and the traditional discussion) of near-multicollinearity needs to be ammended.

Our results agree with Spanos and McGuirk’s (2002) conclusion. Following the rules of thumb outlined by Belsey, Kuh and Welsch (1980), we expect a U-shaped κ that is symmetric and monotonic around $\rho = 0$, with $\kappa > 1000$ as $|\rho|$ gets close to 1. It turns out however, as shown in Figures 2.24 and 2.25 that in no instance do we see a κ that is U-shaped, monotonic or symmetric around $\rho = 0$ as $|\rho| \rightarrow 1$. Even in cases when ρ is as high as ± 0.99 , in no instance is $\kappa > 1000$, which brings into question the rules of

Figure 2.24: The *Condition Number* with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$

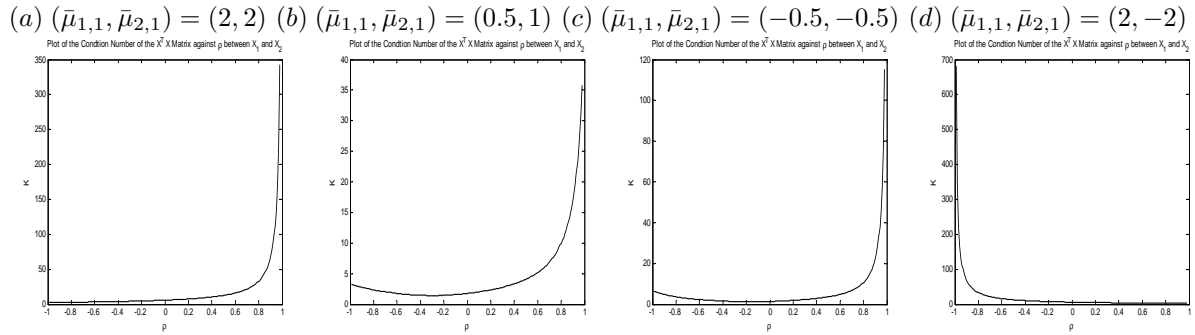
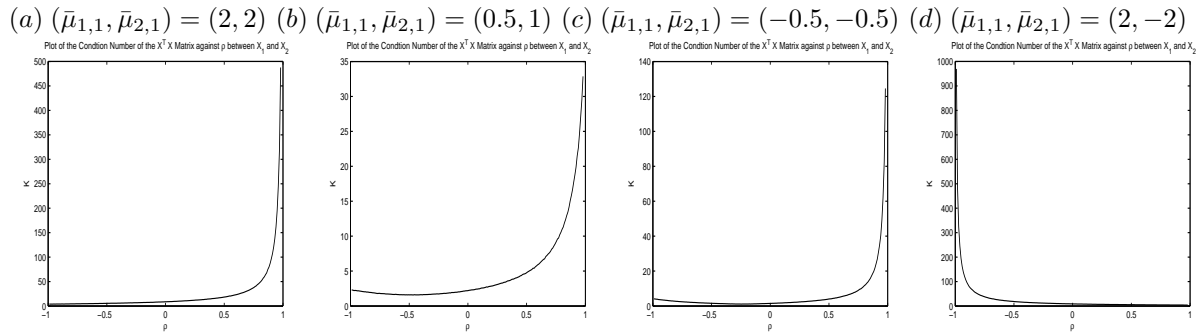


Figure 2.25: The *Condition Number* with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$



thumb outlined by Belsey, Kuh and Welsch (1980) that are traditionally used as cutoffs for detecting the presence of near-multicollinearity.

In Figures 2.26 -2.29 we show what happens to the $VIF(x_1)$ and $VIF(x_2)$ for different mean pairs and $p = 0.6$ and $p = 0.95$. Again, contrary to the “generally accepted” rules of thumb outlined by Marquardt (1970) that a $VIF > 10$ is indicative of severe near-multicollinearity, our results show that VIF s are likely ineffective in diagnosing near-multicollinearity.

Figure 2.26: The Variance Inflation Factors (VIF) of x_1 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$

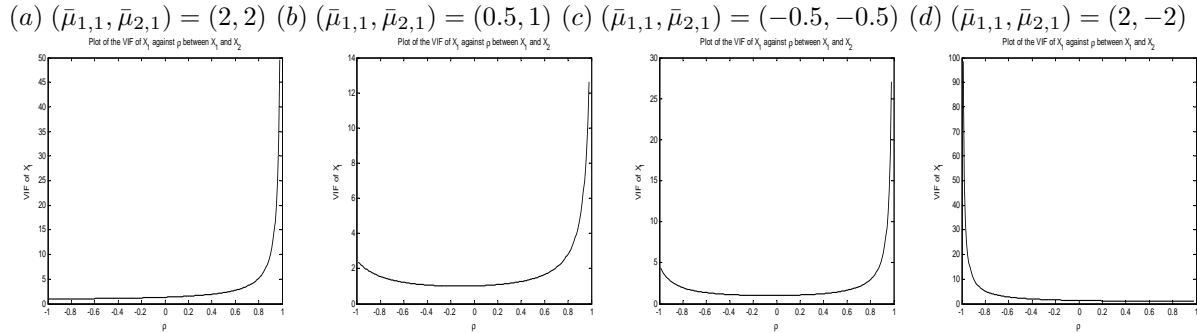
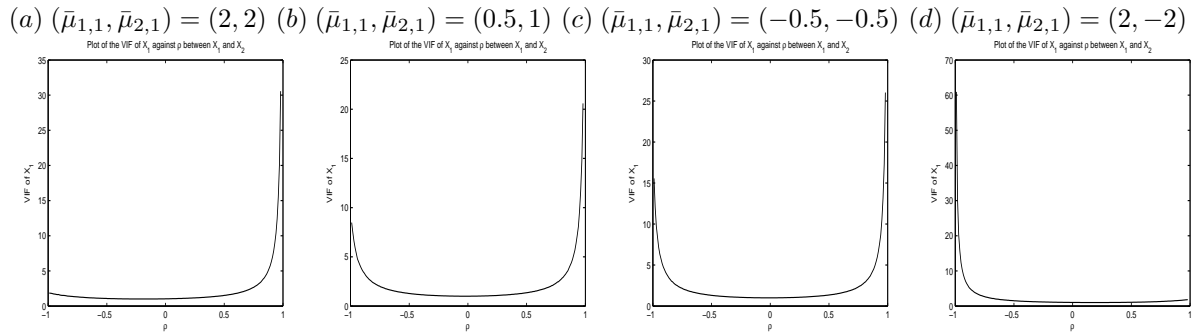


Figure 2.27: The Variance Inflation Factor (VIF) of x_1 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$



Figures 2.26 - 2.29 are for the most part L-shaped (not U-shaped, symmetric and monotonic as the traditional account suggests). When p is high (0.95), VIF s seem to perform better in detecting near-multicollinearity (see panels (b) and (c) of Figures 2.27 and 2.29), evidenced by a $VIF(\cdot) > 10$. However, in panel (a) of each Figure, only extreme cases of near-multicollinearity when $\rho \rightarrow 1$ are detectable, while in panel (d), only extreme cases of near-multicollinearity can be detected when $\rho \rightarrow -1$. This pattern follows in the other panels, suggesting that the VIF s do not provide a definitive approach to detecting near-multicollinearity.

Figure 2.28: The Variance Inflation Factor (VIF) of x_2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.6$

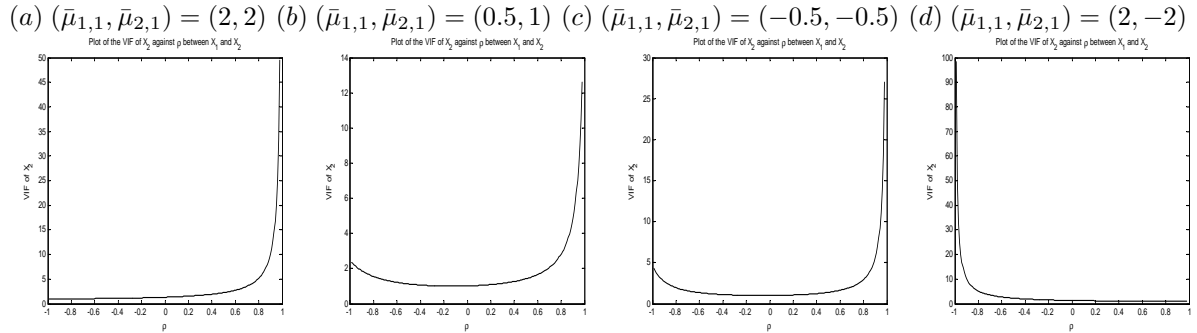
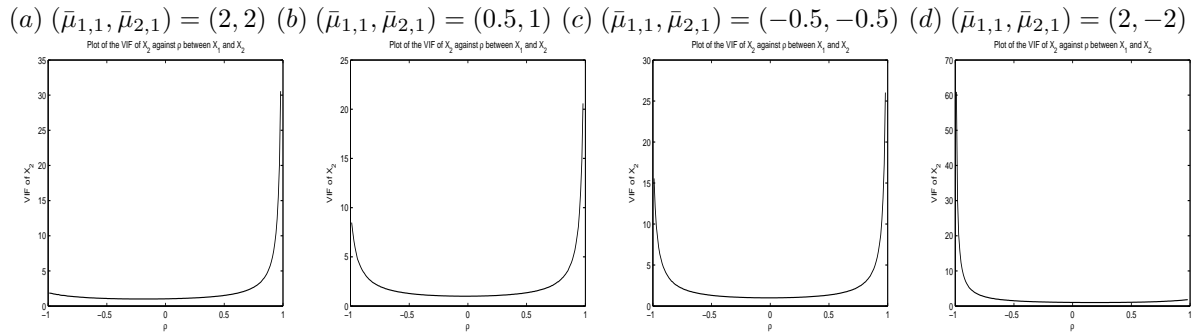


Figure 2.29: The Variance Inflation Factor (VIF) of x_2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$



2.5.3.11 Marginal Effects

For the logistic regression model, in general inferences concerning the magnitude of a covariate cannot be made directly using the β_i coefficients. Marginal effects provide a mechanism to measure the magnitude of a change in a covariate on the likelihood of an event occurring (i.e $Prob(Y = 1)$)

Figure 2.30: A: The Marginal Effects of x_1 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $P = 0.6$

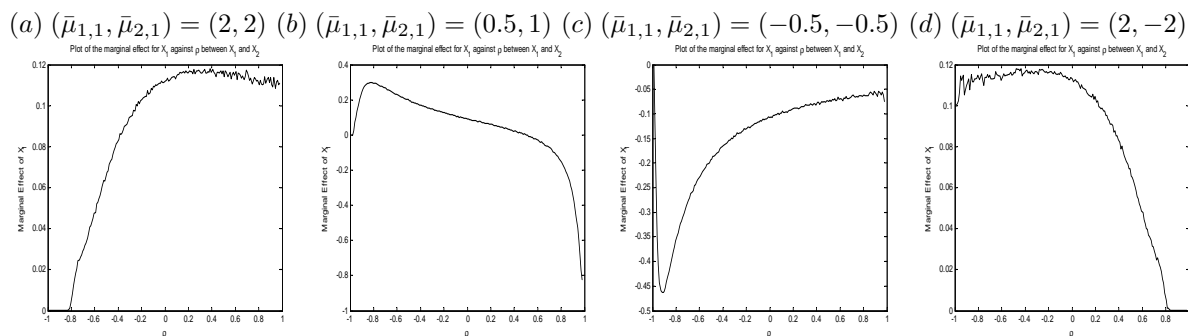
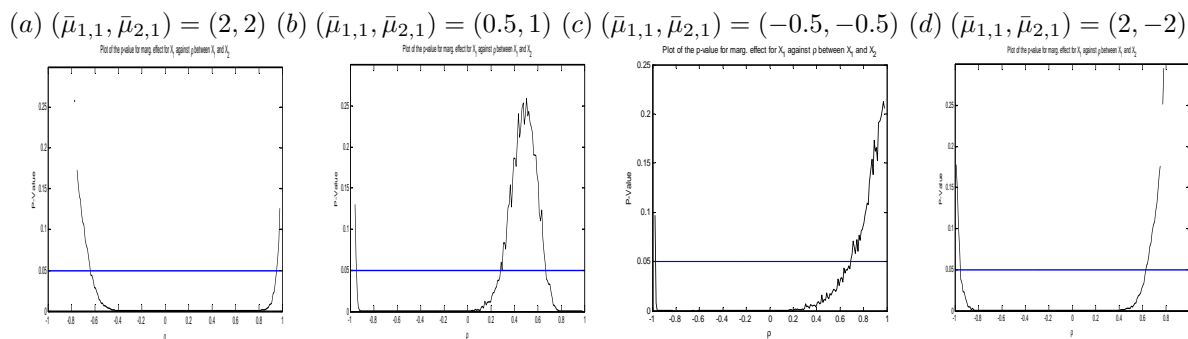


Figure 21: B: The p - values of the Marginal Effects of x_1 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $P = 0.6$

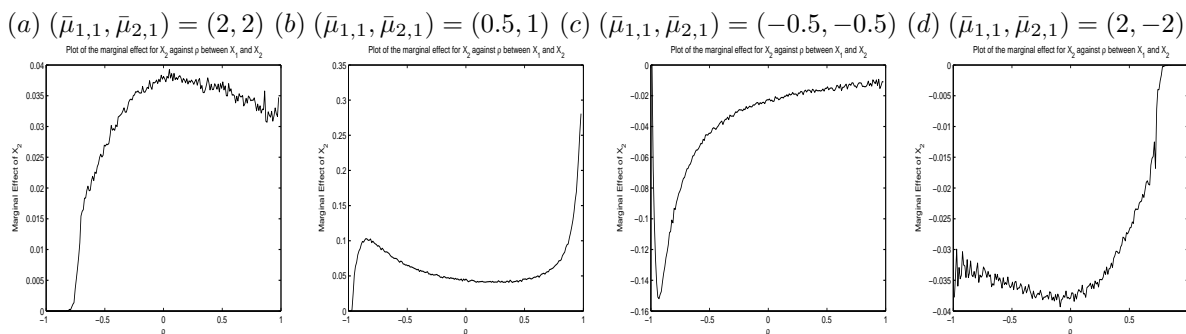


The marginal effects measure the change in predicted probability that results from corresponding changes in the independent variables (Greene, 2000). The marginal effects of the logistic regression model can be written as (see Bergtold, Spanos and Onukwugha, 2010):

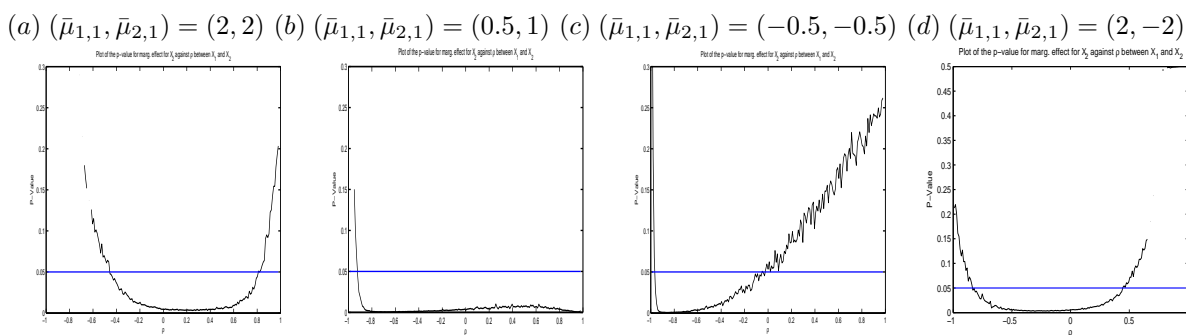
$$\frac{\partial h(\mathbf{X}_i; \beta)}{\partial \mathbf{X}_i} = h(\mathbf{X}_i; \beta) \cdot (1 - h(\mathbf{X}_i; \beta)) \cdot \frac{\partial \eta(\mathbf{X}_i; \beta)}{\partial \mathbf{X}_i} \quad (2.22)$$

where all variables and functions are as previously defined.

Figure 2.31: A: The Marginal Effects of x_2 with varying ρ for different $(\bar{\mu}_{1,1}, \bar{\mu}_{2,1})$: $p = 0.95$



B: The p - values of the Marginal Effects of x_2 with varying ρ : $P = 0.95$



Greene (2000) notes that $\frac{\partial h(\mathbf{X}_i; \beta)}{\partial \mathbf{X}_i}$ can be evaluated at the sample mean of the random variable (marginal effects at the means or MEMs) or evaluated for each random variable at every observation and then averaged across all observations (average marginal effects of AMEs). For binary variables, $\frac{\partial h(\mathbf{X}_i; \beta)}{\partial \mathbf{X}_i}$ is interpreted as the change in predicted probability when a regressor changes in value from '0' to '1', holding all other regressors constant at their observed values (Greene 2000).

