ESSAYS ON THE DYNAMICS OF AND FORECASTING ABILITY WITHIN THE U.S. ENERGY SECTOR

A Dissertation

by

KANNIKA DUANGNATE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	James W. Mjelde
Co-Chair of Committee,	David A. Bessler
Committee Members,	Ximing Wu
	Ke-Li Xu
Head of Department,	C. Parr Rosson III

December 2015

Major Subject: Agricultural Economics

ABSTRACT

Advances in production technology are increasing the availability of natural gas in the U.S. Examining how these technological advancements influence the dynamics in the natural gas and energy sectors is the subject here. Tests for parameters constancy in cointegrated vector autoregressive models are applied to investigate the possible existence of structural changes in pricing relationships among North America natural gas spot markets. Results suggest that long-run pricing relationships among eight natural gas spot prices are not constant over the period of 1994 to 2014. Two potential breaks are found, during 2000 and 2009. Possible contributing factors to structural changes occurring during 2000 are expensive and volatile natural gas prices, the U.S. Federal Energy Regulatory Commission Order No. 637, and changes in imports. The major contributing factor to the structural change during 2009 is likely the shale gas revolution.

Prequential analysis is applied to determine how the presence of these structural breaks affects probabilistic forecasts of out-of-sample data of natural gas returns. Models using longer periods as the based estimation period, forecast returns better. Threshold cointegration examines the effects of the structural breaks on transaction costs between natural gas markets. Pairwise transaction costs differ between the 2009 pre- and post-break periods. During the post-break period, five of seven pairwise transaction costs decrease, while the remaining two pair-wise transaction costs increase relative to the pre-break period. Alterations in natural gas flows as the result of the shale gas revolution partially explain the transaction costs changes.

Finally, a data-rich methodology is used to investigate how the number of factors derived from a large number of time series influences inferences and probabilistic forecasting performance concerning natural gas production. Factor-augmented vector autoregressive models and prequential analysis are applied to data series of the U.S. energy and macroeconomic variables. The number of factors minimally affects inferences from factor-augmented vector autoregressive models, but considerably affects probabilistic forecasting performance. Exploiting estimated factor improves the forecasting ability, but including too many factors tends to exacerbate probabilistic forecasts performance.

DEDICATION

To my family, for their love and support

ACKNOWLEDGEMENTS

I would like to express my gratitude to my committee co-chairs, Dr. Mjelde and Dr. Bessler, and my committee members, Dr. Wu and Dr. Xu for their guidance and support throughout the course of this research. Appreciations also go to my friends, the department faculty and staff, and Thai Aggies for making my time at Texas A&M University a great experience.

I am thankful to Dr. Park, Dr. Lo, and Dr. Zivot for providing a GAUSS program to estimate threshold vector error correction models and to Dr. Han and Inoue for providing a Matlab program to test for parameter instability in factor models. I am also grateful to Dr. Andrews for providing a GAUSS program to generate critical values used in testing for parameter instability.

I am indebted to the Office of the Higher Education Commission, Thailand, for granting me a scholarship to pursue my doctoral study, and to the staff at the Office of Educational Affairs, Royal Thai Embassy, Washington DC, for their support. Also, I would like to express my special acknowledgement to my friends and the department faculty at the Department of Economics, Chiang Mai University, Thailand, for their encouragement. Finally, thanks to my grandfather, mother, father, and sister, for their love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Natural gas is one of major energy sources in the United States (U.S.), contributing to multiple sectors of the economy (National Energy Technology Laboratory 2013). In 2014, approximately 26.82 trillion cubic feet (Tcf) of natural gas were consumed in the U.S. (U.S. Energy Information Administration (U.S. EIA) 2015c). Electric power, industrial, residential, and commercial sectors consumed approximately 30%, 29%, 19%, and 13% of this total consumption amounts (U.S. EIA 2015c). The majority of natural gas consumed in the U.S. is from the domestic production (U.S. EIA 2014d). Until 1986, U.S. natural gas production and consumption were nearly equal (U.S. EIA 2014d). After 1986, consumption has been greater than production; the U.S. has become a net importer of natural gas (U.S. EIA 2014d).

Beginning about 2006, domestic natural gas production increased because of more efficient, cost-effective drilling and completion techniques, which are a combination of horizontal drilling and hydraulic fracturing, especially in the production of natural gas from shale formations (U.S. EIA 2014d, 2015d). The growth in domestic production led to declining natural gas prices resulting in decreased imports, increased exports, and increased consumption, particularly in the electric power and industrial sectors (U.S. EIA 2013b, 2014c).

Are these technological advances, which are causing natural gas production and consumption increases, altering the economic dynamics of the U.S. energy sector? This is the general question addressed. Specially, the overall objective is to investigate whether and how the dynamics in the natural gas sector and the energy sector are influenced by the technological advancements in shale gas production. To achieve this general objective, time series econometric methods are implemented to explore natural gas daily pricing relationships, changes in transaction costs between natural gas spot markets, and dynamic effects of natural gas gross withdrawals on the U.S. energy sector. Four almost self-contained essays, Chapters II though V, address these issues.