While marginal effects have been broadly examined (see Greene (2000), Verlinda (2006) among others), to the best of our knowledge, the examination of marginal effects in the presence of multicollinearity in the logistic regression model has not received comparable attention. In light of the fact that the β_i s are functions of ρ , and hence change as ρ changes, it will be interesting to know how the marginal effects change as

$|\rho| \rightarrow 1$. Figures 2.30 to 2.31 show the marginal effects and associated p -values for mean pairs $(\mu_{1i}, \mu_{2i}) = (2, 2), (0.5, 1), (-0.5, -0.5),$ and $(2, -2)$. These Figures show that the marginal effects differ immensely for each mean pair as $|\rho| \rightarrow 1$. That is, there is no distinctive behavior or predictable response as $|\rho| \rightarrow 1$ | This result is not surprising given the unstable nature of the β_i s and that $h(\mathbf{X}_i; \beta)$ is a function of β_i .

2.6 Extensions

This section extends the analysis of near-multicollinearity to the logistic regression model with continuous normal covariates with heterogeneous variance-covariance structure, as well as logistic regression models with binary covariates. We also examine near-multicollinearity in the k -regressor case of the logistic regression model.

2.6.1 The Logistic Regression Model with Continuous Normal Covariates, Different Mean Vectors and a Heterogeneous Variance-Covariance Structure

If the variance-covariance matrix structure is heterogeneous (and unequal for $j = 0, 1$) and $f_{X|Y}(\mathbf{X}_i; \theta_j)$ for $Y_i = j$, then the predictor function takes the form (Bergtold, Spanos and Onukwugha, 2010):

$$\eta(\mathbf{x}_i; \beta) = \beta_0 + \sum_{k=1}^K \beta_k x_{k,i} + \sum_{j=1}^K \sum_{l \geq j}^K \beta_{j,l} x_{j,i} x_{l,i} \quad (2.23)$$

In the simple two variable case, this index function takes the form $\eta(\mathbf{x}_i; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + + \beta_{22} x_2^2$, so that the log-odds function becomes:

$$\begin{aligned}
\ln \left(\frac{P(y=1|\mathbf{X};\beta)}{P(y=0|\mathbf{X};\beta)} \right) &= \ln \left(\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i;\theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i;\theta_0)} \right) + \log \left(\frac{p}{1-p} \right) + u_i \\
&= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 + u_i
\end{aligned} \tag{2.24}$$

To see why this is interesting, equation(2.12) can be written to include the heterogeneity inherent in the variance-covariance matrix as follows:

$$\begin{aligned}
f_{X/Y}(x_1, x_2; \theta_j \ j = 0, 1) &= \frac{(1-\rho_j^2)^{-\frac{1}{2}}}{2\pi\sigma_{1j}\sigma_{2j}} \cdot \exp \left\{ -\frac{1}{2} \frac{1}{1-\rho_j^2} \left[\left(\frac{x_1-\mu_{1j}}{\sigma_{1j}} \right)^2 \right. \right. \\
&\quad \left. \left. -2\rho_j \left(\frac{x_1-\mu_{1j}}{\sigma_{1j}} \right) \left(\frac{x_2-\mu_{2j}}{\sigma_{2j}} \right) + \left(\frac{x_2-\mu_{2j}}{\sigma_{2j}} \right)^2 \right] \right\} \\
&= \frac{(1-\rho_j^2)^{-\frac{1}{2}}}{2\pi\sigma_{1j}\sigma_{2j}} \cdot \exp \left\{ -\frac{1}{2(1-\rho_j^2)} \left[\frac{x_1^2-2x_1\mu_{1j}+\mu_{1j}^2}{\sigma_{1j}^2} \right. \right. \\
&\quad \left. \left. \left(2\rho_j \frac{x_1x_2-x_1\mu_{2j}-x_2\mu_{1j}+\mu_{1j}\mu_{2j}}{\sigma_{1j}\sigma_{2j}} \right) + \frac{x_2^2-2x_2\mu_{2j}+\mu_{2j}^2}{\sigma_{2j}^2} \right] \right\}
\end{aligned} \tag{2.25}$$

Combining equations (2.12) and (2.25) gives:

$$\begin{aligned}
\ln \frac{f_1(x_1, x_2; \theta_{1,j})}{f_0(x_1, x_2; \theta_{0,j})} &= \left\{ \frac{2(1-\rho_0^2)^{\frac{1}{2}} 2\sigma_{10}\sigma_{20}}{2(1-\rho_1^2)^{\frac{1}{2}} 2\sigma_{11}\sigma_{21}} \cdot \frac{\exp\{\Delta_1\}}{\exp\{\Delta_0\}} \right\} = \ln \frac{2(1-\rho_0^2)^{\frac{1}{2}} 2\sigma_{10}\sigma_{20}}{2(1-\rho_1^2)^{\frac{1}{2}} 2\sigma_{11}\sigma_{21}} (\Delta_1 - \Delta_0) \\
&= \ln D (\Delta_1 - \Delta_0)
\end{aligned} \tag{2.26}$$

where $D = \frac{2(1-\rho_0^2)^{\frac{1}{2}} 2\sigma_{10}\sigma_{20}}{2(1-\rho_1^2)^{\frac{1}{2}} 2\sigma_{11}\sigma_{21}}$ and Δ_j is as defined in equation (2.12). Plugging in Δ_1 and Δ_0 into (2.26) gives

$$\begin{aligned}
\ln \frac{f_1(x_1, x_2; \theta_j)}{f_0(x_1, x_2; \theta_j)} &= \ln D + \left[\frac{1}{2(1-\rho_0^2)\sigma_{10}^2} - \frac{1}{2(1-\rho_1^2)\sigma_{11}^2} \right] x_1^2 + \left[\frac{1}{2(1-\rho_0^2)\sigma_{20}^2} - \frac{1}{2(1-\rho_1^2)\sigma_{21}^2} \right] x_2^2 \\
&+ \left[\frac{\rho_1}{2(1-\rho_1^2)\sigma_{11}\sigma_{21}} - \frac{\rho_0}{2(1-\rho_0^2)\sigma_{10}\sigma_{20}} \right] x_1 x_2 + \left[\frac{\mu_{11}}{2(1-\rho_1^2)\sigma_{11}^2} + \frac{\rho_0 \mu_{20}}{2(1-\rho_0^2)\sigma_{10}\sigma_{20}} \right. \\
&- \left. \frac{\mu_{10}}{2(1-\rho_0^2)\sigma_{10}^2} - \frac{\rho_1 \mu_{21}}{2(1-\rho_1^2)\sigma_{11}\sigma_{21}} \right] x_1 + \left[\frac{\mu_{21}}{2(1-\rho_1^2)\sigma_{21}^2} + \frac{\rho_0 \mu_{10}}{2(1-\rho_0^2)\sigma_{10}\sigma_{20}} \right. \\
&- \left. \frac{\mu_{20}}{2(1-\rho_0^2)\sigma_{20}^2} - \frac{\rho_1 \mu_{11}}{2(1-\rho_1^2)\sigma_{11}\sigma_{21}} \right] x_2 + \left[\frac{\rho_1 \mu_{11} \mu_{21}}{2(1-\rho_1^2)\sigma_{11}\sigma_{21}} + \frac{\mu_{10}^2}{2(1-\rho_0^2)\sigma_{10}^2} \right. \\
&+ \left. \frac{\mu_{20}^2}{2(1-\rho_0^2)\sigma_{20}^2} - \frac{\mu_{11}^2}{2(1-\rho_1^2)\sigma_{11}^2} - \frac{\mu_{21}^2}{2(1-\rho_1^2)\sigma_{21}^2} - \frac{\rho_0 \mu_{10} \mu_{20}}{2(1-\rho_0^2)\sigma_{10}\sigma_{20}} \right] \Big\}
\end{aligned} \tag{2.27}$$

Equation (2.27) can be rewritten as:

$$\begin{aligned}
\frac{f_{\mathbf{X}|Y=1}(x_1, x_2; \theta_1)}{f_{\mathbf{X}|Y=0}(x_1, x_2; \theta_0)} &= \exp \left\{ \ln D + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \beta_1 x_1 + \beta_2 x_2 + \beta_0 \right\} \\
&= \exp \left\{ \tilde{\beta}_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 \right\}
\end{aligned} \tag{2.28}$$

where $\tilde{\beta}_0 = \ln D + \beta_0$ and

$$\beta_0 = \left[\frac{\rho_1 \mu_{11} \mu_{21}}{2(1-\rho_1^2)\sigma_{11}\sigma_{21}} + \frac{\mu_{10}^2}{2(1-\rho_0^2)\sigma_{10}^2} + \frac{\mu_{20}^2}{2(1-\rho_0^2)\sigma_{20}^2} - \frac{\mu_{11}^2}{2(1-\rho_1^2)\sigma_{11}^2} - \frac{\mu_{21}^2}{2(1-\rho_1^2)\sigma_{21}^2} - \frac{\rho_0 \mu_{10} \mu_{20}}{2(1-\rho_0^2)\sigma_{10}\sigma_{20}} \right], \tag{2.29}$$

$$\beta_1 = \left[\frac{\mu_{11}}{2(1-\rho_1^2)\sigma_{11}^2} + \frac{\rho_0 \mu_{20}}{2(1-\rho_0^2)\sigma_{10}\sigma_{20}} - \frac{\mu_{10}}{2(1-\rho_0^2)\sigma_{10}^2} - \frac{\rho_1 \mu_{21}}{2(1-\rho_1^2)\sigma_{11}\sigma_{21}} \right], \tag{2.30}$$

$$\beta_2 = \left[\frac{\mu_{21}}{2(1-\rho_1^2)\sigma_{21}^2} + \frac{\rho_0 \mu_{10}}{2(1-\rho_0^2)\sigma_{10}\sigma_{20}} - \frac{\mu_{20}}{2(1-\rho_0^2)\sigma_{20}^2} - \frac{\rho_1 \mu_{11}}{2(1-\rho_1^2)\sigma_{11}\sigma_{21}} \right], \tag{2.31}$$

$$\beta_{11} = \left[\frac{1}{2(1 - \rho_1^2)\sigma_{11}^2} - \frac{1}{2(1 - \rho_0^2)\sigma_{10}^2} \right], \quad (2.32)$$

$$\beta_{22} = \left[\frac{1}{2(1 - \rho_0^2)\sigma_{20}^2} - \frac{1}{2(1 - \rho_1^2)\sigma_{21}^2} \right], \quad (2.33)$$

$$\beta_{12} = \left[\frac{\rho_1}{2(1 - \rho_1^2)\sigma_{11}\sigma_{21}} - \frac{\rho_0}{2(1 - \rho_0^2)\sigma_{10}\sigma_{20}} \right], \quad (2.34)$$

It is clear to see that $\tilde{\beta}_0$, β_1 , β_2 , β_{11} , β_{12} and β_{22} , are all functions of ρ_j , $j = 0, 1$ and as a result change as ρ_j changes.

2.6.2 The Logistic Regression Model with Binary Covariates

Assume now that $\mathbf{X} = (x_1, x_2)'$ where \mathbf{X} is distributed bivariate Bernoulli. Then we can show that:

$$\eta(\mathbf{x}_i; \beta) = \ln \left(\frac{f_1(x_1, x_2|Y = j)}{f_0(x_1, x_2|Y = j)} \right) + \log \left(\frac{p}{1 - p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$$

For $\mathbf{X} = (x_1, x_2)'$ distributed conditional bivariate Bernoulli,

$$f_j(x_1|Y = j) = r_j^{x_1}(1 - r_j)^{1-x_1}; \quad j = 0, 1 \quad (2.35)$$

where $r_j = Prob(x_1 = 1|Y = j)$, and:

$$f_j(x_2|x_1 = k, Y = j) = q_{jk}^{x_2}(1 - q_{jk})^{1-x_2}; \quad j = 0, 1, \quad k = 0, 1 \quad (2.36)$$

making:

$$f_j(x_1, x_2|Y = j) = [r_j q_{j1}^{x_2}(1 - q_{j1})^{1-x_2}]^{x_1} \cdot [(1 - r_j) q_{j0}^{x_2}(1 - q_{j0})^{1-x_2}]^{(1-x_1)}; \quad j = 0, 1, \quad (2.37)$$

where $q_{jk} = Prob(x_2 = 1|x_1 = l, Y = j)$; for $k = 0, 1$ and $j = 0, 1$.

From equation (2.37)

$$\begin{aligned} \ln \left(\frac{f_1(x_1, x_2|Y=j)}{f_0(x_1, x_2|Y=j)} \right) &= x_1 \ln [r_1 q_{11}^{x_2}(1 - q_{11})^{(1-x_2)}] + (1 - x_1) \ln [(1 - r_1) q_{10}^{x_2}(1 - q_{10})^{(1-x_2)}] \\ &- x_1 \ln [r_0 q_{01}^{x_2}(1 - q_{01})^{(1-x_2)}] + (1 - x_1) \ln [(1 - r_0) q_{00}^{x_2}(1 - q_{00})^{(1-x_2)}] \\ &= x_1 \ln(r_1) + x_1 x_2 \ln(q_{11}) + x_1(1 - x_2) \ln(1 - q_{11}) + (1 - x_1) \ln(1 - r_1) \\ &+ (1 - x_1) x_2 \ln(q_{10}) + (1 - x_1)(1 - x_2) \ln(1 - q_{10}) - x_1 \ln(r_0) \\ &- x_1 x_2 \ln(q_{01}) - x_1(1 - x_2) \ln(1 - q_{01}) - (1 - x_1) \ln(1 - r_0) \\ &- (1 - x_1) x_2 \ln(q_{00}) - (1 - x_1)(1 - x_2) \ln(1 - q_{00}) \end{aligned} \quad (2.38)$$

Further expanding and rearranging equation (2.38) yields

$$\begin{aligned}
\ln \left(\frac{f_1(x_1, x_2 | Y=j)}{f_0(x_1, x_2 | Y=j)} \right) &= x_1 \ln \left(\frac{r_1}{r_0} \right) + x_1 x_2 \ln \left(\frac{q_{11}}{q_{01}} \right) + x_1 (1-x_2) \ln \left(\frac{1-q_{11}}{1-q_{01}} \right) + (1-x_1) \ln \left(\frac{1-r_1}{1-r_0} \right) \\
&+ x_2 (1-x_1) \ln \left(\frac{q_{10}}{q_{00}} \right) = (1-x_1)(1-x_2) \ln \left(\frac{1-q_{10}}{1-q_{00}} \right) \\
&= x_1 \ln \left(\frac{r_1}{r_0} \right) + x_1 x_2 \ln \left(\frac{q_{11}}{q_{01}} \right) + [x_2 - x_1 x_2] \ln \left(\frac{1-q_{11}}{1-q_{01}} \right) + (1-x_1) \ln \left(\frac{1-r_1}{1-r_0} \right) \\
&+ [x_2 - x_1 x_2] \ln \left(\frac{q_{10}}{q_{00}} \right) + [1 - x_1 - x_2 + x_1 x_2] \ln \left(\frac{1-q_{10}}{1-q_{00}} \right),
\end{aligned} \tag{2.39}$$

giving rise to:

$$\eta(x_1, x_2; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \tag{2.40}$$

where:

$$\beta_0 = \kappa + \ln \left(\frac{1-r_1}{1-r_0} \right) + \ln \left(\frac{1-q_{10}}{1-q_{00}} \right) = \kappa + \ln \left[\frac{(1-r_1)(1-q_{10})}{(1-r_0)(1-q_{00})} \right]$$

$$\begin{aligned}
\beta_1 &= \ln \left(\frac{r_1}{r_0} \right) + \ln \left(\frac{1-q_{11}}{1-q_{01}} \right) - \ln \left(\frac{1-r_1}{1-r_0} \right) - \ln \left(\frac{1-q_{10}}{1-q_{00}} \right) \\
&= \ln \left[\frac{\left(\frac{r_1}{r_0} \right) \left(\frac{1-q_{11}}{1-q_{01}} \right)}{\left(\frac{1-r_1}{1-r_0} \right) \left(\frac{1-q_{10}}{1-q_{00}} \right)} \right] = \ln \left[\frac{r_1(1-r_0)(1-q_{11})(1-q_{00})}{r_0(1-r_1)(1-q_{01})(1-q_{10})} \right]
\end{aligned}$$

$$\beta_2 = \ln \left(\frac{q_{10}}{q_{00}} \right) - \ln \left(\frac{1-q_{10}}{1-q_{00}} \right) = \ln \left[\frac{\left(\frac{q_{10}}{q_{00}} \right)}{\left(\frac{1-q_{10}}{1-q_{00}} \right)} \right] = \ln \left[\frac{q_{10}(1-q_{00})}{q_{00}(1-q_{10})} \right]$$

$$\beta_{12} = \ln \left(\frac{q_{11}}{q_{01}} \right) - \ln \left(\frac{1-q_{11}}{1-q_{01}} \right) - \ln \left(\frac{q_{11}}{q_{01}} \right) + \ln \left(\frac{1-q_{10}}{1-q_{00}} \right) = \ln \left[\frac{q_{11}(1-q_{01})q_{00}(1-q_{10})}{q_{01}(1-q_{11})q_{10}(1-q_{00})} \right]$$

Further expansion of (β_{12}) yields

$$\beta_{12} = \ln \left[\frac{(q_{11} - q_{11}q_{01})(q_{00} - q_{00}q_{10})}{q_{01} - q_{01}q_{11})(q_{10} - q_{10}q_{00})} \right] = \ln \left[\frac{q_{11}q_{00}[1 - q_{10} - q_{01} + q_{01}q_{10}]}{q_{01}q_{10}[1 - q_{00} - q_{11} + q_{00}q_{11}]} \right]$$

It can further be shown that $\beta_0, \beta_1, \beta_2, \beta_{12}$ are all functions of ρ , but the correlations are expressed in terms of *Yule's Q*. *Yule's Q* is a numerical transformation that preserves the rank order of the data and establishes a more traditional range to the index so that -1 represents a perfect negative relationship, 0 , no relationship and 1 , a perfect positive relationship (Bakeman et al. 1996), thereby providing a measure of association to examine dependence among nominal variables. Mathematically, it can be expressed as

$$Yule's\ Q_j \in (-1, 1) = \frac{\rho_{00}\rho_{11} - \rho_{01}\rho_{10}}{\rho_{00}\rho_{11} + \rho_{01}\rho_{10}} \quad (2.41)$$

Since x_1, x_2 are binary, from equation (2.37):

$$\begin{aligned} \rho_{11} &= f_j(x_1 = 1, x_2 = 1) = r_j q_{j1} \\ \rho_{00} &= f_j(x_1 = 0, x_2 = 0) = (1 - r_j)(1 - q_{j0}) \\ \rho_{01} &= f_j(x_1 = 0, x_2 = 1) = (1 - r_j)q_{j0} \\ \rho_{10} &= f_j(x_1 = 1, x_2 = 0) = r_j(1 - q_{j1}) \end{aligned}$$

Substituting $\rho_{00}, \rho_{11}, \rho_{01}, \rho_{10}$ from above into (2.41) and rearranging:

$$Q_j = \frac{q_{j1}(1 - q_{j0}) - q_{j0}(1 - q_{j1})}{q_{j1}(1 - q_{j0}) + q_{j0}(1 - q_{j1})} \quad \text{for } j = 0, 1 \quad (2.42)$$

If $\beta_{12} = \ln(1) = 0$, then it can be mathematically shown that $\frac{Q_1}{Q_2} = 1$. Goodman (1965) points out an important relationship between *Yule's Q* and the so-called *cross-product ratio*. Denote by CRP_j , the *cross-product ratio*. Goodman (1965) shows that $Q_j = \frac{CRP_j - 1}{CRP_j + 1}$, so that $CRP_j = \frac{1 + Q_j}{1 - Q_j}$.

Using equation (2.42), for $j = 0, 1$

$$CPR_j \equiv \frac{\rho_{00}\rho_{11}}{\rho_{01}\rho_{10}} = \frac{r_j(1-r_j)q_{j1}(1-q_{j0})}{r_j(1-r_j)q_{j0}(1-q_{j1})} = \frac{q_{j1}(1-q_{j0})}{q_{j0}(1-q_{j1})} \quad (2.43)$$

so that the ratio of cross-products is:

$$\frac{CPR_1}{CPR_0} = \frac{\frac{q_{11}(1-q_{10})}{q_{10}(1-q_{11})}}{\frac{q_{01}(1-q_{00})}{q_{00}(1-q_{01})}} = \frac{q_{11}(1-q_{10})q_{00}(1-q_{10})}{q_{01}(1-q_{11})q_{10}(1-q_{00})} = \exp(\beta_{12}) \quad (2.44)$$

Consequently, we can rewrite (2.44) as

$$\frac{CPR_1}{CPR_0} = \frac{\frac{1+Q_1}{1-Q_1}}{\frac{1+Q_0}{1-Q_0}} = \frac{(1+Q_1)(1-Q_0)}{(1+Q_0)(1-Q_1)} = \exp(\beta_{12}) \quad (2.45)$$

giving

$$\beta_{12} = \ln \left[\frac{(1+Q_1)(1-Q_0)}{(1+Q_0)(1-Q_1)} \right] \quad (2.46)$$

Similarly, we can show that β_1 is also a function of CPR_j and Q_j as follows

$$\begin{aligned} \beta_1 &= \ln \left(\frac{r_1(1-r_0)(1-q_{11})(1-q_{00})}{r_0(1-r_1)(1-q_{01})(1-q_{10})} \cdot \frac{q_{01}q_{10}}{q_{01}q_{10}} \cdot \frac{q_{11}q_{00}}{q_{11}q_{00}} \right) \\ &= \ln \left(\frac{r_1(1-r_0)}{r_0(1-r_1)} \cdot \frac{CPR_0}{CPR_1} \cdot \frac{q_{11}(1-q_{00})}{q_{01}(1-q_{10})} \right) \end{aligned}$$

From which it can be shown that

$$\beta_1 = \ln \left(\frac{r_1(1-r_0)}{r_0(1-r_1)} \cdot \frac{(1+Q_0)(1-Q_1)}{(1+Q_1)(1-Q_0)} \cdot \frac{q_{11}(1-q_{00})}{q_{01}(1-q_{10})} \right) \quad (2.47)$$

Equivalently,

$$\begin{aligned} \beta_2 &= \ln \left(\frac{q_{10}(1-q_{00})}{q_{00}(1-q_{10})} \right) = \ln \left(\frac{q_{10}(1-q_{00})q_{11}(1-q_{01})q_{01}(1-q_{11})}{q_{00}(1-q_{10})q_{11}(1-q_{01})q_{01}(1-q_{11})} \right) \\ &= \ln \left(\frac{CRP_0}{CRP_1} \cdot \frac{q_{11}(1-q_{01})}{q_{01}(1-q_{11})} \right) \end{aligned}$$

Rearranging the above equation gives the equation for β_2 in terms in terms of *Yule's Q* as follows

$$\beta_2 = \ln \left[\frac{(1 + Q_0)(1 - Q_1)q_{11}(1 - q_{01})}{(1 + Q_1)(1 - Q_0)q_{01}(1 - q_{11})} \right] \quad (2.48)$$

2.6.3 The Multivariate Normal k -regressor Case with Homogeneous Covariance

Consider the binary response variable Y with mean p and a vector of k explanatory variables denoted as \mathbf{X} . Assume that the inverse conditional distribution of \mathbf{X} given $Y = j$ is multivariate normal with homogenous covariance matrix, i.e.

$$f_{\mathbf{X}|Y}(\mathbf{X}_i, \theta_j) = (2\pi)^{-\frac{k}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{X}_i - \mu_j)' \mathbf{V}^{-1} (\mathbf{X}_i - \mu_j)\right\}$$

where μ_j is the vector of means for \mathbf{X}_i conditional on $Y_i = j$ for $j = 0, 1$ and V is the covariance matrix for \mathbf{X}_i . Given the predictor function takes the form:

$$\eta(\mathbf{X}_i; \beta) = \ln \left(\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i; \theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i; \theta_0)} \right) + \ln \left(\frac{p}{1-p} \right)$$

Then:

$$\begin{aligned} \ln \left(\frac{f_{\mathbf{X}|Y=1}(\mathbf{X}_i; \theta_1)}{f_{\mathbf{X}|Y=0}(\mathbf{X}_i; \theta_0)} \right) &= \ln \left(\frac{\exp\{-\frac{1}{2}(\mathbf{X}_i - \mu_1)' \mathbf{V}^{-1} (\mathbf{X}_i - \mu_1)\}}{\exp\{-\frac{1}{2}(\mathbf{X}_i - \mu_0)' \mathbf{V}^{-1} (\mathbf{X}_i - \mu_0)\}} \right) = \\ &= \ln \left(\frac{\exp\{-\frac{1}{2}(\mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_i - 2\mathbf{X}_i' \mathbf{V}^{-1} \mu_1 + \mu_1' \mathbf{V}^{-1} \mu_1)\}}{\exp\{-\frac{1}{2}(\mathbf{X}_i' \mathbf{V}^{-1} \mathbf{X}_i - 2\mathbf{X}_i' \mathbf{V}^{-1} \mu_0 + \mu_0' \mathbf{V}^{-1} \mu_0)\}} \right) \\ &= -\frac{1}{2}(-2\mathbf{X}_i' \mathbf{V}^{-1} \mu_1 + 2\mathbf{X}_i' \mathbf{V}^{-1} \mu_0 + \mu_1' \mathbf{V}^{-1} \mu_1 - \mu_0' \mathbf{V}^{-1} \mu_0) \\ &= [(\mu_1 - \mu_0)' \mathbf{V}^{-1}] \mathbf{X}_i + \left[\frac{1}{2} \mu_0' \mathbf{V}^{-1} \mu_0 - \frac{1}{2} \mu_1' \mathbf{V}^{-1} \mu_1 \right]. \end{aligned} \quad (2.49)$$

The predictor function in this case takes the form:

$$\eta(\mathbf{X}_i; \beta) = \beta_0 + \beta' \mathbf{X}_i. \quad (2.50)$$

Using equation (2.49), the coefficients will take the following parameterizations in terms of the parameters of the inverse conditional distributions:

$$\beta_0 = \ln\left(\frac{p}{1-p}\right) + \left[\frac{1}{2}\mu_0' \mathbf{V}^{-1} \mu_0 - \frac{1}{2}\mu_1' \mathbf{V}^{-1} \mu_1\right]; \text{ and}$$

$$\beta = [(\mu_1 - \mu_0)' \mathbf{V}^{-1}].$$

Now consider the case where each element of the random vector \mathbf{X}_i is standardized by dividing by its respective variance, so that $Var(X_{m,i}) = 1$ for $m = 1, \dots, k$. Then re-partition the covariance matrix \mathbf{V} , so that $\mathbf{V} = \begin{pmatrix} 1 & \rho_m' \\ \rho_m & P_{-m} \end{pmatrix}$, where $cov(X_m, X_s) = \rho_{m,s}$ since $Var(X_{m,i}) = 1$ for $m = 1, \dots, k$; ρ_m is the $(k-1 \times 1)$ vector of correlation coefficients between X_m and all the other explanatory variables; and P_{-m} is the remainder of the covariance matrix \mathbf{V} with 1s along the diagonal and correlation coefficients as the off-diagonal elements. The re-partition of \mathbf{V} allows us to isolate the parameterization of the coefficient β_m for $m = 1, \dots, k$. Following Spanos and McGuirk (2002) and invoking the use of the Schur lemma (p. 371):

$$\mathbf{V}^{-1} = \begin{pmatrix} \gamma^{-1} & -\gamma^{-1} \rho_m' P_{-m}^{-1} \\ -P_{-m}^{-1} \rho_m \gamma^{-1} & P_{-m}^{-1} \rho_m \gamma^{-1} \rho_m' P_{-m}^{-1} + P_{-m}^{-1} \end{pmatrix},$$

where $\gamma = 1 - \rho_m P_{-m}^{-1} \rho_m'$. Thus:

$$\beta = \begin{pmatrix} \beta_m \\ \beta_{-m} \end{pmatrix} = \begin{pmatrix} \gamma^{-1}(\mu_{1,m} - \mu_{0,m}) - \gamma^{-1}\rho'_m P_{-m}^{-1}(\mu_{1,-m} - \mu_{0,-m}) \\ -P_{-m}^{-1}\rho_m \gamma^{-1}(\mu_{1,m} - \mu_{0,m}) + [P_{-m}^{-1}\rho_m \gamma^{-1}\rho'_m P_{-m}^{-1} + P_{-m}^{-1}](\mu_{1,-m} - \mu_{0,-m}) \end{pmatrix}, \quad (2.51)$$

where $\mu_{j,m}$ denotes the m^{th} element of μ_j for $j=0,1$; and $\mu_{j,-m}$ represents the $(k-1 \times 1)$ vector of μ_j without $\mu_{j,m}$ for $j = 0,1$.

Examining equation (2.51), it becomes apparent that β_m is a function of the correlation coefficients (i.e. ρ_m) between X_m and all the other explanatory variables. Furthermore, it is affected by the correlation between the other explanatory variables via P_{-m}^{-1} , as well. As with the case with two binary variables, these correlations show up in both the numerator and denominator of these parameterizations. Thus, the **ceteris paribus** clause is not applicable in the k regressor case either. Furthermore, given the direct relationship between the binary and multivariate case shown here, the general findings from the simulations are likely applicable here, as well. That is, the magnitude of β_m and the $var(\beta_m)$ will likely follow each other (e.g. as β_m increases, $var(\beta_m)$ increases). Furthermore, the impact of near multicollinearity will likely be problem and data specific. That is, *a priori* knowledge of the correlations between explanatory variables may or may not lead to issues with systematic or erratic volatility.

All three extensions can be used to further examine other cases of near-multicollinearity in the logistic regression model. However, this is beyond the scope of this paper, and so we leave this for future work.

2.7 Conclusion

The traditional account of near-multicollinearity heavily relies upon a **ceteris paribus** clause which allows for the examination of the impacts $se(\hat{\beta})$, and $\tau(\hat{\beta})$ as $|\hat{\rho}| \rightarrow 1$, while

holding the parameter estimates $\hat{\beta}$ and $\hat{\sigma}^2$ constant. However, as shown by Spanos and McGuirk (2002) in the context of the linear regression model, $\hat{\beta}, \hat{\sigma}^2$ as well as R^2 and their associated variances and t – ratios are all functions of $\hat{\rho}$ and therefore change as $|\hat{\rho}| \rightarrow 1$. The changes in these statistics are rather different than the traditional account implies. Making use of *statistical reparametization*, Spanos and McGuirk (2002) show that contrary to the argument of the traditional account $se(\hat{\beta}), \tau(\hat{\beta})$ and associated statistics are non-monotonic functions of ρ , and therefore propose an ammendment to the traditional account of near multicollinearity.

Given the similarities between the logistic regression model and the linear regression model, it is no surprise that the traditional specification of near-multicollinearity in the logistic model closely mirrors the traditional specification in the linear regression model. In light of the work done by Spanos and McGuirk (2002), the primary objective of this paper has been to revisit the statistical specification of near-multicollinearity in the logistic regression model. Using the *PR approach*, we mathematically derive the parameters of the logistic regression model (β_i) and related statistics, and explicitly show that they are all functions of ρ . Monte carlo simulations confirm that the parameters, $se(\hat{\beta})$, as well as $\tau(\hat{\beta})$ all vary with ρ . The implications of our findings are that $\beta_i, se(\hat{\beta})$, and $\tau(\hat{\beta})$ fluctuate in a non-symmetric, non-monotonic way as $|\rho| \rightarrow 1$, and hence our analysis emphatically calls into question the much-invoked ceteris paribus clause to near-multicollinearity in the logit model.

We also consider what happens to traditional multicollinearity diagnostics such as the *VIFs* and κ . The usual account suggests that high near-multicollinearity results in high $R^2, VIF(\cdot)$, and κ . In other words, as the degree of multicollinearity increases ($|\rho| \rightarrow 1$), one expects monotonic, U-shaped graphs centered around $\rho = 0$. Our analysis finds no evidence of this. This conforms to what Spanos and McGuirk (2002) state, that “ (multicollinearity) diagnostics such as $VIF(\cdot)$ and κ can be completely

ineffective for detecting points of systematic volatility” (p 392).

What do these findings imply for future research? So far, we have only considered the simple 2 – *regressor* case of the logistic regression model with continuous (normal) covariates with different mean vectors, but homogeneous variance-covariance structure. In section 2.6, we extend the analysis to logistic regression models with continuous (normal) covariates, different mean vectors, but heterogeneous variance-covariance structure, as well as logistic regression models with binary covariates and nonlinear index/predictor functions. Our objective is to examine how nonlinearities affect the surfaces of relevant statistics in the presence of near-multicollinearity via simulations similar to those described in section 2.5. An equally important feature of the PR approach is that it allows for the parsimonious examination and specification of near-multicollinearity in logistic regression models with k regressors as shown in section 2.6.3.