Dynamics in Daily Natural Gas Pricing

Starting with the natural gas sector, daily natural gas pricing relationships are explored in Chapters II and III. The objective of the study presented in Chapter II is to investigate the possible existence and effects of structural changes with unknown break points among North America natural gas spot markets. The questions not only whether structural changes exist among North America natural gas spot markets, but also what may be inducing such changes are addressed. To achieve the objective, tests for parameter constancy in a cointegrated vector autoregressive model introduced by Hansen and Johansen (1999) are applied to investigate potential existence of structural changes in long-run pricing relationships among North America natural gas spot markets. Evidence from the tests suggests two possible structural changes, one during 2000 and another during 2009. Possible contributing factors to the structural change during 2000 are expensive and volatile natural gas prices, the U.S. Federal Energy Regulatory Commission Order No. 637, and changes in imports. The likely major contributing factor to the break occurring during 2009 is the shale gas revolution. Natural gas pricing dynamics are examined by dividing the data into three sub-periods that correspond to the structural changes and performing innovation accounting analysis (impulse response functions and forecast error variance decompositions) for each sub-period. Dynamics in daily natural gas pricing differ across the three sub-periods.

Because of the potential presence of structural breaks found in Chapter II, the objective in Chapter III is to determine whether and how the potential presence of structural breaks affects out-of-sample probability forecasting performance using the prequential forecasting approach introduced by Dawid (1984). To achieve this objective, calibration measures (calibration plots and chi-squared goodness-of-fit test statistics), root mean-squared error, the Brier score and its decompositions, and the ranked probability score are applied for model assessments. Different in-sample data periods provide different probability forecasts. Models having better forecasting performance are the models which incorporate a larger time period. Interestingly, the larger time period includes a period of potential structure changes.

Transaction Costs

The objective in Chapter IV is to examine the presence of threshold cointegration between market pairs before and after the potential break (associated with the shale gas revolution) in the long-term pricing relationship among the North America natural gas spot markets. Threshold cointegration allows for non-linear long-run relationships (Balke and Fomby 1997). Under the law of one price, this non-linear relationship is explained by transaction costs. Including transportation costs, transaction costs are costs incurred when participating in a market or between markets. Transaction costs between market pairs differ before and after the structural shift that occurred in 2009. After 2009, five of seven pairwise transaction costs decrease, while the remaining two increase. Changes in natural gas flows as the result of the shale gas revolution are most likely the cause of the differences in transaction costs.

Dynamics in the U.S. Energy Sector

Unlike the vector autoregressive model, a factor-augmented vector autoregressive (FAVAR) model allows for the incorporation of richer information data sets. The objective in Chapter V is to investigate whether and how the number of unobservable components from a data-rich model influences inferences and probabilistic forecasting performance of various models. Models include two FAVAR (varying number of factors), a traditional vector autoregressive (VAR), and a univariate autoregressive (AR) models. The FAVAR approach, proposed by Bernanke, Boivin, and Eliasz (2005), is employed to characterize factors stimulating dynamics in the U.S. energy sector. Innovation accounting analysis (impulse response functions and forecast error variance decompositions) are applied to discover dynamic responses among estimated factors and natural gas gross withdrawals. Then, the prequential forecasting approach introduced by Dawid (1984) is applied to evaluate predictive distributions for out-of-sample data. It appears that the number of factors has only a minor impact on the inferences from dynamic responses, but has a considerable impact on the probabilistic forecasting performance.

CHAPTER II

TESTS FOR LONG-RUN PARAMETER CONSTANCY IN COINTEGRATED VAR MODELS OF NATURAL GAS SPOT PRICES

Since deregulation of natural gas wellhead prices and pipeline regulatory reform, the natural gas market has become more efficient (DeVany and Walls 1994a; Joskow 2013). Natural gas prices are driven by market supply and demand conditions, as illustrated by findings of market competitiveness and allocative efficiency (DeVany and Walls 1993, 1994b; Walls 1994a, 1994b; Doane and Spulber 1994; King and Cuc 1996; Serletis and Rangel-Ruiz 2004; Cuddington and Wang 2006; Park, Mjelde, and Bessler 2008; Mohammadi 2011; U.S. Energy Information Administration (U.S. EIA) 2014c). Supply side factors influencing natural gas prices consist of variations in domestic natural gas production, the volume of imported and exported gas, and the level of gas storage (U.S. EIA 2014c). Demand side factors consist of weather variability, economic growth, and other energy prices (U.S. EIA 2014c).

The majority of natural gas consumed in the U.S. is derived from domestic production (U.S. EIA 2014d). U.S. dry natural gas[1] production has noticeably increased in recent years with production in 2014 being approximately 39% higher than in 2006 (U.S. EIA 2015d). Horizontal drilling in conjunction with hydraulic fracturing[2] are behind the increasing natural gas production (U.S. EIA 2011, 2014d). The result of increasing domestic supply is declining prices. Natural gas spot prices at Henry Hub

[1] Dry natural gas is also known as consumer-grade natural gas (U.S. EIA 2015b)
[2] Hydraulic fracturing is fracturing of rock at depth with fluid pressure (U.S. EIA 2015b).

were \$5 to \$8 per million British thermal units (MMBtu) during 2003 to 2008, but decreased to \$2 to \$4 per MMBtu during 2009 to 2013 (U.S. EIA 2014f). Declining prices contribute to decreases in imported natural gas and increases in exports (U.S. EIA 2013b, 2013d).