Chapter 3

The Spatial Dynamics of Growth and Income Inequality: Empirical Evidence Using U.S. County Level Data.

3.1 INTRODUCTION

The connection between income inequality and economic growth has captured the attention of economists for over 50 years, since the seminal works of Kuznets (1955) and Kaldor (1957). Kuznets hypothesized, in what became known as the inverted-U hypothesis that as a country transitions from an agrarian sector to urban industrialization, inequality at first increases, and later decreases leading to an inverted-U relationship between growth and inequality. This empirical fact documented by Kuznets (1955) withstood the test of time until the late 1970s after which it failed to provide an explanation for the trend of rising inequality that characterized income distribution in the United States thereafter.

Many studies exist that look into the evolution of inequality and growth in the

U.S.¹. Justifications for the inverted-U relationship between inequality and growth – the notion that inequality initially rises, but then eventually declines as societies progress economically – are related to neoclassical arguments, which highlight a trend toward income convergence over time. Over the past three decades, however, the growth literature has increasingly focused on endogenous growth theories. Thus it is not surprising that the decrease in the significance of the inverted-U hypothesis coincides with the development of endogenous growth theories in the 1980s, which more often than not emphasize a trend toward income divergence over time – in other words, more income inequality.

The analysis of income inequality in the U.S. until recently has been limited to the state level. Data limitations have limited the extent to which such a study can be broadened to a more disaggregated level. Williamson (1965) studies income inequality at the regional level and shows that during the process of economic growth and development, societies go through three stages of development of interregional inequality. In the first stage, inequality increases, followed by a second stage of stable, yet high inequality, and a final stage of neoclassical convergence of incomes – thus decreasing inequality. Thus at the intra-country/interregional level, Williamson shows that inequality follows the predictions of Kuznets' inverted-U hypothesis.

This is a fascinating topic because various studies have found conflicting results when examining the inter-country and even inter-state relationship between inequality and growth. While it is the case that a significant number of these studies have found a negative relationship between inequality and growth (see Alesina and Perotti, 1994; Keefer and Knack, 2000; Persson and Tabellini, 1994), others have found a positive correlation between inequality and growth (see Partridge, 2005; Ngarambe, Goetz and Debertin, 1998). The fact that prior studies have found such different results gives

¹The next section provides a brief review of this literature

reason to believe that the more disaggregated nature of county-level data should provide more insights into this seemingly never-ending inequality-growth debate.

This paper differs from previous work on this subject area in two significant ways. First, unlike prior work which have focused on cross-sectional, country-level data, this article examines the correlation between income inequality and economic growth using a panel of income distribution data for 3,109 counties or county-equivalent administrative units of the U.S.². Together, these counties account for over 99 percent of all counties in the U.S. One advantage of using county-level data is that most of the measurement errors that usually plague cross-country studies, such as intertemporal and international comparability of data are reduced, since these data are collected and compiled using the same statistical techniques for the same time periods within the U.S. The non-spatial dynamic effects of inequality on growth are examined using the System Generalized Method of Moments (SYSGMM) suggested by Arellano and Bover (1995) and Blundell and Bond (1998).

The second contribution of this work is that the use of county-level data allows for the examination of the spatial interactions (or lack thereof) which may exist between counties. These spatial interactions are captured with the aid of a spatial contiguity weights matrix. We assume that dependence among counties only arises from the disturbance process, hence the estimation of a spatial error model. An unconditional maximum likelihood function similar to Elhorst (2005) is fully developed, and used to estimate the dynamic spatial panel regression model. The results from the non-spatial approach are compared to those of the spatial model, and trends or outliers are analyzed. Therefore, the consideration of county-level characteristics and data provide a more comprehensive analysis of the link between income inequality and economic growth, and pushes this inquiry one step closer toward more tenable results. This

²Louisiana is divided into parishes and Alaska into boroughs not counties. These are considered county equivalents

advantage allows for the examination of a richer set of potential reactions of growth to inequality and vice versa.

The results from the non-spatial dynamic models indicate that increases in income inequality are significantly related with decreases in long run economic growth. These results are shown to be robust across different model specifications and sample sizes. However, these results are biased since they do not correct for the spatial effects that may exist between counties. Incorporating spatial effects, the results from the spatial error model indicate that the bias in the parameter for inequality amounts to about 2.66 percent, while that for initial income amounts to about 21.51 percent. The bias in the parameter for education is relatively small (1.19 percent), while the difference in the parameter for wages between the spatial and non-spatial models is essentially zero.

The rest of the paper is structured as follows. The next section reviews the development of the major themes and provides a brief background of the approaches and results of the literature on the relationship between income inequality and economic growth. Section 3.3 provides a description of the data used and the methodology employed for the empirical analysis. Using the data from section 3.3, section 3.4 estimates and provides regression results for the non-spatial model, while section 3.5 estimates the spatial regression model. Finally, a concluding section attempts to identify trends and outliers, as well as unanswered questions and puzzles that arise.

3.2 Background and Literature Review

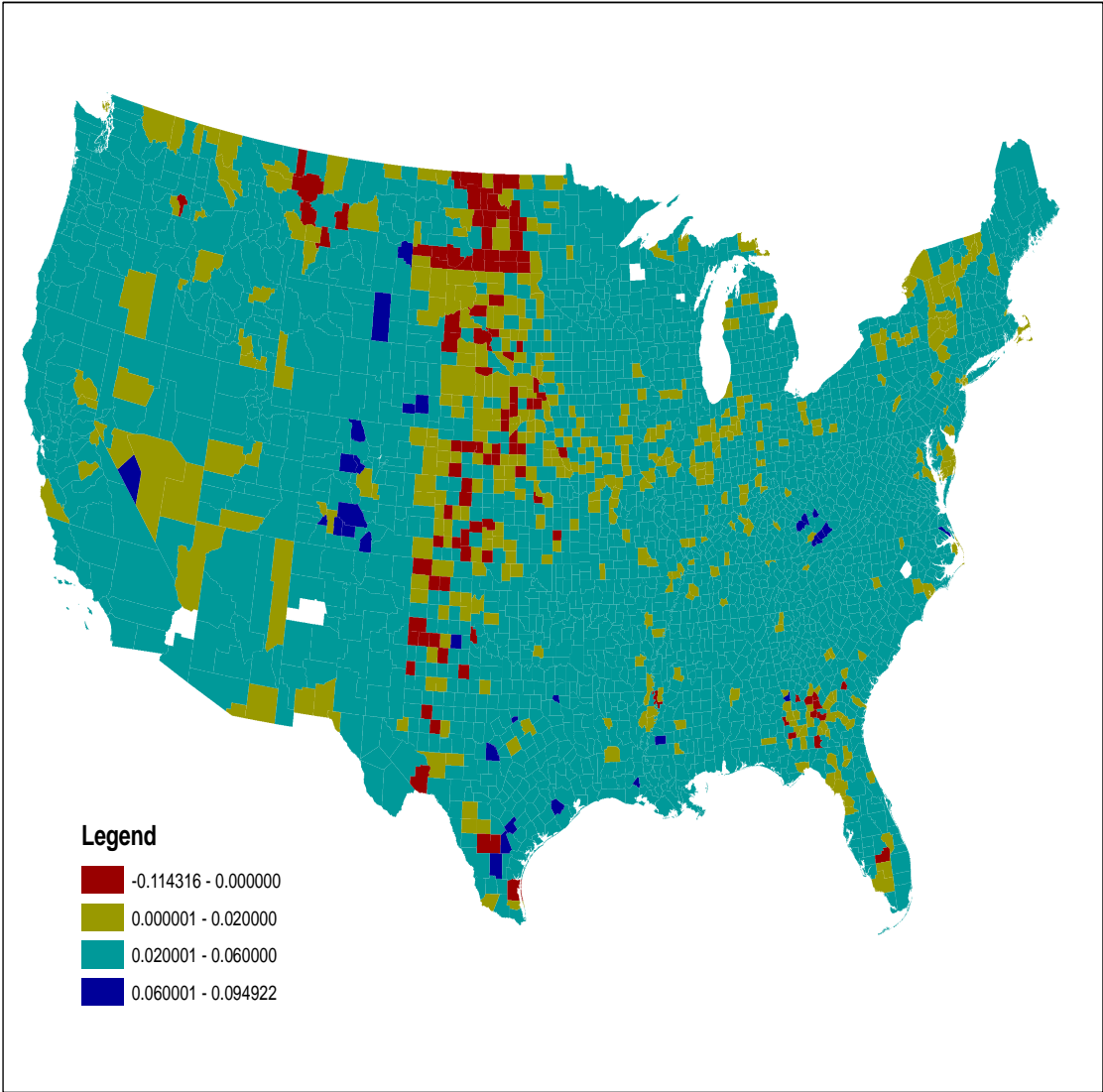
3.2.1 Trends in U.S. County-Level Growth and Inequality

Over the last forty years, per capita real GDP growth in the U.S. averaged approximately 2 percent per year. Although one would expect that this growth is necessarily good, it has been accompanied by a significant reduction in the share of income earned

by the bottom 90 percent of households. In fact, many families have seen little increases in their incomes, and these increases have been primarily the result of the growth of two-earner households. Figures 3.1 and 3.2 show the average annual per capita real GDP growth rates of U.S. counties in the 1970s and 2000s respectively.

A glimpse of Figure 3.1 shows that in the 1970s, real per capita GDP grew approximately 3 percent annually (between 0.02 – 0.06 in most regions). However, several

Figure 3.1: Spatial Variation in U.S. County-Level Per Capita Income Growth: 1970 – 1980



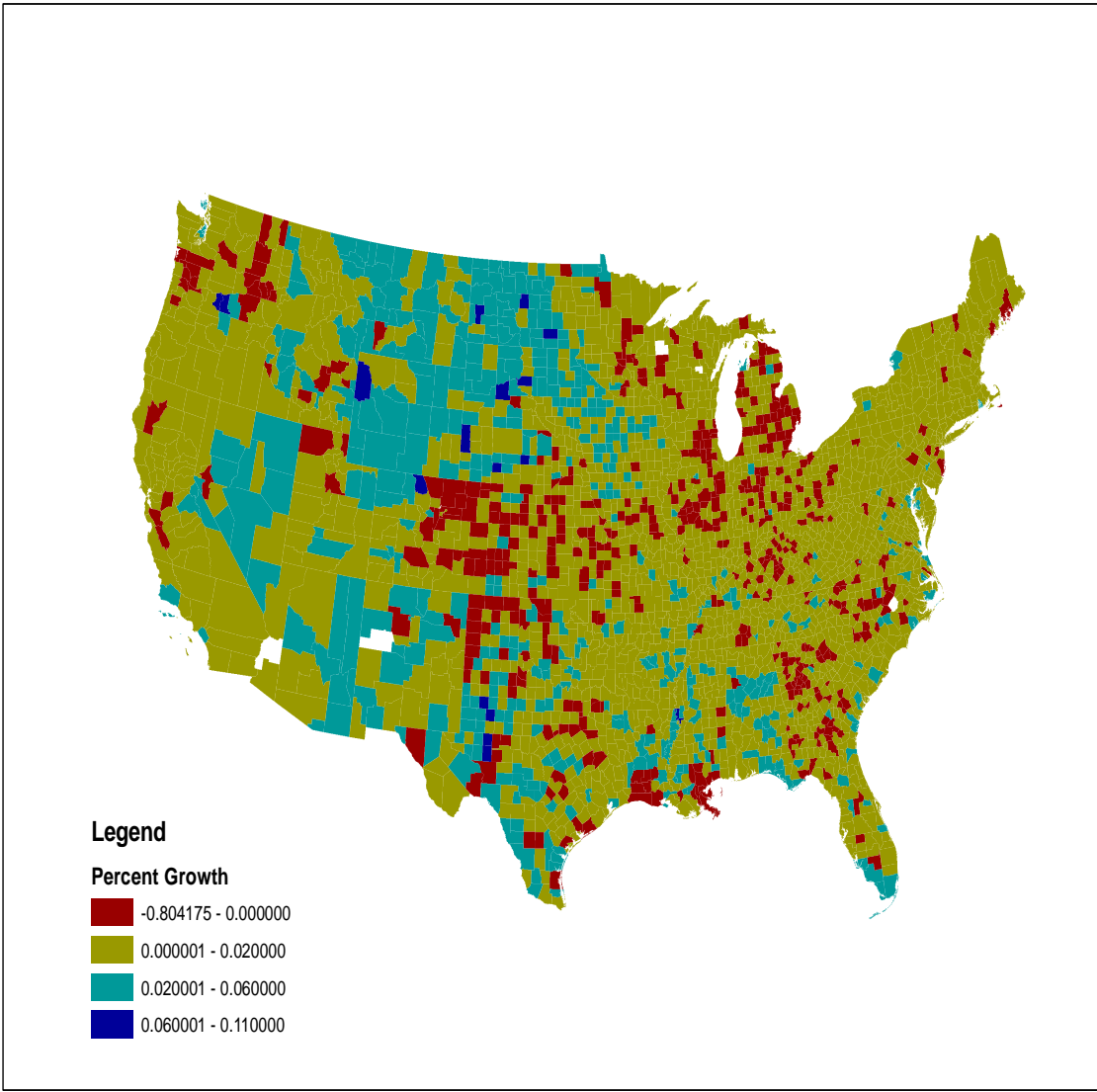
exceptions must be noted. A particularly striking feature of the map is that counties in the Great Plains region experienced relatively low growth. As a matter of fact, many of these counties experienced negative growth, with only a few showing modest positive growth (between 0 and 0.02). Many counties in the East and West coasts recorded positive growth rates. Nonetheless, overtime, this pattern would change dramatically.

Three decades later (Figure 3.2), while some counties in the Great Plains region still had low growth rates of per capita GDP, others had progressed in terms of growth. In the Great Plains region, except for counties in the central Midwestern states, many northern and southern counties experienced positive growth (between 0.02 - 0.06). On the other hand, the East and West Coasts, which had recorded positive per capita growth in the 1970s witnessed reduced growth rates (0.0 - 0.02) in the 2000s. Equally important to note are counties in the Gulf Coast region which saw declining per capita growth in the 2000s. Underlying these patterns are the immense differences between counties. Thus, relating these differences to county characteristics such as migration patterns, average wage per job per county, the degree of urbanization of counties, as well as the level of income inequality within counties, can prove relevant in helping to determine the sources of long run growth of these counties.

These trends seem to suggest a pattern of mobility and eventual convergence of real GDP per capita, thus signifying movements toward economic equality across U.S. counties. If this were indeed the case, one would expect that this growth would in fact eventually benefit most residents of these growing counties. However, if anything, the benefits of this growth seem to go disproportionately to the wealthy segment of the population, leading to more inequality overtime.

Figures 3.3 and 3.4 show the trends in income inequality in U.S. counties in 1970 and 2000. We use the Gini coefficient as our measure of inequality. This coefficient calculates the share of the area that encompasses the triangle defined by the line of

Figure 3.2: Spatial Variation in U.S. County-Level Per Capita Income Growth: 2000 – 2007



perfect equality and the line of perfect inequality of the Lorenz curve. The Gini coefficient lies between 0 and 1. A Gini coefficient of 0 implies equal income for all earners, while a Gini coefficient of 1 implies that one individual had all the income, while the rest of the population had nothing. Thus, higher Gini coefficients reflect more uneven distribution of incomes, and lower Ginis reflect societies characterized by more equal distributions of income and wealth (Atkinson, 1970).

Figure 3.3 shows that in the 1970s, in general, inequality was low (less than 0.38) across many U.S. counties. Nonetheless, several exceptions to this pattern are noticeable. High inequality (over 0.38) is present in counties in southern and Gulf states such as Arizona, New Mexico, and Texas, with particularly high inequality in Edwards (0.569) and Kenedy (0.579) counties in Texas. Isolated patches of high inequality are also apparent in the Dakotas, particularly in Dewey (0.54) and Shannon (0.52) counties of South Dakota and Emmons (0.45) county in North Dakota.

Figure 3.3: Spatial Variation in U.S. County-Level Income Inequality: 1970

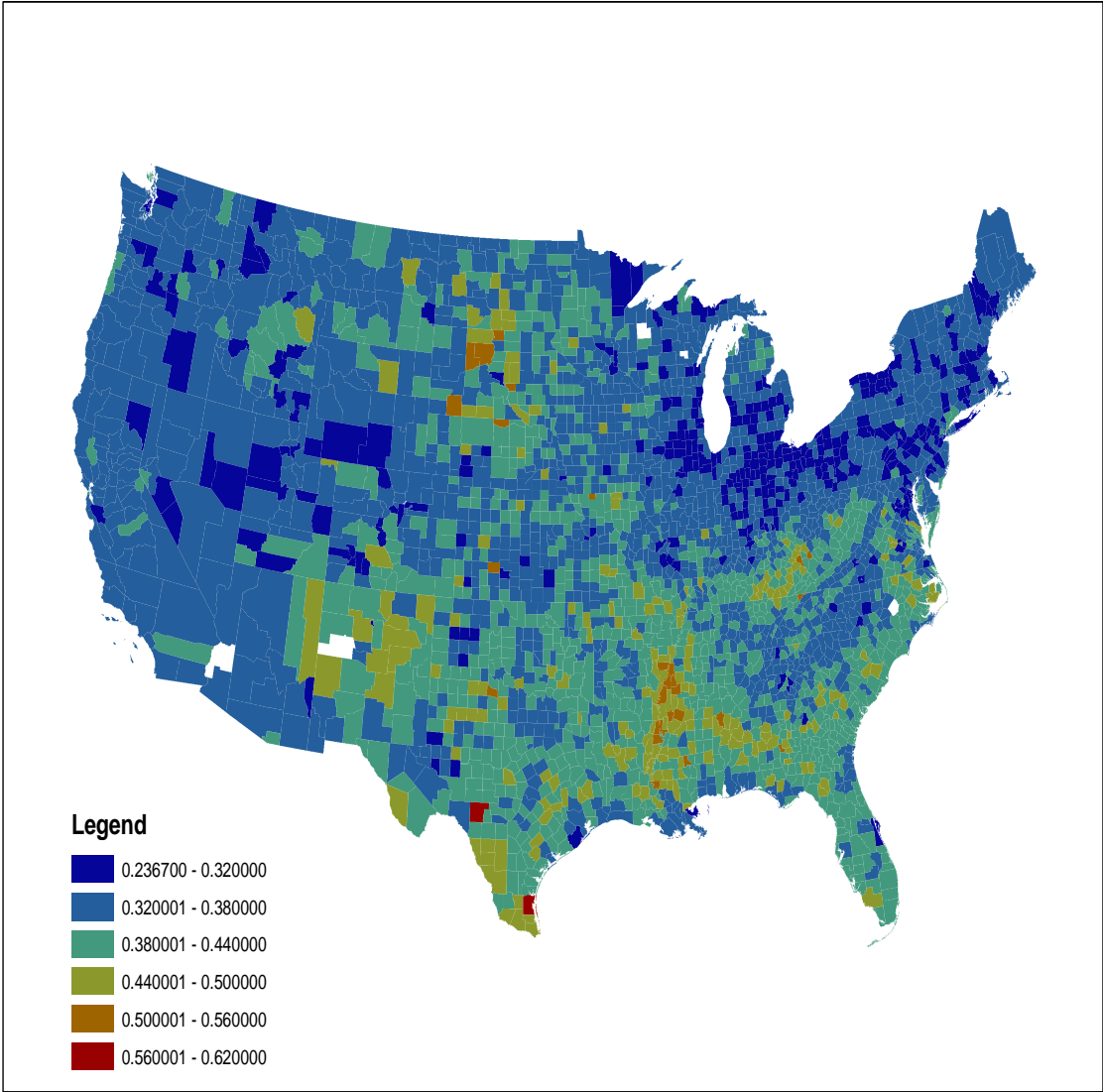
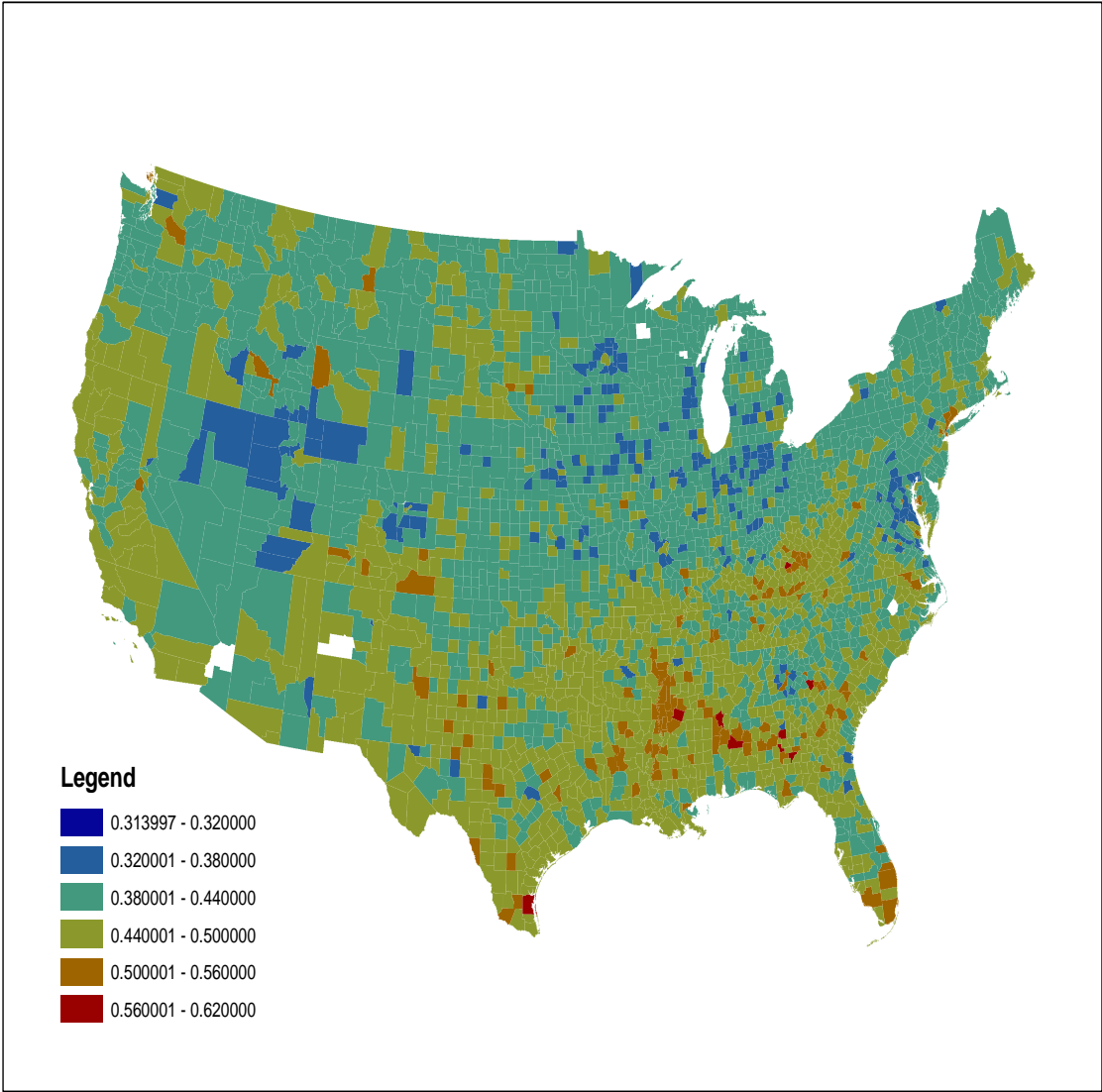


Figure 3.4 shows the distribution of family income thirty years later. Clearly, it is noticeable that by 2000, inequality had risen across many counties, even in the hitherto low-inequality counties. The West and Northeastern coasts, which in the 1970s recorded low inequality, were, by 2000, stricken by severe income inequality (greater than 0.44). Counties in the southern states, as well as those in the Gulf coast states which were already experiencing high inequality in 1970, saw higher income inequality in 2000.

Figure 3.4: Spatial Variation in U.S. County-Level Income Inequality : 2000



Several studies, using varying methods of analysis have sought to examine the effect that such changes in inequality might have on long run economic growth rates. A review of this literature is examined next.

3.2.2 Literature Review

The effects of income inequality on economic performance have been a contentious issue among economists and policymakers for a long time. Different models and methods of analysis, have yielded different results, sometimes sharply different, sometimes modestly. In this subsection, we provide a summary and a brief overview of the key approaches in the inequality-growth literature. In so doing, we try to be representative, rather than comprehensive, as we attempt to report completely on the findings and issues.

The genesis of this profoundly important debate can be traced back to the highly influential works of Kuznets (1955) and Kaldor (1957). The former shows that as an economy progresses economically, inequality initially rises, and then subsequently decreases, leading to an inverted-U process between inequality and economic growth. The latter study shows that the marginal propensity to save in wealthier economies exceeds that in poorer economies. Thus if GDP growth is positively related to the fraction of income that is saved, then economies with more unequal income distributions should grow unassailably more swiftly than their counterparts characterized by more equal income distributions.

Several studies, using varying techniques and approaches, seem to be in accordance with Kaldor's findings. In particular, Forbes (2000), using the data on inequality assembled by Deininger and Squire (1996) and an Arellano-Bond Generalized Method of Moments technique with various robustness tests, shows a positive and significant relationship between inequality and growth. Forbes, whose study focuses exclusively

on the short and medium terms, notes clearly that this sturdy positive relationship between inequality and growth could fade, and/or possibly reverse in the long run. Li and Zou (1998), who also use the aforementioned data assembled by Deininger and Squire (1996), estimate a regression using both fixed-effects and random-effects models. Given their model specifications, they find, in most cases, statistically significant evidence of a positive inequality-growth relation.

Nonetheless, unlike the studies mentioned above, other papers have found a significant negative relationship between income inequality and growth. Alesina and Rodrik (1994) use a cross-country political economy model to test whether initial inequality is statistically significant in forecasting long term growth. To do this, they regress, among other variables, the Gini coefficients for land and income on the average growth rate from 1960 to 1985³. They show that regardless whether the land Gini or the income Gini are entered into the regressions singly or collectively, a statistically significant negative correlation between inequality and growth is found. Alesina and Perotti (1996) analyze this inequality-growth relationship through the role of political instability on investment. They use the bivariate simultaneous equations model below:

$$INV = \alpha_0 + \alpha_1 SPI + \alpha_2 PRIM + \alpha_3 PPPIDE + \alpha_4 PPPI + \epsilon_1$$

$$SPI = \beta_0 + \beta_1 PRIM + \beta_2 GDP + \beta_3 INV + \beta_4 MIDCLASS + \epsilon_2$$

where, *INV*, *SPI*, *PRIM*, *PPPIDE*, *PPPI*, *GDP* and *MIDCLASS* are indexes for total investment; socio-political instability; primary school enrollment ratio in 1960; the PPP value of the investment deflator in 1960 relative to the U.S; the deviation of the PPPI relative to the sample mean; the initial per capita income; and the share of total income of the third and fourth quintiles; respectively. Using the above equations they conclude that “income inequality increases political instability, which in turn decreases investment. After an extensive battery of robustness tests, we can conclude that these

³These variables are the initial level of per capita income and the primary school enrollment ratio.

results in our sample of 70 countries are quite solid”. Keefer and Knack (2000) and Panizza (2002) also find a negative relationship between growth and inequality.

Barro (2000) challenges the orthodoxy of a negative relationship between growth and inequality. Using a broad panel of countries, he finds no clear general pattern between the two variables. More precisely, he shows that for rich countries – countries with per capita GDP over \$2000 (1985 U.S. dollars) – the inequality-growth relationship is positive, whereas for poor countries – those with per capita GDP below \$2000 – the relationship is negative. Nevertheless, the overall effects of inequality on growth and investment, Barro concludes, are weak.

The study of the relationship between inequality and growth is not limited to cross-national studies. Panizza (2002) points out that the problem with most cross-country studies is based on the quality and comparability of the inequality data. He argues that

even though the data set assembled by Deininger and Squire (1996) greatly improved the quality of the available data on income inequality, this data set is far from being problem free (page 16).

A possible solution to this problem proposed by Panizza (2002) is to use regional data. Using U.S. cross-state data for the period 1940 to 1980 and the GMM approach, he finds a negative, but non-robust correlation between the Gini index and regional growth.

Partridge (1997, 2007) also studies inequality and growth at the sub-national level. Similar to Panizza (2002), he emphasizes that redistribution policies that either enhance or inhibit growth need not only occur at the national level, but at the sub-national level, as well. In his 1997 paper, Partridge examines the inequality-economic growth relationship using data on the 48 contiguous U.S. states, and a model similar to Persson and Tabellini (1994) and Alesina and Rodrik (1994) briefly discussed above. Partridge finds, using the Gini coefficient as the inequality measure, a positive relation between initial inequality and subsequent economic growth in states with high inequality. However, he

also finds that in states in which the middle quintile's share of income is larger, growth subsequently tends to increase. The positive relationship is in stark contrast to the works of Persson and Tabellini (1994), and Alesina and Rodrik (1994), whose models Partridge extend. To reconcile these differences, Partridge concedes that a weakness of examining regional or state level data is that human and physical capital mobility could hinder the extent to which regional/state governments engage in income redistribution, making it necessary to take into account changes in migration patterns from high inequality-stricken states to low inequality-endowed states.

A limited number of within-country studies have been carried out at the county-level due limitations of existing data, or the unavailability of data altogether. However, the recent development of new datasets, and resourceful use of previously available datasets, allows for the possibility of measuring the county-level links between economic growth and income inequality. Ngarambe, Goetz, and Debertin (1998) is an examination of the joint determinants of inequality and economic growth for 1,257 counties in states in the U.S. for the 1970 and 1980 decades. They check for reverse causality between the two variables using a two stage least squares regression approach. Their empirical results show that in the 1970s, increased regional income disparities in the U.S. South significantly decreased growth, whereas in the 1980s, increased inequality complemented growth.

None of these studies, however, considers an explicit role of space on growth and inequality. Over the last few years, some in the profession have welcomed the view that the location of spatial units can determine their growth, and this is reflected in the general consensus that direct and indirect linkages between regions are crucial to the understanding of their growth dynamics and income distribution. Thus recently, the use of spatial econometric methods to analyze the role of location on growth is

gaining significant ground.⁴ This article is the first of its kind that uses detailed U.S. county-level panel data to explicitly uncover the role of geographical location on income inequality and economic growth.

3.3 Data and Methodology

The primary sources of data for the estimation of the model come from the U.S. Census Bureau and the Bureau of Economic Analysis Regional Economic Information System (BEA–REIS). The data set contains observations for 3,109 counties of the U.S. from 1970 - 2007. The growth model estimated considers the determinants of the average annual per capita income growth over 10-year horizons (except for the 2000-2007 period where the average annual growth rate is over a 7-year period) at the county level.

Data on per capita personal incomes and the ensuing growth rates come from the BEA – REIS. Following the convention in most economic growth and inequality studies, the average annual per capita income growth over ten-year periods is used as the dependent variable⁵. On the one hand, while this long-term perspective is constrained by data limitations, on the other hand, such a long-term perspective follows the norm of most growth studies that strive to explain long- rather than short-run variations and reduces annual serial correlation from business cycle fluctuations.

Before proceeding, it is worth pausing to note that this set-up (averaging over n periods) is not without problems. Attanassio, Picci and Scorcu. (2000) identify four drawbacks with this set-up. Firstly, they argue that annual data provide information that is lost when averaging. Their second criticism stems from the fact that because the length of business cycles varies over time and across space, and the interval over which these averages are computed is arbitrarily fixed, there is no guarantee that business

⁴Abreu et al. (2005) provide a comprehensive review of this literature.

⁵See Barro (2000) and Partridge (1997)

cycles are cut in the right way. In addition, averaging gets rid of the possibility of considering cross-sectional heterogeneity in the parameters. Finally, they insist that if averaging indeed measures the long-run effects, it prevents the analysis of short-run effects which usually include the interesting dynamic interplay of forces acting in opposite directions and/or different magnitudes. Hence averaging reduces, if not annuls the effects of such short-run dynamics.

The choice of independent variables that affect growth are well grounded in the literature, and their inclusion in the model shown hereinafter rests on the availability, and/or reliability of data for the counties of interest. These independent variables are dated at the start of each decade. This decreases the problem of endogeneity. Thus, by using lagged variables, there should not be any direct reverse causality issues between inequality and growth. By dating the variables at the start of each decade, this model also follows the convention in the literature that posits that economic growth converges to an equilibrium path based on initial conditions (Durlauf and Quah, 1999). To this end, the level of per capita income at the start of each decade is included as a regressor to account for convergence across counties (Barro and Sala-i-Martin, 1991, 1992; Goetz and Hu, 1996). The use of this variable makes it possible to examine if poorer counties were “catching-up” to wealthier ones during the period considered.