In 2014, total U.S. natural gas consumption was four percent greater than total U.S. natural gas production (U.S. EIA 2015c, 2015d). The U.S., however, is projected to become a net exporter (U.S. EIA 2014b). Total domestic natural gas production is projected to outpace total domestic natural gas consumption by 2019; production is projected to be 18% larger than total domestic natural gas consumption by 2040 (U.S. EIA 2014b). In response to the abundant domestic gas supplies and relatively low natural gas prices, U.S. industrial natural gas consumption has also been increasing since reaching a low in 2009. Industrial natural gas consumption in 2014 was approximately 24% higher than that in 2009 (U.S. EIA 2015c). In the 1970s, 1980s, and early 1990s, the energy resources for the majority of U.S. electricity generation were primarily coal and nuclear power; economic, environmental, technological, and regulatory changes have caused natural gas to be the new fuel of choice for most of new power plants (National Energy Technology Laboratory 2013). In 2014, approximately 27% of U.S. electricity was generated by natural gas; it was 76% higher than U.S. electricity generated by natural gas in 2001 (U.S. EIA 2015c).

These developments in the natural gas industry may be altering the sector's supply and demand relationships. The objective of this study is to investigate the possible existence and effects of structural changes with unknown break points among

North America natural gas spot markets. This study addresses the questions not only whether structural changes exist among North America natural gas spot markets but also what may be inducing such changes. Associations between identified break points and actual events are identified. The U.S. EIA (2011) notes, "Although the U.S. Energy Information Administration's (U.S. EIA) National Energy Modeling System (NEMS) and energy projections began representing shale[3] gas resource development and production in the mid-1990s, only in the past 5 years has shale gas been recognized as a "game changer" for the U.S. natural gas market." This study aims to determine if the "game changer" has changed the pricing relationships among North America natural gas spot markets. If so, changing dynamic pricing relationships may influence trading and natural gas policy, such as pipeline systems. As such, the study is of interest not only to those interested in energy markets, but also researchers interested in modeling energy issues, market structural changes and time series analysis.

Literature on Price Dynamics in Natural Gas Markets

Many studies suggest that the deregulation of natural gas has improved natural gas market performance (DeVany and Walls 1993, 1994b; Walls 1994a, 1994b; Doane and Spulber 1994; King and Cuc 1996; Kleit 1998; Serletis and Rangel-Ruiz 2004; Cuddington and Wang 2006; Park, Mjelde, and Bessler 2008; Mohammadi 2011; Apergis, Bowden, and Payne 2015). Market integration is one of fundamental issues that have been used in many studies to monitor market performances. DeVany and

[3] Shale is a fine-grained, sedimentary rock composed of mud from flakes of clay minerals and tiny fragments (silt-sized particles) of other materials (U.S. EIA 2015b).

Walls (1993, p. 1) state, “The relationship between commodity prices at geographically dispersed locations is evidence of market performance.” If spatially separated markets for a homogenous good are integrated into one market, their prices will be interrelated and the “law of one price” holds under the constraints of transaction costs (transportation and/or arbitrage costs). Under the assumptions of no asymmetric information and no limitations on the transportable volume of the product, if transaction costs are zero, arbitrage will establish a single price in all spatially dispersed markets (King and Cuc 1996). If transaction costs are non-zero, arbitrage will guarantee that price differences will be equal to these costs (King and Cuc 1996).

Many empirical techniques have been employed to investigate price dynamics through market integration. DeVany and Walls (1993) and Walls (1994b), for example, employ the two-series cointegration model introduced by Engel and Granger (1987). DeVany and Walls (1993, p. 2) claim, “Cointegration provides a way to test for arbitrage-free pricing in time varying series.” Two non-stationary series are cointegrated (move together in the long-run) if they have a linear combination that is stationary (Engel and Granger 1987). This means that, if arbitrage is effective, prices, after considering transaction costs, converge to a single price (DeVany and Walls 1993). DeVany and Walls (1993) evaluate competition between natural gas spot markets located throughout the U.S using daily natural gas price data from 1987 to 1991. They find that the natural gas markets had become more competitive as most of market-pairs were not cointegrated in 1987 but more than 65% of the markets had become cointegrated by 1991. Walls (1994b) using daily natural gas price data from 1990 to

1991 finds that natural gas markets are strongly integrated within the production field, but much less integrated between the field and city markets. At some markets, Chicago and to a lesser extent California, natural gas prices closely follow field prices. The equalization of marginal values of natural gas across all production and consumption locations is suggestive of allocative efficiency.

Instead of using Engel and Granger's cointegration model, Walls (1994a) uses cointegration techniques developed by Johansen (1988, 1991) to measure market linkages in the U.S. natural gas sector. Walls (1994a, p. 189) argues, "The cointegration methodology developed by Johansen (1988, 1991) is the most fruitful way to test for spatial market linkages" as it overcomes the inference limitation of the Engel-Granger cointegration procedure. Using daily natural gas spot prices at twenty nodes located within six regions of the U.S. for 1989-1990, Walls (1994a) finds that natural gas spot markets at dispersed locations are connected.

Considering price correlation coefficients, DeVany and Wall (1994b) and Doane and Spulber (1994) find that the deregulation results increased competitiveness in natural gas markets. Concerned that there is no unique criterion describing the suitable level of correspondence between two price series, Doane and Spulber (1994) also rely on Granger causality and cointegration tests in addition to employing price correlation tests. They find consistent results among the three tests.