The data on county-level income inequality come from Nielsen (2002). Deininger and Squire (1996) assemble a cross-country “high quality” data set on inequality. They argue that any consistent and comprehensive data set on inequality must meet the following criteria: Firstly, the data must use households or individuals as the unit of observation; secondly, the sample must be representative of the population; and finally the inequality measures must be based on comprehensive coverage of different income and population groups. In this article, we argue that the Gini coefficients calculated by Nielsen (2002) meet these criteria in that Nielsen (2002) calculates Ginis based on

representative household incomes and income groups from Census data.

Other control variables considered include initial levels of human capital stock, and the average wage per job per county. The role of human capital in determining subsequent economic growth is theoretically and empirically well grounded in the literature, and so will not be emphasized here (see Romer, 1989; Barro, 2001). The percentage of college-educated individuals per county aged 25 years and older is used as a proxy for the stock of human capital in that county. County-level education data are obtained from the U.S. Department of Agriculture, Economic Research Service (USDA-ERS).

Regional wages and regional growth are also inextricably related. The average wage per job is included to capture labor market trends within counties. Wage data are collected from the BEA-REIS. We hypothesize on the one hand that if the average wage per job reflects labor costs, then wage rates inversely affect the growth of counties. On the other hand, relatively high wages or other factors that promote higher levels of labor compensation may push inefficient firms into becoming relatively more efficient. Furthermore, higher labor costs might lead to the adoption of new and better technology, which enhances productivity. The higher productivity due to higher levels of compensation permits the high-wage firm to stay competitive. Therefore, changes in productivity would merely compensate for, if not exceed labor costs. As such, higher wages might in fact enhance growth. Thus while it is known that local wages affect regional growth, the direction of this effect is not particularly well known. The model accounts for county-specific time-invariant characteristics as well as time-specific effects with the inclusion of county and period dummies respectively. To this end, the base growth model is expressed as:

$$growth_{it} = \alpha_1 y_{i,t-1} + \alpha_2 Ineq_{i,t-1} + \beta_1 Edu_{i,t-1} + \beta_2 Wage_{i,t-1} + \delta_t + \varphi_i + \varepsilon_{it} \quad (3.1)$$

where $y_{i,t-1}$ is the natural log of per capita income of county i in period $t - 1$. In this model, the left hand side is the average annual growth rate of per capita income of county i in period $t - 1$. $Ineq_{i,t-1}$, $Edu_{i,t-1}$, and $Wage_{i,t-1}$ are the Gini coefficient, education, and (the natural log of the) average wage per job, respectively. δ_t denotes the unobservable time effects, and φ_i accounts for time-invariant county effects and ε_{it} is the disturbance term. Table 3.1 provides the descriptive statistics, as well as a summary of the data sources and brief definitions of the variables in the model.

Table 3.1: Summary Statistics and Data Sources

Variables, Definitions and Sources	Year	Mean	Stdev	Min	Max
y : Real per capita personal income ^a	1970	9.289	0.237	8.421	10.283
	1980	9.827	0.232	8.642	10.813
	1990	9.957	0.218	8.949	11.165
	2000	10.121	0.226	9.019	11.457
$Ineq$: Gini coefficient of inequality ^b	1970	0.373	0.046	0.237	0.579
	1980	0.368	0.037	0.265	0.521
	1990	0.379	0.039	0.013	0.561
	2000	0.434	0.038	0.314	0.605
Edu : College completion rate ^c	1970	0.073	0.040	0.011	0.386
	1980	0.115	0.054	0.028	0.478
	1990	0.135	0.066	0.037	0.534
	2000	0.165	0.078	0.049	0.637
$Wage$: Average wage per county per job ^a	1970	9.765	0.222	9.000	11.105
	1980	10.147	0.212	9.561	11.560
	1990	10.089	0.207	9.533	11.205
	2000	10.194	0.203	9.621	11.315

Notes:

a. Bureau of Economic Analysis Regional Economic Information System (BEA – REIS)

b. Nielsen (2002)

c. United States Department of Agriculture Economic Research Service (USDA – ERS)

3.4 Estimation and Results

3.4.1 The Non-Spatial Dynamic Model

For the ease of the ensuing discussion, (3.1) can be rewritten as:

$$growth_{it} = \alpha_1 y_{i,t-1} + \alpha_2 Ineq_{i,t-1} + X'_{i,t-1} \beta + \delta_t + \varphi_i + \varepsilon_{it} \quad (3.2)$$

where $X'_{i,t-1}$ contains the set of control variables for the stock of human capital, and the average wage per job per county. This equation can be estimated using OLS, or the traditional fixed and random effects methods specified in the panel data literature. In fact, a Hausman specification test indicates that the random effects model is not appropriate in this case. More precisely, a χ^2 of 843.98 rejects the random effects model in favor of the fixed effects model at any level of significance. However, it is well known that $y_{i,t-1}$ may be correlated with the error term, ε_{it} and therefore fixed and random effects estimation lead to inconsistent results. In addition, the time-invariant county-specific characteristics φ_i may be correlated with the control variables, rendering OLS estimators biased. More so, Arellano and Bover (1995) show that the presence of the lagged endogenous variable causes the OLS estimate of this variable to be downward biased and inconsistent, even if the error terms are not serially correlated.

To avoid these problems, many dynamic panel studies of inequality and growth use the Arellano and Bond fixed effects GMM estimator, which uses lagged levels of all the control variables as instruments. Arellano and Bond (1991) show that for short dynamic panels ($N \rightarrow \infty$, and T is fixed), equation (3.2) is first differenced to eliminate the individual effects φ_i which are the major source of the bias in the OLS estimator. This gives:

$$y_{it} - y_{i,t-1} = \varsigma(y_{i,t-1} - y_{i,t-2}) + \alpha_2(Ineq_{i,t-1} - Ineq_{i,t-2}) + (\mathbf{X}'_{i,t-1} - \mathbf{X}'_{i,t-2})\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}) \quad (3.3)$$

where $\varsigma = (\alpha_1 + 1)$. It should be kept in mind that (3.3) controls for the time dummies by taking into account deviations of the variables from their period means. Arellano and Bond (1991) argue that a more consistent estimator can be obtained by using instruments whose validity is based on the orthogonality between lagged values of y_{it} and the error term ε_{it} . Thus when $t \geq 3$, the choice of the instruments depends on correlations between $y_{it} - y_{i,t-1}$ and each of $y_{it-1} - y_{i,t-2}$ and $\varepsilon_{it} - \varepsilon_{i,t-1}$. For example, when $t = 3$, equation ((3.3)) becomes,

$$y_{i3} - y_{i,2} = \varsigma(y_{i,2} - y_{i,1}) + \alpha_2(Ineq_{i,2} - Ineq_{i,1}) + (\mathbf{X}'_{i,2} - \mathbf{X}'_{i,1})\beta + (\varepsilon_{i,3} - \varepsilon_{i,2})$$

Therefore $y_{i,1}$ is a valid instrument for $y_{i,2} - y_{i,1}$, since they are correlated, and $y_{i,1}$ is not correlated with $\varepsilon_{i,3} - \varepsilon_{i,2}$, unless these errors are serially correlated. At $t = 4$, by the same argument, it can be shown that $y_{i,1}$ and $y_{i,2}$ are valid instruments. Also, since $Ineq_{i,t-1}$ and $\mathbf{X}'_{i,t-1}$ are predetermined, the values of these variables in the subsequent periods are correlated with the current error terms. That is, $E(Ineq_{i,t}\varepsilon_{it}) \neq 0$ and $E(\mathbf{X}'_{i,t}\varepsilon_{it}) \neq 0$ for $s > t$ and 0 otherwise. Consequently, in this case, at time s only $Ineq_{i,1}, \dots, Ineq_{i,s-1}$ and $\mathbf{X}'_{i,1}, \dots, \mathbf{X}'_{i,s-1}$ will be valid instruments in (3.3).

Blundell and Bond (1998) argue that using lagged levels are poor instruments for first differences. They show that as $\alpha_1 \rightarrow 0$ in (3.2), the Arellano and Bond GMM estimator performs poorly. To be precise, they observe that when T is small, the first differenced GMM estimator has a large downward finite sample bias. To this end, Blundell and Bond (1998) propose the use of the System GMM which estimates two

sets of equations - one set in levels that uses lagged first-differenced instruments, and another set in first differences that uses lagged-level instruments. Studies by Blundell and Bond (2000) and Blundell et al. (2000) show that the System GMM estimator for equations such as equation (3.2) outperforms the first differenced GMM estimator. To estimate the dynamic panel data model in this article, we therefore use the System GMM estimator. Table 3.2 presents the estimates of the System GMM (SYSGMM). For the purpose of comparability, the OLS, fixed effects (FE), random effects (RE), and the Arellano and Bond GMM (A&B GMM) estimates are also presented.

Table 3.2: Estimation Results for the Non-spatial Model

Model	OLS	RE	FE	A&B GMM	SYSGMM
$y_{i,t-1}$	-0.0409 (0.0008)	-0.0121 (0.0009)	-0.0519 (0.0020)	-0.3636 (0.0255)	-0.3518 (0.0264)
$Ineq_{i,t-1}$	-0.0342 (0.0038)	-0.0039* (0.0043)	-0.0511 (0.0062)	-0.0561 (0.0137)	-0.0492 (0.0140)
$Edu_{i,t-1}$	0.1100 (0.0035)	0.0595 (0.0032)	0.1237 (0.0082)	0.0641 (0.0208)	0.0762 (0.0207)
$Wage_{i,t-1}$	-0.0104 (0.0010)	0.0015* (0.0010)	-0.0027* (0.0022)	-0.0196 (0.0045)	-0.0251 (0.0041)
R^2	0.3294	0.4798	0.5559		
<i>Observations</i>	12430	12430	12430	6216	6216

Notes:

- i.* * implies insignificant at the 5 percent level.
- ii.* For the fixed effects model, R^2 is the within- R^2 ; the random effects model, R^2 is the overall- R^2 .
- iii.* The standard errors are reported in parentheses. These errors are robust to heteroskedasticity of unknown form.
- iv.* The GMM estimates reported are all two step.

Table 3.2 shows that regardless of the estimation technique, changes in inequality have significant impacts on the long-run growth dynamics of U.S. counties. In particular, since we are interested in the estimates from the SYSTEM GMM estimator, the coefficient on inequality, - 0.0492 (and standard error of 0.0140) indicates that there exists a statistically significant negative relationship between inequality at the start of each decade, and growth throughout the decade, thus confirming the hypothesis that income redistribution policies that seek to reduce inequality will have a significant pos-

itive relationship on the growth of a county. This finding is in accordance with some of the cross-country, as well as the within-country studies mentioned above.

Following the predictions of most theoretical models and prior empirical findings in the literature, the significantly negative coefficient estimate of - 0.3518 for initial income is evidence of convergence of incomes across U.S. counties. In addition, the coefficient on education, used as a proxy for the stock of human capital, is positive and significant as expected; indicating that an increase in the educated population of a county is correlated with faster growth of that county. Wages are also shown to be negatively related to economic growth.

3.4.2 Robustness of the Non-Spatial Model

The results suggested in Table 3.2 indicate a strong negative relationship between changes in inequality and changes in economic growth across various model specifications, but more importantly so for the system GMM model. This negative sign of inequality coefficient is in contrast with some of the well cited panel data studies of inequality and growth, notably Forbes (2000) and Li and Zou (1998). As such, we perform a set of sensitivity analyses to check if these results are robust across different samples. The results of our sensitivity analyses are presented in Table 3.3. First, we re-estimate the SYSGMM model without the inclusion of the period dummies. SYSGMM 1 shows the estimates from this model specification. The coefficient of lagged income is still negative and statistically significant, providing support for hypothesis of convergence across counties shown in Table 3.2. The coefficient of initial inequality is still negative and significant, while that for college education is positive and significant as expected. In this specification, the coefficient on wages wages are still negative and significant.

Next, we drop the wage variable from the model, and assume a dynamic model

in which growth is determined only by the initial levels of income, inequality, and education. The period dummies are also included. Column 2 (SYSGMM 2) shows the estimates from this specification. The results are again significant, and of the same signs as before. To test how steady these results are, an interaction term between inequality and education is included (SYSGMM 3). The results from this specification are slightly different. While the coefficient on inequality is still negative and significant, that for education is now negative. More surprising is the negative and significant coefficient of the interaction term between inequality and education of - 0.6230. This result suggests that counties that strive to enhance their growth by implementing policies aimed at better education should attempt to adopt income redistribution policies that decrease inequality as well.

Table 3.3: Sensitivity Analysis

Model	SYSGMM 1	SYSGMM 2	SYSGMM 3	SYSGMM 4
$y_{i,t-1}$	-0.1091 (0.0102)	-0.3416 (0.0277)	-0.3408 (0.0279)	-0.3515 (0.0267)
$Ineq_{i,t-1}$	-0.0299 (0.0098)	-0.0479 (0.0142)	-0.1453 (0.0231)	-0.1436 (0.0229)
$Edu_{i,t-1}$	0.0400 (0.0157)	0.0544 (0.0201)	0.2341 (0.0560)	0.2036 (0.0561)
$Edu * Ineq_{i,t-1}$			-0.6425 (0.1308)	-0.6230 (0.1300)
$Wage_{i,t-1}$	-0.0099 (0.0035)			-0.0255 (0.0041)
<i>Observations</i>	9325	9325	9325	9325

Notes:

SYSGMM 1. Estimates the SYSGMM model without the period dummies

SYSGMM 2. Includes period dummies; the wage variable is dropped.

SYSGMM 3. Includes period dummies and interaction between inequality and education.

SYSGMM 4. All variables and period dummies included.

One of the major criticisms of the models presented thus far is that they do not account for the spatial relationships that exist between counties. The space-specific time-invariant variables that affect economic growth and inequality can differ significantly across counties, and failure to incorporate these variables into a model of growth

and inequality may lead to biased outcomes. One way to attempt to capture the role of space was the introduction of the county-specific dummies, φ_i . However, if the data in the model are pooled incorrectly, then the county-specific dummies cannot correct for geographic differences entirely. Moreover, first-differencing essentially eliminates these county-specific, time-invariant county characteristics. Consequently, there is need to explicitly suggest an alternative specification that corrects for spatial disparities between counties.

In addition, the use of predetermined variables can be problematic because these variables are either chosen based almost entirely on data availability. However, it is known that the process of growth and inequality started long ago (m periods ago, $m \rightarrow \infty$) for which data are unavailable. Therefore, there is need to specify a growth – inequality model that not only includes the explicit role of space, but indeed incorporates procedures that make it possible for researchers to test the appropriateness of their assumptions about the initial observations as well.

3.5 The Role of Space in the Panel Growth Model

This section examines, using a spatial econometric technique, the correlation between inequality in a county with a certain level of geographic proximity to other counties and the economic growth process in these proximate counties. Inequality in one county can affect the economic growth process in a neighboring county through several channels, such as through the movement of labor, goods, services, and technology from one county to its neighbor.

Despite the fact that the study of dynamic panel models has received considerable attention lately, an extension of these models to include spatial effects has been quite limited due to their difficulty of implementation using the current econometric software

packages. Following Tobler’s First Law of Geography which states that “everything is related to everything else, but near things are more related than distant things”; it makes sense, at least from an intuitive point of view, to consider the role of space when examining the inequality-growth relationship at the county level. In what follows, a brief general review of the construction of the index of spatial dependence – the so-called spatial weights matrix – is presented, followed by its implementation in the inequality growth model.