King and Cuc (1996) apply time-varying parameter (Kalman Filter) analysis, which allows for dynamic structure changes, to evaluate the level of price convergence in North America natural gas spot market. Using bid-week prices of natural gas for 17 markets in the U.S. and Canada from January 1986 to September 1995, they find that price convergence in all North America natural gas spot markets has increased since deregulation; yet, there exist an east-west split in natural gas pricing. Similarly, Cuddington and Wang's (2006) empirical results from autoregressive models of pairwise price differentials suggest that markets in the East and Central regions are highly integrated, but these markets are separated from the more roughly integrated Western market. In response to King and Cuc's (1996) findings, Serletis (1997) employs both Engle and Granger's (1987) approach and Johansen's (1988) maximum likelihood approach to investigate whether there exist an east-west split in North American natural gas markets using monthly spot price data from June 1990 to January 1996. They find the east-west separation does not exist.

Citing potential limitations of using correlations/cointegration approaches, Kleit (1998) estimates transactions cost directly to measure the effects of deregulation. Using monthly data from 1984 to 1993, he finds that transactions costs to and from the Louisiana, Oklahoma, and Texas regions have decreased, but transactions costs from the Rocky Mountain area have increased because of the deregulation.

Vector autoregressive (VAR) models and vector error correction models (VECM) also have been used to analyze price dynamics in the natural gas pricing literature. Using a VAR model, DeVany and Walls (1996) show that arbitrage-free prices and price dynamics depend on the market structure. Their study suggests the law of one price holds over most of the natural gas markets. Serletis and Rangel-Ruiz (2004) employ a VECM to investigate the strength of shared dynamics between West Texas

Intermediate crude oil prices, Henry Hub natural gas prices, and AECO Alberta natural gas markets using daily data from 1990 to 2001. They find evidence of decoupling of crude oil and natural gas prices and a high degree of interconnectedness between U.S. Henry Hub and AECO Alberta natural gas prices since deregulation. The study also indicates that natural gas prices in North America are largely defined by the U.S. Henry Hub price. Considering more diverse natural gas spot markets, Park, Mjelde, and Bessler (2008) employ a VECM to study dynamic interactions among North America natural gas spot markets and each market's role in price discovery for 1998-2007. They find that natural gas spot markets in North America are highly integrated but the degree of integration varies among the markets. Natural gas markets in Oregon, Illinois, and Louisiana are found to be the most significant markets for price discovery. With 11 natural gas markets, Olsen, Mjelde, and Bessler (2014) study interaction of natural gas prices in the U.S. and Canada using a VECM. Their results support earlier studies' findings that markets are integrated but the degree of integration varies among the markets; the closer markets are located, the higher degree of integration. It appears that eastern markets provide relatively more information to western markets than western markets provide to eastern markets; there is no east-west split in the U.S. Unlike previous studies, Olsen, Mjelde, and Bessler (2014)'s findings suggest that AECO, Alberta, is less important for price discovery than other Canadian markets.

In the literature, changes/shifts in natural gas markets are found and assorted factors affecting changes/shifts are identified. Mohammadi (2011) examines long-run relations and short-run dynamics of upstream-downstream pricing behavior in the U.S.

natural gas industry. Using monthly data from January 1984 to August 2009, he finds that natural gas markets are integrated but subject to regime shifts and asymmetric adjustments, suggesting market imperfections. Two regime shifts are found; one in the late 1990 regarding the implementation of regulatory reform and another during 2005-2009, a period of high volatility in energy prices. His results also indicate that shocks from both demand and supply sides play important roles in determining short-run price movements while demand shocks are the primary factor of natural gas prices in the long run.

Considering weekly data from March 1994 to September 2011, Lin and Wesseh (2013) find the existence of regime-switching in the natural gas market. They suggest that the shift in the early part of 1994 was because of oil shortages, whereas, the shift between 1998 and 2002 was related to Hurricane Mitch. The largest shift, however, is attributed to hurricane-related gas shortages in North America and the financial crisis in 2008. Similarly, Apergis, Bowden, and Payne's (2015) find a structural break occurred between 1994 and 1995; they identified the cause of this break as a response to the deregulation of natural gas industry. Allowing endogenous structural breaks, Apergis, Bowden, and Payne (2015) consider cointegration between city-gate and residential retail natural gas prices in the 50 U.S. states and find that degree of market integration of the post-break period is relatively higher than that of the pre-break period. Wakamatsu and Aruga (2013) examine whether U.S. shale gas production affects the structure of the U.S. and Japanese natural gas markets using monthly data from May 2002 to May 2012. They find a structural break of natural gas prices and consumption around 2005 suggesting that the shale gas production starting in 2005 led to a change in the relationships between the U.S. and Japanese natural gas markets. The U.S. and Japanese markets had been connected before 2005; but the U.S. natural gas market has become more independent since the shale gas revolution.

History of Tests for Parameter Constancy

Regression analysis in time-series analyses generally assumes that that the regression relationship is constant over time. The classical test for parameter constancy is the Chow test. Chow (1960) splits the sample into two sub-periods and estimates parameters for each sub-period to test whether sets of coefficients in two linear regressions are equal. Unfortunately to use the Chow test, the time of change point must be known. To overcome this drawback, Quandt (1960) suggests calculating the likelihood ratio test statistics (Chow Statistics) at each date of the data to find an unknown change point; the date which maximum statistic is obtained is determined as the change point. Using the cumulative sum (CUSUM) technique of residuals from recursively estimating the model, Brown, Durbin, and Evans (1975) proposed the CUSUM and CUSUM of square tests. Unfortunately, a nuisance parameter exists in both Quandt's (1960) and Brown, Durbin, and Evans' (1975) tests. An unknown break date is not identified under the null hypothesis (the parameter appears only under the alternative hypothesis). Consequently, these tests, which treat an unknown change point as a parameter, do not follow standard large sample asymptotic distributions (Andrews 1993). With an unknown change point, Andrews (1993) proposes parameter instability tests based on generalized method of moment estimators. He derives the asymptotic null distributions of the supremum test statistics. Andrews (1993) shows that his tests have nontrivial asymptotic local power against all alternatives of parameter instability, even though a one-time change is allowed. Bai and Perron (1998, 2003) estimate multiple structural changes at unknown dates using ordinary least squares. They present an algorithm to attain global minimizers of the sum of squared residuals based on dynamic programming.

Tests for parameter constancy in cointegrated systems are presented in literature. Saikkonen and Lüthepohl (2000) propose tests for the cointegrating rank of a vector autoregressive process that allows for a simple shift in the mean of the data-generation process. Johansen, Mosconi, and Nielsen (2000) provide a cointegration analysis in the presence of structural breaks in the deterministic trend. Hansen (2003) generalizes the cointegrated vector autoregressive model of Johansen (1988) such that structural changes can be tested in any subset of parameters. Saikkonen and Lüthepohl (2000), Johansen, Mosconi, and Nielsen (2000), and Hansen (2003) assume the break points are known a priori. Tests for unknown break points include Hansen (1992), Seo (1998), Hansen and Johansen (1999), and Lüthepohl, Saikkonen, and Trenkler (2004).

Based on the cointegration model proposed by Granger (1981) and developed by Engle and Granger (1987), Hansen (1992) introduces tests for parameter instability in regression with I(1) processes by making use of the fully modified estimation method of Philips and Hansen (1990). Hansen (1992) suggests that it is required to know the stochastic process of regressors before applying the tests, as the asymptotic distributions of the test statistics are dependent on the stochastic process of regressors. Seo (1998) proposes tests for structural change of the cointegrating vector and the adjustment vector in the error correction model. Normalization of the cointegration space is required in Seo's (1998), otherwise the cointegrating vector cannot be identified even though the cointegration space is identified. Considering tests for parameter constancy in the cointegrated vector autoregressive model, Hansen and Johansen (1999) suggest graphical analysis using recursive estimation to evaluate the constancy of the long-run parameters. Unlike Seo's (1998), Hansen and Johansen's (1999) tests do not require an identification of the individual cointegration vectors. Lüthepohl, Saikkonen, and Trenkler (2004) propose a cointegration rank test of a vector autoregressive process in which a simple shift in the mean is allowed.

Methodology

Literature on natural gas price dynamics usually imposes the assumption that the relationships among markets are constant over time. Failure to consider the potential existence of structural changes may cause bias and unreliable inferences.

Tests for Structural Changes in Cointegrated VAR models

In most empirical cointegration studies examining parameter instability, the long-run parameters are assumed to be constant while the short-run dynamics and the adjustment parameters from the error correction models are tested for parameter instability (Hansen and Johansen 1999). Hansen and Johansen's (1999) tests allow the long-run parameters to vary, but the short-term dynamics are constant over time. Two different techniques involving recursive estimation are presented for testing parameter instability in cointegrated VAR models. The first test assesses the time paths of the non-zero eigenvalues instead of all parameters in the model. The second test investigates the cointegration relations based on the Lagrange Multiplier (LM) type test. Both tests' results can be presented graphically.

A p-dimensional, k^{th} order VAR model expressed as a k-1^{th} order VECM is

$$(2.1) \quad \Delta X_t = \mu + \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t$$

$$= \alpha\beta^{*\prime} X_{t-1}^{*} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + e_t, \qquad t = 1, \dots, T,$$

where $X_t^{*} = (1', X_t')$ and $\beta^{*} = (\delta', \beta')$ (Hansen and Johansen 1999). $\mu = \alpha\delta'$ is constant term, α, β, δ, and Γ are parameters, 1 is a vector of one, X_t is a p x 1 vector of variables, and p and k are integers. The error terms e_t are assumed to be independent and Gaussian with mean zero and covariance matrix Ω, and the initial values $X_{-k+1}, \dots, X_0$ are fixed as the base sample (Hansen and Johansen 1999). $\alpha\beta'$ is expected to have reduced rank such that α and β are (p x r) matrices of full rank where r ($< p$) is the number of cointegrating vectors (Hansen and Juselius 1995). The long-run structure is identified by the cointegration space spanned by β while the short-run structure is identified through α and Γ_i (Johansen 1995). The constant term μ is restricted to satisfy the condition $\alpha_{\perp}'\mu = 0$ such that no deterministic trend is allowed in the model (Hansen and Johansen 1999).

Equation (2.1) can be rewritten as

$$(2.2) \quad Z_{0t} = \alpha\beta^{*\prime} Z_{1t} + \Gamma Z_{2t} + e_t, \qquad t = 1, \dots, T,$$

where $Z_{0t} = \Delta X_t$, $Z_{1t} = X_{t-1}^{*}$, $Z_{2t} = \Delta X'_{t-1}, \dots, \Delta X'_{t-k+1}$, and $\Gamma = (\Gamma_1, \dots, \Gamma_{k-1})$ (Hansen and Johansen 1999). Maximum likelihood estimation using all the data involves a reduced rank regression of Z_{0t} on Z_{1t} and Z_{2t} (Hansen and Johansen 1999).