3.5.1 The Spatial Weights Matrix

The proximity between the counties is measured using a spatial weights matrix, denoted W . W is an $N \times N$ matrix, where N represents the number of geographical units (counties) in space. A row and column exists for each county and the value in each cell represents the spatial proximity between that county and another. The weights of the matrix can be binary (1 or 0) if the units are spatially contiguous, or they can take continuous values if, for example, an inverse distance measure between the two counties is used. Thus this matrix allows for the consideration of the level of spatial interaction (spatial autocorrelation) and spatial structure (spatial heterogeneity) between any two or more counties. In other words, each element of the matrix, $\omega_i(d_{il})$, measures the spatial relationships between locations i and l , and can be interpreted as the effect of a variable in location i on location l . The matrix can be more formally presented as:

$$W = \begin{bmatrix} 0 & \omega_i(d_{12}) & \cdots & \omega_i(d_{1N}) \\ \omega_i(d_{21}) & 0 & \cdots & \omega_i(d_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ \omega_i(d_{N1}) & \omega_i(d_{N2}) & \cdots & 0 \end{bmatrix} \tag{3.4}$$

where ω_i , ($i = 1, \dots, N$) denotes the (known) real characteristic roots of W , and d is equal to 1 or 0 if the counties are contiguous. The proper dependence representation of W has been problematic. Consequently, many spatial econometricians have resorted to a W which is most empirically expedient or conventional. However, because misspecification of W can distort the spatial analysis, thus affecting the Maximum Likelihood Estimators (MLE), a review of the literature on the specification of W in spatial econometric models by Griffith (1996) provides some rules of thumb for specifying W . He concludes that:

1. "It is better to posit some reasonable geographic weights matrix specification than to assume all entries are zero (the independent observations situation of conventional statistics), the extreme case of under-specification". In other words, it is important to hypothesize a suitable W , rather than attempt to ignore the role of space, because such under-specification might smother the standard errors of an estimator, yet increase the mean squared error for the model.
2. "It is better to use a surface partitioning that falls somewhere between a regular square and a regular hexagonal tessellation".
3. "Relatively large numbers of areal units should be employed in the spatial statistical analysis". Griffith suggests $N > 60$. Simulation experiments by Stetzer (1982) show that small sample sizes magnify misspecification problems of W , while large sample sizes decrease misspecification errors.
4. "Low-order spatial statistical models should be given preference over high-order ones".
5. "In general, it is better to employ a somewhat under-specified than a somewhat over-specified geographic weights matrix" (p. 355).

Based on these simple rules of thumb, a spatial contiguity matrix seems appropriate for this study. N is large ($N = 3,109$) so that the spatial contiguity matrix used is 3109×3109 , and a spatial error model of order 1 is employed for the analysis.

3.5.2 The Dynamic Spatial Panel Growth Model

We present the dynamic model to include spatial error autocorrelation. The spatial error autocorrelation model is chosen against the spatial lag model because the spatial error autocorrelation model reduces the negative impacts of spatially correlated omitted variables. This model also corrects for the bias that results from the use of spatially correlated aggregate variables, as well as spatially correlated measurement errors that usually plague most empirical studies. Moreover, Lesage and Pace (2009) note that the spatial error model has an expectation equal to that of the conventional regression model where independence between the dependent variable observations is part of the maintained hypothesis. The spatial lag model on the other hand assumes dependence between the variables, making the coefficients of the spatial lag model difficult to interpret. Lesage and Pace (2009) also argue that in small samples, there may be an efficiency gain from correctly modeling spatial dependence in the disturbance process. To this end, assume now that

$$\varepsilon_{it} = \rho W \varepsilon_{it} + u_{it} = (I_N - \rho W)^{-1} u_{it} \quad (3.5)$$

where ρ is the coefficient in the spatial autoregressive structure for the disturbance term ε_{it} and $u_{it} \sim N(0, \sigma^2 I_N)$. The model in equation (3.2) can then be rewritten to include spatial error autocorrelation as follows:

$$growth_{it} = \alpha_1 y_{i,t-1} + \alpha_2 Ineq_{i,t-1} + X'_{i,t-1} \beta + \delta_t + \varphi_i + (I_N - \rho W)^{-1} u_{it} \quad (3.6)$$

Taking first differences and controlling for the time dummies by taking deviations of the variables from their period means, the dynamic model can now be written as:

$$y_{it} - y_{i,t-1} = \varsigma(y_{i,t-1} - y_{i,t-2}) + \alpha_2(Ineq_{i,t-1} - Ineq_{i,t-2}) + (\mathbf{X}'_{i,t-1} - \mathbf{X}'_{i,t-2})\beta + (I_N - \rho W)^{-1}(u_{it} - u_{it-1}) \quad (3.7)$$

The model in equation (3.7) can be estimated by GMM using appropriate instruments. However, Elhorst (2003) notes that a major drawback to the GMM approach is that it tends to overestimate the coefficient of the spatial autoregressive parameter ρ , since ρ is unbounded from above using GMM. To estimate (3.7), an unconditional maximum likelihood estimation approach is used (Hsiao, Pesaran and Tahmiscioglu, 2002 and Elhorst, 2005). Hsiao, Pesaran and Tahmiscioglu (2002) show that MLE tends to dominate the GMM approach in terms of the bias and root mean squared errors of the estimators as well as the size and power of the test statistics. They also show that as N increases, the MLE are more consistent and efficient than the GMM estimators. Before proceeding with the analysis, it is imperative that the Maximum Likelihood function be presented.

3.5.3 The Likelihood Function.

It is well grounded in the literature that the likelihood function for dynamic panel models is contingent upon the assumptions made about the initial observations (Hsiao 2003). The predetermined explanatory variables are assumed to be generated by a stationary process. This assumption is safe because first differencing transforms the process from a non-stationary to a stationary one. For simplicity, equation (3.7) is now written as:

$$\Delta y_{it} = \varsigma \Delta y_{i,t-1} + \alpha_2 \Delta Ineq_{i,t-1} + \Delta \mathbf{X}'_{i,t-1} \beta + (I_N - \rho W)^{-1} \Delta u_{it} \quad (3.8)$$

(3.8) is well defined for Δy_{it} for $t = 2, 3, \dots, T$ but not for Δy_{i1} because Δy_{i0} is not observed. In other words, the absence of pre-sample data for the explanatory and explained variables needed for first differencing the observations for each county implies that Δy_{i0} cannot be observed. In order to derive the likelihood function for the entire sample, it is imperative that the probability function for Δy_{i1} be specified as well. Following Hsiao, Pesaran and Tahmiscioglu (2002) and Elhorst (2003, 2005), the following assumptions are made about Δy_{i1} :

1. The control variables $Ineq_{i,t-1}$ and $\mathbf{X}'_{i,t-1}$ that affect Δy_{i1} are “weakly exogenous” forcing variables; and
2. Either $|\varsigma| < 1$, and the dynamic process started sometime in the past, say m periods ago. This assumption ensures that: $E(\Delta y_{i1}) = 0$, $Var(\Delta y_{i1}) = \sigma_u^2 2 / (1 + \varsigma)$ for $t = 3, 4, \dots, T$ and $i = 1, 2, \dots, N$. Or
3. The dynamic growth process started sometime in the past, not too far from the 0^{th} period and the expected changes in the initial observations are the same for all $i = 1, 2, \dots, N$. In this case, m is finite and so it can be shown that $E(\Delta y_{i1}) = \pi_{i0} I_N$, and $Var(\Delta y_{i1}) = \sigma_u^2 \frac{2}{(1+\varsigma)} (1 + \varsigma^{2m-1})$, for $t = 3, 4, \dots, T$ and $i = 1, 2, \dots, N$. π_{i0} is an unknown and fixed parameter to be estimated, and I_N is the identity matrix.⁶ By recursive substitution, Δy_{i1} can be written as:

$$\begin{aligned} \Delta y_{i1} = & \varsigma^m \Delta y_{i,t-(m-1)} + \sum_{j=0}^{m-1} \varsigma^j \alpha_2 \Delta Ineq_{i,t-(m-1)} \\ & + \sum_{j=0}^{m-1} \varsigma^j \Delta (\mathbf{X}'_{i,t-(m-1)}) \beta + \sum_{j=0}^{m-1} \varsigma^j (I_N - \rho W)^{-1} \Delta u_{i,t-j} \end{aligned} \quad (3.9)$$

⁶ π_{i0} is identical for all $i = 1, 2, \dots, N$, and so π_0 is used in the remainder of the paper. Similarly for $\pi_{i,1}, \pi_{i,2}, \dots, \pi_{i,T}$.

Given the assumption that $Ineq_{i,t-1}$ and $\mathbf{X}'_{i,t-1}$ are generated by a stationary process, as well as assumptions 2 or 3, it follows that $E(\Delta Ineq_{i,t-1}) = 0$ and $E(\Delta \mathbf{X}'_{i,t-1}) = 0$. Consequently, $E(\Delta y_{i1}) = \varsigma^m \Delta y_{i,t-(m-1)}$. However, since the observations for $\Delta Ineq_{i,t-1}$ and $\Delta \mathbf{X}'_{i,t-1}$ for $t = 1$ are not observable, $Var(\Delta y_{i1})$ is unknown, and so is the probability function of Δy_{i1} . Various techniques for approximating the optimal $\Delta Ineq_{i,t-1}$ and $\Delta \mathbf{X}'_{i,t-1}$ for $t = 1$ have been suggested.⁷ In this article, we follow the approach proposed by Bhargava and Sargan (1983).

Consistent with Bhargava and Sargan (1983), since the model contains vectors of $px1$ time-varying explanatory variables, under the assumption that changes in the initial observations are the same for all counties such that $E(\Delta y_{i1}) = \pi_0 I_N$, ($i = 1, 2, \dots, N$), it follows that the predictor of $\Delta Ineq_{i,t-1}$ and $\Delta \mathbf{X}'_{i,t-1}$ is $\pi_0 I_N + \pi'_t (\Delta Ineq_{i,t-1} + \Delta \mathbf{X}'_{i,t-1}) + \xi$ for $(t = 1, 2, \dots, T)$. In this set-up, $\xi \sim N(0, \sigma_\xi^2 I_N)$, π_0 is a constant associated with the mean of Δy_{i1} and π_t is the $(pT + 1) \times 1$ vector of parameters. By construction, we can now show that

$$\begin{aligned} \Delta y_{i1} = & \pi_0 I_N + \Delta Ineq_{i0} \pi_1 + \Delta \mathbf{X}'_{i0} \pi_1 + \dots \\ & + \Delta Ineq_{i,T-1} \pi_T + \Delta \mathbf{X}'_{i,T-1} \pi_T + \Delta u_{i1} \end{aligned} \quad (3.10)$$

where $\Delta u_{i1} = \xi + \sum_{j=0}^{m-1} \varsigma^j (I_N - \rho W)^{-1} \Delta u_{i,1-j}$. Under the assumption that $Ineq_{i,t-1}$ and $\mathbf{X}'_{i,t-1}$ are “weakly exogenous” forcing variables, it follows then that $E(\Delta u_{i1}) = 0$, $E(\Delta u_{i1} \Delta u'_{i2}) = -\sigma^2 (I_N - \rho W)^{-1} ((I_N - \rho W)')^{-1}$, $E(\Delta u_{i1} \Delta u'_{it}) = 0$ for $t = 3, 4, \dots, T$ and

⁷See for example Bhargava and Sargan (1983), Blundell and Smith (1991), Ridder and Wansbeek (1989), and Nerlove and Balestra (1996).

$$\begin{aligned}
E(\Delta u_{i1} \Delta u'_{i2}) &= \sigma_{\xi}^2 I_N + \sigma^2 \left\{ \frac{2}{1+\zeta} (1 + \zeta^{2m-1}) \right\} ((I_N - \rho W)^{-1} ((I_N - \rho W)')^{-1}) \\
&\equiv \sigma^2 (I_N - \rho W)^{-1} \{ \lambda^2 \sigma^2 (I_N - \rho W)^{-1} ((I_N - \rho W)')^{-1} \\
&+ \left[\frac{2}{1+\zeta} (1 + \zeta^{2m-1}) \right] I_N \} ((I_N - \rho W)')^{-1}
\end{aligned} \tag{3.11}$$

where $\lambda^2 = \sigma_{\xi}^2 / \sigma^2$. In order to correctly identify the parameters in (3.10), $N > pT + 1$, without which the number of parameters used to derive the matrix of the pre-sample distribution of the initial values must be reduced. Also note that since $|\zeta| < 1$, $\lim_{T \rightarrow \infty} \left\{ \frac{2}{1+\zeta} (1 + \zeta^{2m-1}) \right\} = \frac{2}{1+\zeta}$, which is not a free parameter, in which case the estimation of (3.11) is not possible. Therefore, to estimate equation (3.11), we make use of assumption 3.

Denote by ϑ , the $N \times N$ matrix given by $\vartheta = \lambda^2 (I_N - \rho W)^{-1} ((I_N - \rho W)')^{-1} + \left[\frac{2}{1+\zeta} (1 + \zeta^{2m-1}) \right] I_N$. Also denote by H_{ϑ} , the $NT \times NT$ matrix given by

$$H_{\vartheta} = \begin{pmatrix} \vartheta & -I_N & 0 & \cdots & 0 & 0 & 0 \\ -I_N & 2I_N & -I_N & \cdots & 0 & 0 & 0 \\ 0 & -I_N & 2I_N & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2I_N & -I_N & 0 \\ 0 & 0 & 0 & \cdots & -I_N & 2I_N & -I_N \\ 0 & 0 & 0 & \cdots & 0 & -I_N & 2I_N \end{pmatrix} \tag{3.12}$$

Then the covariance matrix of Δu_{it} can be written as:

$$Var(\Delta u_{it}) = \sigma^2 \{ [I_T \otimes (I_N - \rho W)^{-1}] H_{\vartheta} [I_T \otimes ((I_N - \rho W)')^{-1}] \} \tag{3.13}$$

Now define $\theta = (\alpha'_2, \beta', \varsigma, \pi')'$, $\gamma = (\rho, \lambda, \frac{2}{1+\varsigma}(1 + \varsigma^{2m-1}))'$, and $\zeta = (\theta, \sigma^2, \gamma')'$. Then using the properties of equations (3.5) and (3.12), Elhorst (2003, 2005) shows that the log-likelihood function for estimating the ML estimator of θ is given by:

$$\begin{aligned}
\log L(\zeta) &= -\frac{NT}{2} \log(2\pi\sigma^2) + \sum_{i=1}^N \log(1 - \rho\omega_i) \\
&- \frac{1}{2} \sum_{i=1}^N \log\left\{1 - T + \frac{2T}{1+\varsigma}(1 + \varsigma^{2m-1}) + T\lambda^2(1 - \rho\omega_i)^2\right\} \\
&- \frac{1}{2\sigma^2} \Delta u_{it}(\theta)' H_{\theta}^{-1} \Delta u_{it}(\theta)
\end{aligned} \tag{3.14}$$

where

$$\Delta u_{it}(\theta) = \begin{pmatrix} (I_N - \rho W)(\Delta y_{i1} - \pi_0 I_N - \Delta Ineq_{i0} \pi_1 - \Delta \mathbf{X}'_{i0} \pi_1 - \dots) \\ (I_N - \rho W)(\Delta y_{i2} - \varsigma \Delta y_{i1} - \alpha_2 \Delta Ineq_{i1} - \Delta \mathbf{X}'_{i1} \beta) \\ \vdots \\ (I_N - \rho W)(\Delta y_{iT} - \Delta y_{i,T-1} - \alpha_2 \Delta Ineq_{i,T-1} - \Delta \mathbf{X}'_{i,T-1} \beta) \end{pmatrix}$$

To derive the profile likelihood function, the parameters of θ must be estimated by concentrating out these parameters from their first-order conditions. An iterative procedure for the numerical solution of the maximization problem must be used to get these parameters. Also, since an appropriate value of m is needed, we set $m = 1$. This implies that any differences between the results from this specification and the specification from column 6 (SYSGMM) of Table 3.2 are the result of spatial effects. Table 3.4 presents the results from this specification.

Table 3.4: Estimation Results for the Spatial Dynamic Model

Variable	$y_{i,t-1}$	$Ineq_{i,t-1}$	$Edu_{i,t-1}$	$Wage_{i,t-1}$
Coefficient	-0.1367 (0.0277)	-0.0758 (0.0120)	0.0881 (0.0193)	-0.0162 (0.0042)

NOTES:

i. The standard errors are reported in parentheses. These errors are robust to heteroskedasticity of unknown form.

As noted earlier, the estimates of the SYSGMM approach are biased because they do not incorporate the explicit role that spatially localized patterns have on income distribution and growth dynamics. Table 3.4 shows the magnitude of this bias. The bias in the parameter for inequality is about 2.66 percent, while that for initial income amounts to about 21.51 percent. The bias in the inequality parameter is relatively small because even though T is short, N is relatively large. Education is still positively correlated with growth, although, the bias in the parameter for this variable is relatively small (1.19 percent). The result for wages is virtually unchanged.

3.6 Conclusion

This chapter reconsiders the correlations between economic growth and income inequality using U.S. county-level data. We first examine the non-spatial dynamic relationship between inequality and growth. Contrary to previous studies that use dynamic panel data techniques such as Forbes (2000) and, Li and Zou (1998), we find a significant negative relationship between changes in the gini coefficient of inequality and changes in the growth rates of counties in subsequent periods. We also find that this relationship is stable across different sample sizes and model specifications.

However, we argue that the results reported using the non-spatial dynamic panel specification may be biased because the geographical location of counties can prove fundamental in determining their growth dynamics and income distribution. As such, we specify a spatial error model to capture the spatial dynamic relationship between inequality and growth. We find that the magnitude of the bias in the parameter for inequality is about 2.66 percent, while that for initial income amounts to about 21.51 percent. We argue that the bias in the inequality parameter is relatively small because even though N is relatively large, T is small. Education is still positively correlated

with growth. However, the bias in the parameter for this variable is relatively small (1.19 percent).