Regression of Z_{0t} and Z_{1t} on Z_{2t} yields residuals $R_{0t}^{(T)}$ and $R_{1t}^{(T)}$, which are defined (Hansen and Johansen 1999) as

(2.3) $R_{0t}^{(T)} = Z_{0t} - M_{02}^{(T)}\left\{M_{22}^{(T)}\right\}^{-1} Z_{2t}$,

(2.4) $R_{1t}^{(T)} = Z_{1t} - M_{12}^{(T)}\left\{M_{22}^{(T)}\right\}^{-1} Z_{2t}$, and

(2.5) $R_{et}^{(T)} = e_t - M_{e2}^{(T)}\left\{M_{22}^{(T)}\right\}^{-1} Z_{2t}$,

where $M_{ij}^{(t)} = \sum_{s=1}^{t} Z_{it}Z_{jt}'$ and $M_{ej}^{(t)} = \sum_{s=1}^{t} e_t Z_{jt}'$ for $i, j = 0, 1, 2$.

The analysis in which the parameter Γ has been eliminated is based on the following regression equation (Hansen and Johansen 1999)

(2.6) $R_{0t}^{(T)} = \alpha\beta^{*\prime} R_{1t}^{(T)} + R_{et}^{(T)}, \qquad t = 1, \dots, T.$

The product moment matrices for *i*, *j* = 0, 1, e are defined as

(2.7) $S_{ij}^{T(t)} = \frac{1}{t}\sum_{s=1}^{t} R_{it}^{(T)} R_{ij}^{(T)\prime}$

$$= \frac{1}{t}[M_{ij}^{(t)} - M_{i2}^{(T)}\left\{M_{22}^{(T)}\right\}^{-1} M_{2j}^{(t)} - M_{i2}^{(t)}\left\{M_{22}^{(T)}\right\}^{-1} M_{2j}^{(T)}$$

$$+M_{i2}^{(T)}\left\{M_{22}^{(T)}\right\}^{-1} M_{22}^{(t)}\left\{M_{22}^{(T)}\right\}^{-1} M_{2j}^{(T)}],$$

and when *t* = *T*, $S_{ij} = S_{ij}^{T(T)}$(Hansen and Johansen 1999).

Solving the following problem,

(2.8) $|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0,$

yields eigenvalues $1 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_p > 0$ and $\hat{\lambda}_{p+1} = 0$ and eigenvectors $\hat{V} = (\hat{v}_1, \dots, \hat{v}_{p+1})$ which are normalized as $\hat{V}'S_{11}\hat{V} = I$ (Hansen and Johansen 1999).

The maximum likelihood estimators of β^* and α are given by $\hat{\beta}^* = (\hat{v}_1, \dots, \hat{v}_r)$ and $\hat{\alpha} = S_{01}\hat{\beta}^*$. Note that unless additional restriction(s) on β^* is (are) imposed, β^* is unidentified; only the space spanned by the vectors in β^* is estimable; further, the last row of $\hat{\beta}^*$ contains $\hat{\delta}$ such $\hat{\mu} = \hat{\alpha}\hat{\delta}'$ (Hansen and Johansen 1999).

Hansen and Johansen's (1999) tests are based on recursively estimating the VECM adding one observation at time. If the recursive estimation is based on equation (2.2), which is the Z-representation using $S_{ij} = S_{ij}^{T(t)}$, all parameters are estimated recursively given that all parameters can vary overtime (Hansen and Johansen 1999). If the recursive estimation is based on equation (2.6), which is the R-representation using $S_{ij} = S_{ij}^{T(T)}$, the constancy of parameters β is analyzed given that the short-run dynamics are constant over time (Hansen and Johansen 1999).

As noted by Hansen and Johansen's (1999), the estimated eigenvalues may not provide sufficient information to determine whether the long-run parameters are constant over time or not. The test for parameter constancy by Nyblom (1989) may be preferred to the fluctuation test of the eigenvalues. Hansen and Johansen (1999) show the Nyblom statistic for parameter instability has the same asymptotic distribution as that in Hansen's (1992) when it is applied to the cointegrated VAR model. In this study, the Lagrange Multiplier (LM) type test of Nyblom (1989) is applied to identify the potential existence of structural changes in pricing relationship among North America natural gas spot markets.

It is emphasized that the null hypothesis of the test is parameter constancy and a specific alternative is not stipulated (Hansen and Johansen 1999). Hansen and Johansen

(1999, p. 307) note, "We regard the recursive analysis as a misspecification test where the purpose is to detect possible instabilities in the parameters when there is no prior knowledge of structural breaks or time dependencies in the parameters."

Cointegrating vectors such that $r < p$ is required for testing for parameter constancy of β^* (Dennis 2006). Cointegrating rank and lag length are determined simultaneously using Schwarz loss measure. This method provides better large sample results in Monte Carlo simulations than the trace test, which determines the cointegrating rank given the lag order (Wang and Bessler 2005). Schwarz loss (SL) measure is

(2.9) $SL = \ln(\det(\Sigma)) + (p \text{ x } k)\text{xln}\,(T)/T,$

where Σ is the estimated variance-covariance matrix of error term with p is number of series in a considered vector, k is lag order in each equation, T is the number of total observations in each series, det(Σ) is the determinant of the variance-covariant matrix and ln is the natural logarithm (Wang and Bessler 2005).