Bibliography

- [1] Abreu, M., De Groot, H. L. F., and Florax, R. J. (2005): “Space and growth: a survey of empirical evidence and methods”, *Région et Développement*, 13-44
- [2] Alesina, A., and Perotti, R. (1994): “Income Distribution, Political Instability and Investment”, *European Economic Review*, 1170-1189.
- [3] Alesina, A., and Rodrik, D. (1994): “Distributive Politics and Economic Growth”, *Quarterly Journal of Economics*, 465-490.
- [4] Atkinson, A. (1970): “On the Measurement of Inequality”, *Journal of Economic Theory*, 244-263.
- [5] Attanasio, O. P., Picci, L., and Scorcu, A. E. (2000): “Saving, Growth and Investment: A Macroeconomic Analysis Using a Panel of Countries”, *Review of Economics and Statistics*, 1-30.
- [6] Bachmeier, L. J. (2008): “Monetary Policy and the Transmission of Oil Shocks”, *Journal of Macroeconomics*, 1738-1755.
- [7] Bachmeier, L. J., and Cha, I. (2011): “Why Don’t Oil Shocks Cause Inflation? Evidence From Disaggregate Inflation Data”, *Journal of Money, Credit and Banking*, Forthcoming.

- [8] Bakeman, R., McArthur, D., and Quera, V. (1996): “Detecting group differences in sequential association using sampled permutations: Log odds, kappa, and phi compared”, *Behavioral Research Methods*, 446-457.
- [9] Barro, R. J. (2001): “Human Capital and Growth”, *American Economic Review*, 12-17.
- [10] Barro, R. J. (2000): “Inequality and Growth in a Panel of Countries”, *Journal of Economic Growth*, 5-32.
- [11] Barro, R. J., and Sala-i-Martin, X. (1992): “Convergence”, *Journal of Political Economy*, 407-443.
- [12] Barro, R. J., and Sala-i-Martin, X. (1991): “Convergence Across States and Regions”, *Brookings Papers and Economic Activity*, 107-182.
- [13] Bec, F., and Bassil, C. (2009): “Federal Funds Rate Stationarity: New Evidence”, *Economics Bulletin*, 867-872.
- [14] Belsley, D. A., Kuh, E., and Welsch, R. E. (1980): *Regression Diagnostics: Influential Data and Sources of Collinearity*, New York: Wiley.
- [15] Benkwitz, A., Lutkepohl, H., and Neumann, M. (2000): “Problems related to confidence intervals for impulse responses of autoregressive processes”, *Econometric Reviews*, 69-103.
- [16] Bergtold, J. S., Spanos, A., and Onukwugha, E. (2010): “Bernoulli Regression Models: Revisiting the Specification of Statistical Models with Binary Dependent Variables”, *Journal of Choice Modelling*, 1-28.

- [17] Bernanke, B. S., Gertler, M. L., and Watson, M. (1997): “Systematic Monetary Policy and the Effects of Oil Price Shocks”, *Brookings Papers on Economic Activity*, 91-157.
- [18] Bhargava, A., and Sargan, J. D. (1983): “Estimating Dynamic Random Effects Models From Panel Data Covering Short Time Periods”, *Econometrica*, 1635-1659.
- [19] Bierens, H. J. (2007, April 5). Multicollinearity. University Park, Pennsylvania, USA.
- [20] Blanchard, O. J., and Gali, J. (2010): “The Macroeconomic Effects of Oil Price Shocks: Why are the 2000s so Different from the 1970s”, in J. Gali, & M. Gertler, *International Dimensions of Monetary Policy* (pp. 373-428). Chicago: University of Chicago Press.
- [21] Blanchard, O. J., and Quah, D. (1989): “The Dynamic Effects of Aggregate Demand and Supply Disturbances”, *American Economic Review*, 655-673.
- [22] Blundell, R., and Smith, R. J. (1991): “Conditions Initiales et Estimation Efficace dans les Modèles Dynamiques sur Données de Panel : Une Application au Comportement d’investissement des Entreprises”, *Annales d’Economie et de Statistique*, 109-123.
- [23] Blundell, R., Bond, S., and Windmeijer, F. (2000): “Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator”, in Baltagi, B. (ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels. Advances in Econometrics* (pp.15). Amsterdam: JAI Press, Elsevier.

- [24] Campbell, J. Y., and Shiller, R. J. (1987): “Cointegration and Tests of Present Value Models”, *Journal of Political Economy*, 1062-1088.
- [25] Cochrane, J. H., and Piazzesi, M. (2002): “The Fed and Interest Rates - A High-Frequency Identification”, *American Economic Review*, 90-95.
- [26] Crowder, W. J. (2006): “The Interaction of Monetary Policy and Stock Returns”, *Journal of Financial Research*, 523-535.
- [27] Davidson, R., and MacKinnon, J. G. (1981): “Several Tests for Model Specification in the Presence of Alternative Hypotheses”, *Econometrica*, 781-793.
- [28] Diebold, F., and Mariano, R. S. (1995): “Comparing Predictive Accuracy”, *Journal of Business and Economic Statistics*, 253-263.
- [29] Driesprong, G., Jacobsen, B., and Maat, B. (2008): “Striking Oil: Another Puzzle?”, *Journal of Financial Economics*, 307-327.
- [30] Durlauf, S. N., and Quah, D. T. (1999): “The New Empirics of Economic Growth”, in J. B. Taylor, & M. Woodford, *Handbook of Macroeconomics* (pp. 235-308). Elsevier.
- [31] Eddelbuettel, D., and Francois, R. (2011) “Rcpp: Seamless R and C++ Integration”, *Journal of Statistical Software*, 1-18.
- [32] Elhorst, J. P. (2003): “Specification and Estimation of Spatial Panel Data Models”, *International Regional Science Review*, 244-268.
- [33] Ewing, B. T., and Thompson, M. A. (2007): “Dynamic Cyclical Comovements of Oil Prices with Industrial Production, Consumer Prices, Unemployment, and Stock Prices”, *Energy Policy*, 5535-5540.

- [34] Forbes, K. (2000): “A Reassessment of the Relationship between Inequality and Growth”, *American Economic Review*, 869-887.
- [35] Goetz, S. J., and Hu, D. (1996): “Economic Growth and Human Capital Accumulation: Simultaneity and Expanded Convergence Tests”, *Economics Letters*, 355-362.
- [36] Goodman, L. A. (1965): “On the Multi-Variable Analysis of 3 Dichotomous Variables”, *American Journal of Sociology*, 290-301.
- [37] Gourieroux, C. (2000): *Econometrics of Qualitative Dependent Variables*, Cambridge, UK: Cambridge University Press.
- [38] Greene, W. H. (2000): *Econometric Analysis*, Upper Saddle River: Prentice Hall, Inc.
- [39] Griffith, D. A. (1996): “Some Guidelines For Specifying the Geographic Weights Matrix Contained in Spatial Statistical Models”, in S. L. Arlinghaus, *Practical Handbook of Spatial Statistics* (pp. 65-82), Boca Raton: CRC Press.
- [40] Hamilton, J. D. (2010): “Nonlinearities and the Macroeconomic Effects of Oil Prices”, *NBER Working Papers*, 16186.
- [41] Hamilton, J. D. (1983): “Oil and the Macroeconomy Since World War II”, *Journal of Political Economy*, 228-248.
- [42] Hamilton, J. D. (1994): *Time Series Analysis, Princeton*: Princeton University Press.
- [43] Hamilton, J. D. (2003): “What is an Oil Shock?” *Journal of Econometrics*, 593-617.

- [44] Hoerl, A. E., and Kennard, R. W. (1970): “Ridge Regression: Biased Estimation for Nonorthogonal problems”, *Technometrics*, 55-67.
- [45] Hoerl, A., and Kennard, R. (1988): “Ridge regression”, *Encyclopedia of Statistical Sciences*, 129-136.
- [46] Hoerl, A., and Kennard, R. W. (1970): “Ridge Regression: Applications to Nonorthogonal problems”, *Technometrics*, 69-82.
- [47] Hosmer, D. W., and Lemeshow, S. (1989): *Applied Logistic Regression*, New York: John Wiley and Sons, Inc.
- [48] Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (2002): “Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Models Covering Short Time Periods”, *Journal of Econometrics*, 107-150.
- [49] James, W., and Stein, C. (1960): “Estimation With Quadratic Loss”, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 361-379). Berkeley: University of California Press.
- [50] Jones, C., and Kaul, G. (1996): “Oil and the Stock Markets”, *Journal of Finance*, 463-491.
- [51] Kaldor, N. (1957): “A Model of Economic Growth”, *Economic Journal*, 591-624.
- [52] Keefer, P., and Knack, S. (2000): “Polarization, Politics, and Property Rights : Links Between Inequality and Growth”, *Public Choice*, 127-154.
- [53] Kilian, L., and Lewis, L. (2011): “Does the Fed Respond to Oil Price Shocks”, *The Economic Journal*, Forthcoming.

- [54] Kilian, L., and Park, C. (2009): “The Impact of Oil Price Shocks on the U.S. Stock Market”, *International Economic Review*, 1267-1287.
- [55] Kilian, L., and Vigfusson, R. J. (2011): “Nonlinearities in the Oil Price-Output Relationship”, *CEPR Discussion Papers*, 8174.
- [56] Kilian, L. (2008): “The Economic Effects of Energy Price Shocks”, *Journal of Economic Literature*, 871-909.
- [57] Kuznets, S. (1955): “Economic Growth and Income Inequality”, *American Economic Review*, 1-28.
- [58] Le Cessie, S., and Van Houwelingen, J. C. (1992): “Ridge estimators in Logistic Regression”, *Applied Statistics*, 191-201.
- [59] Li, H., and Zou, H.-f. (1998): “Income Inequality is not Harmful for Growth: Theory and Evidence”, *Review of Development Economics*, 318-334.
- [60] Marquardt, D. W. (1970): “Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation”, *Technometrics*, 591-612.
- [61] Mason, R., Gunst, R. F., and Webster, J. T. (1975): “Regression Analysis and Problems of Multicollinearity”, *Communications in Statistics*, 277-292.
- [62] McCullagh, P., and Nelder, J. A. (1983): *Generalized Linear Models*, London: Chapman and Hall.
- [63] Menard, S. (2002): *Applied Logistic Regression Analysis*, Thousand Oaks: Sage.
- [64] Menard, S. (2010): *Logistic Regression: From Introductory to Advanced Concepts and Applications*, Los Angeles: Sage.

- [65] Miller, I. J., and Ratti, R. A. (2009): “Crude Oil and Stock Markets: Stability, Instability, and Bubbles”, *Energy Economics*, 559-568.
- [66] Morgan, W. A. (1939): “A Test for Significance of the Difference Between Two Variances in a Sample from a Normal Bivariate Population”, *Biometrika*, 13-19.
- [67] Nerlove, M., and Balestra, P. (1996): “Formulation and Estimation of Panel of Econometric Models For Panel Data”, in M. L., & P. Sylvestre, *The Econometrics of Panel Data* (pp. 3-22), Kluwer: Dordrecht.
- [68] Ngarambe, O., Goetz, S. J., and Debertin, D. L. (1998): “Regional Economic Growth and Income Distribution: County-Level Evidence from the U.S. South”, *Journal of Agricultural and Applied Economics*, 325-337.
- [69] O’Brien, R. M. (2007): “A Caution Regarding Rules of Thumb for Variance Inflation Factors”, *Quality and Quantity*, 673-690.
- [70] Panizza, U. (2002): “Income Inequality and Economic Growth: Evidence from American Data”, *Journal of Economic Growth*, 25-41.
- [71] Partridge, M. (2005): “Does Income Distribution Affect U.S. State Economic Growth?”, *Journal of Regional Science*, 363-394.
- [72] Partridge, M. (1997): “Is Inequality Harmful for Growth? Comment”, *American Economic Review*, 1019-1032.
- [73] Partridge, M. (2007): “The Elusive Inequality-Economic Growth Relationship: Are there Differences between Cities and the Countryside?”, *Annals of Regional Science*, 375-400.

- [74] Peng, C., Lee, K., and Ingersoll, G. (2002): “An introduction to logistic regression analysis and reporting”, *Journal of Educational Research*, 3-13.
- [75] Persson, T., and Tabellini, G. (1994): “Is Income Inequality Harmful for Growth”, *American Economic Review*, 600-621.
- [76] Rawlings, J. O., Dickey D., A., and Pantula S., G. (1998): *Applied Regression Analysis*, New York: Springer-Verlag.
- [77] Reis, R., and Blinder, A. S. (2005): “Understanding the Greenspan Era”, *Proceedings, Federal Reserve Bank of Kansas City*, 11-96.
- [78] Ridder, G., and Wansbeek, T. (1989): “Dynamic Models For Panel Data”, in F. V. Ploeg, *Advanced Lectures in Quantitative Economics* (pp. 557-583), London: Academic Press.
- [79] Romer, P. M. (1989): “Human Capital and Growth: Theory and Evidence”, *NBER Working Paper*, 251-286.
- [80] Rotemberg, J. J., and Woodford, W. (1996): “Imperfect Competition and the Effects of Energy Price Increases”, *Journal of Money, Credit and Banking*, 549-577.
- [81] Schaefer, R. L. (1985): “Alternative estimators in logistic regression when the data are collinear”, *Journal of Statistical Computation and Simulation*, 75-91.
- [82] Schaefer, R. L., Roi, L. D., and Wolfe, R. A. (1984): “A Ridge Logistic Estimator,” *Communications in Statistics - Theory and Methods*, 99-113.

- [83] Scrucca, L. and Weisberg, S. (2004). A simulation study to investigate the behavior of the log-density ratio under normality. *Communications in Statistics - Computation and Simulation*, 159–178.
- [84] Sensier, M., and Van Dijk, D. (2004): “Testing for changes in volatility of US macroeconomic time series”, *Review of Economics and Statistics*, 833-839.
- [85] Shen, J., and Gao, S. (2008): “A Solution to Separation and Multicollinearity in Multiple Logistic Regression”, *Journal of Data Science*, 515-531.
- [86] Shikida, C. D., and Figueirado, E. A. (2010): “Is the Federal Funds rate stationary? New evidence from P-ADF Tests”, *Revista de Economia e Administração*, 94-101.
- [87] Snee, R. D., and Marquardt, D. W. (1984): “Collinearity Diagnostics Depend on the Domain of Prediction, the Mode, and the Data”, *American Statistician*, 83-87.
- [88] Spanos, A. (1999): *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge, UK: Cambridge University Press.
- [89] Spanos, A. (1986): *Statistical Foundations of Econometric Modeling*, Cambridge, UK: Cambridge University Press.
- [90] Spanos, A., and McGuirk, A. (2001): “The Model Specification Problem from a Probabilistic Reduction Perspective”, *American Journal of Agricultural Economics*, 1168-1176.

- [91] Spanos, A., and McGuirk, A. (2002): “The Problem of Near-Multicollinearity Revisited: Erratic vs Systematic Volatility”, *Journal of Econometrics*, 365-393.
- [92] Stein, C. (1955): “Inadmissibility of the Usual Estimator for the Mean of a Multivariate Distribution”, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (pp. 197-206), Berkeley: University of California Press.
- [93] Stetzer, F. (1982): “Specifying Weights in Spatial Forecasting Models: The Results of Some Experiments”, *Environment and Planning*, 571-584.
- [94] Stock, J., and Watson, M. (1999): “Forecasting inflation”, *Journal of Monetary Economics*, 293-335.
- [95] Urgan, N. N., and Tez, M. (2008): “Liu Estimators in Logistic Regression when the Data are Collinear”, *20th EURO Mini Conference* (pp. 323-327), Neringa, Lithuania: Vilnius.
- [96] Verlinda, J. A. (2006): “A Comparison of Two Common Approaches for Estimating Marginal Effects in Binary Choice Models”, *Applied Economics Letters*, 77-80.
- [97] Warnock, F. E., and Warnock, V. C. (2000): “Explaining the Increased Variability in Long-term Interest Rates”, *Economic Quarterly*, 71-96.
- [98] Webster, J. T., Gunst, R. F., and Mason, R. L. (1974): “Latent Root Regression Analysis”, *Technometrics*, 5 13-522.

- [99] Williamson, J. G. (1965): “Regional Inequality and the Process of National Development: A Description of the Patterns”, *Economic Development Cultural Change*, 3-45.
- [100] Wooldridge, J. M. (2002): *Introductory Econometrics: A Modern Approach*, Cincinnati: South Western College Publishing.