For the base sample $t \in \{1, \dots, n\},\ n = T_0, \dots, T,$ let β^* be normalized on $\bar{c}$ such that $\hat{\beta}_c^{*(n)} = \hat{\beta}^{*(n)}\left(\bar{c}'\hat{\beta}^{*(n)}\right)^{-1}$ and define $\hat{\alpha}_c^{(n)} = \hat{\alpha}^{(n)}\hat{\beta}^{*(n)\prime}\bar{c}$, such that $\alpha_c^{(n)}\ \beta_c^{*(n)\prime} = \alpha^{(n)}\ \beta^{*(n)\prime}$. A maximum test for constancy of β^* (the difference between $\hat{\beta}^{*(n)}$ and $\hat{\beta}^{*(T)}$) is given by a sequence of test statistics (Dennis 2006),

(2.10) $Q_T^{(n)} = \left(\frac{n}{T}\right)^2 trace\left\{\left(V^{(T)}\right)^{-1} S^{n\prime}\left(M^{(T)}\right)^{-1} S^n\right\},$

where $V, M,$ and S are given by

(2.11) $V^{(T)} = \hat{\alpha}_c^{(T)\prime}\left(\hat{\Omega}^{(T)}\right)^{-1}\hat{\alpha}_c^{(T)},$

(2.12) $M^{(T)} = T^{-1}c_\perp' S_{11} c_\perp$, and

(2.13) $S^{(n)} = c_{\perp}'(S_{01}^{(n)} - \hat{\alpha}_c^{(T)}\hat{\beta}_c^{*(T)\prime}S_{11}^{(n)})'(\widehat{\Omega}^{(T)})^{-1}\hat{\alpha}_c^{(T)}$

$$= c_{\perp}'(S_{01}^{(n)} - \hat{\alpha}^{(T)}\hat{\beta}^{*(T)\prime}S_{11}^{(n)})'(\widehat{\Omega}^{(T)})^{-1}\hat{\alpha}^{(T)}.$$

In Hansen and Johansen (1999), the test statistic $Q_T^{(n)}$ is based on a first order approximation of the score function $S^{(n)} = M^{(n)}\left(Tc_{\perp}'\left(\hat{\beta}_c^{(n)} - \hat{\beta}^{(T)}\right)\right)'$. In this study, the score function as given in equation (2.13) is directly used following Bruggeman, Donati, and Warne (2003) because the first order approximation may lead to ill-behaved $Q_T^{(n)}$ in some situation, such as when the sample size is small (Dennis 2006).

Innovation Accounting Techniques

Based on the equivalent level VAR, innovation accounting techniques including impulse response functions and forecast error variance decompositions are used to analyze dynamic responses. Impulse response functions illustrate how each series in the model responds to a one-time shock in every series while forecast error variance decompositions illustrate how the forecast error for each series at any horizon is decomposed into shocks in each series (Doan 2000).

In equation (2.1), the innovation terms, e_t, are assumed to be independent but contemporaneous correlations among the elements are allowed. If the elements of innovation term are contemporaneously uncorrelated, then innovation accounting procedures can be performed using the moving average representation of the estimated VAR (Hamilton 1994). Nevertheless, contemporaneous correlations usually exist when considering economic data. Following Bernanke (1986) to obtain contemporaneously uncorrelated innovations, the observed innovations, e_t, are modeled as a function of more fundamental driving sources of variation, ε_t, which are independent (orthogonal) to other sources of variation

(2.14) $e_t = A^{-1}\varepsilon_t$,

where A is a matrix representing how each non-orthogonal innovation is caused by the orthogonal variation in each equation (Bernanke 1986). Usual innovation accounting procedure can be preformed by first re-expressing the estimated VECM as a VAR and then pre-multiplying the VAR by A. To obtain an identified model, zero restrictions on A are investigated using directed acyclic graphs (DAGs) on innovations from the estimated VECM. This procedure has been used by Swanson and Granger (1997), Hoover (2005), Park, Mjelde, and Bessler (2008), and Lai and Bessler (2015).

Directed Acyclic Graphs (DAGs)

Directed acyclic graphs (DAGs) help assigning contemporaneous causal flows to a set of observational variables. In DAGs, $X_i \rightarrow X_j$ indicates that X_i causes X_j; $X_k - X_l$ indicates that X_k and X_l are connected by information flows, but the algorithm cannot determine whether X_k causes X_j or vice versa[4]. Several algorithms have been developed to create DAGs. One of the most widely used algorithms is the PC algorithm named after the inventor, Peter and Clark (Spirtes and Glymour 1991). The PC algorithm is based on the notion of conditional independence under the assumptions of causal sufficiency, Markov condition, faithfulness condition, and Gaussian data. Causality sufficiency requires that there are no omitted variables causing two or more of the included variables. Markov condition requires that probabilities of variables can be expressed by conditioning just

[4] For more information on DAGs, see Pearl (2000) and Spirtes, Glymour, and Scheines (2000).

on variables of direct cause. Faithfulness condition is satisfied when correlation between two variables is zero because there is no edge between any two variables and no cancellation of structural parameters. The PC algorithm searches for zero correlations and conditional correlations to remove edges between variables and finds v-structures, i.e. $X_i \rightarrow X_j \leftarrow X_k$, to identify causation among a set of variables (Spirtes, Glymour, and Scheines 2000). Similar to the PC algorithm, the GES (greedy equivalency search) algorithm assumes the same conditions; however, the GES algorithm is a score-based algorithm, which searches over equivalent classes scoring each to find the best model (selected via loss metrics[5]).

The linear-Gaussian approach usually generates a set of possible models, which are equivalent in their conditional probability structure, resulting in indistinguishable causal flows (Shimizu et al. 2006). To over this problem, the assumption of Gaussian data must be relaxed such that the higher-order moments are used to identify the causal patterns (Shimizu et al. 2006). Shimizu et al. (2006) introduce a linear non-Gaussian acyclic model (LiNGAM), which allows the full causal model to be estimated. The LiNGAM algorithm is based on independent component analysis (Shimizu et al. 2006). In this study, the LiNGAM is applied to the residual series from the estimated VECM to create the orthogonal innovations for the estimation of impulse response functions and forecast error variance decompositions.

[5] For more information on the GES algorithm, see Chickering (2002, 2003).

Data

Eight natural gas spot prices in Canada and United States are considered: AECO Hub, Alberta, Canada; Chicago City Gate, Illinois; Dominion South Point, Pennsylvania; Henry Hub, Louisiana; Malin, Oregon; Oneok, Oklahoma; Opal, Wyoming; and Waha Hub, Texas. These markets allow for regional diffusion. Further, the markets have been identified in previous studies as important pricing markets. AECO Hub is included to represent Canada because in 2012, almost 99% of U.S. pipeline-imported natural gas came from Canada and 61% of pipeline natural gas exports went to Canada (U.S. EIA 2013d). Henry Hub is an important market for pricing of the North America natural gas spot and future markets (Serletis and Rangel-Ruiz 2004; U.S. EIA 2013a). Natural gas futures prices (NYMEX) are based on delivery at the Henry Hub (U.S. EIA 2014f). Park, Mjelde, and Bessler (2008) find Chicago is a dominant market for price discovery in North America natural gas spot markets. Texas, Pennsylvania, Louisiana, Oklahoma, and Wyoming were the top five national natural gas producing states in 2014 (U.S. EIA 2014d). Because of limited availability of historical prices at Ellisburg-Leidy Hub, prices at Dominion South Point are used to represent the Pennsylvania area. Description of each price series is provided in Appendix A.

Weekday nominal prices of natural gas from May 3, 1994 to October 31, 2014 are obtained from Bloomberg L.P. (2015). A missing value is replaced by a prior day's price. Each price is the closing price for a specific location for natural gas to be delivered on the next day. All prices are in U.S. dollars[6] per MMBtu (a unit of heat equal to one million British thermal units). Each price series contains 5,349 observations.

All price series are natural logarithm-transformed before any estimation. Summary statistics on the natural logarithms of each price series are presented in table 2.1, whereas, the data are plotted in figure 2.1. Each price series are non-Gaussian[7] as the null hypothesis of Jarque-Berra normality test on each series is rejected.

Testing for unit root (non-stationarity) in logarithm levels, Augmented Dickey-Fuller (ADF) statistics (Said and Dickey 1984) of all eight natural gas spot prices but Opal are greater than -3.430, which is the test critical value at 1% level (table 2.2). Failure to reject the null hypothesis of unit root implies that natural gas spot prices at all markets except Opal are non-stationary at the 1% level (table 2.2). Under the null hypothesis of unit root, the ADF test may have lower power against the alternative hypothesis of stationarity (DeJong et al. 1992). Kwiatkowski-Philips-Schmidt-Shin (KPSS) test statistics (Kwiatkowski et al. 1992) under the null hypothesis of stationary are also presented in table 2.2. KPSS test statistics of all prices in levels are greater than 0.739, which is the test critical value at 1% level, implying that the null hypotheses of stationarity are rejected. Based on KPSS tests, the eight prices in levels are

[6] A study considering potential effect of exchange rates and using Canadian currency is left for future research.

[7] The data used appears to be non-Gaussian but the tests for parameter constancy in the co-integrated VAR model are based on the assumption of Gaussianity. The robustness of tests for parameter constancy in the co-integrated VAR model under the assumption of Gaussian data when applying non-Gaussian data is left for future research.

Table 2.1. Summary Statistics on the Natural Logarithms of Natural Gas Spot Price at Eight Markets

Statistics	AECO Hub	Chicago	Dominion South	Henry	Malin	Oneok	Opal	Waha Hub
Mean	1.0833	1.3603	1.3832	1.3384	1.2475	1.2556	1.1002	1.2671
Maximum	2.7200	3.7153	3.2189	2.9642	4.0298	3.4592	3.3666	3.1987
Minimum	-0.7134	0.2070	0.1398	0.0296	-0.0726	0.0583	-1.8971	0.0770
Std. Dev.	0.6664	0.5116	0.5152	0.5164	0.5884	0.5084	0.5872	0.5102
Skewness	-0.3346	0.1607	0.2266	0.1181	-0.0919	0.0123	-0.2456	-0.0035
Kurtosis	2.2152	2.5735	2.3079	2.3510	2.5521	2.3205	2.5613	2.3505
Jarque-Bera	237.0803	63.5509	152.5372	106.3110	52.2469	103.0467	96.6488	94.0249
Probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

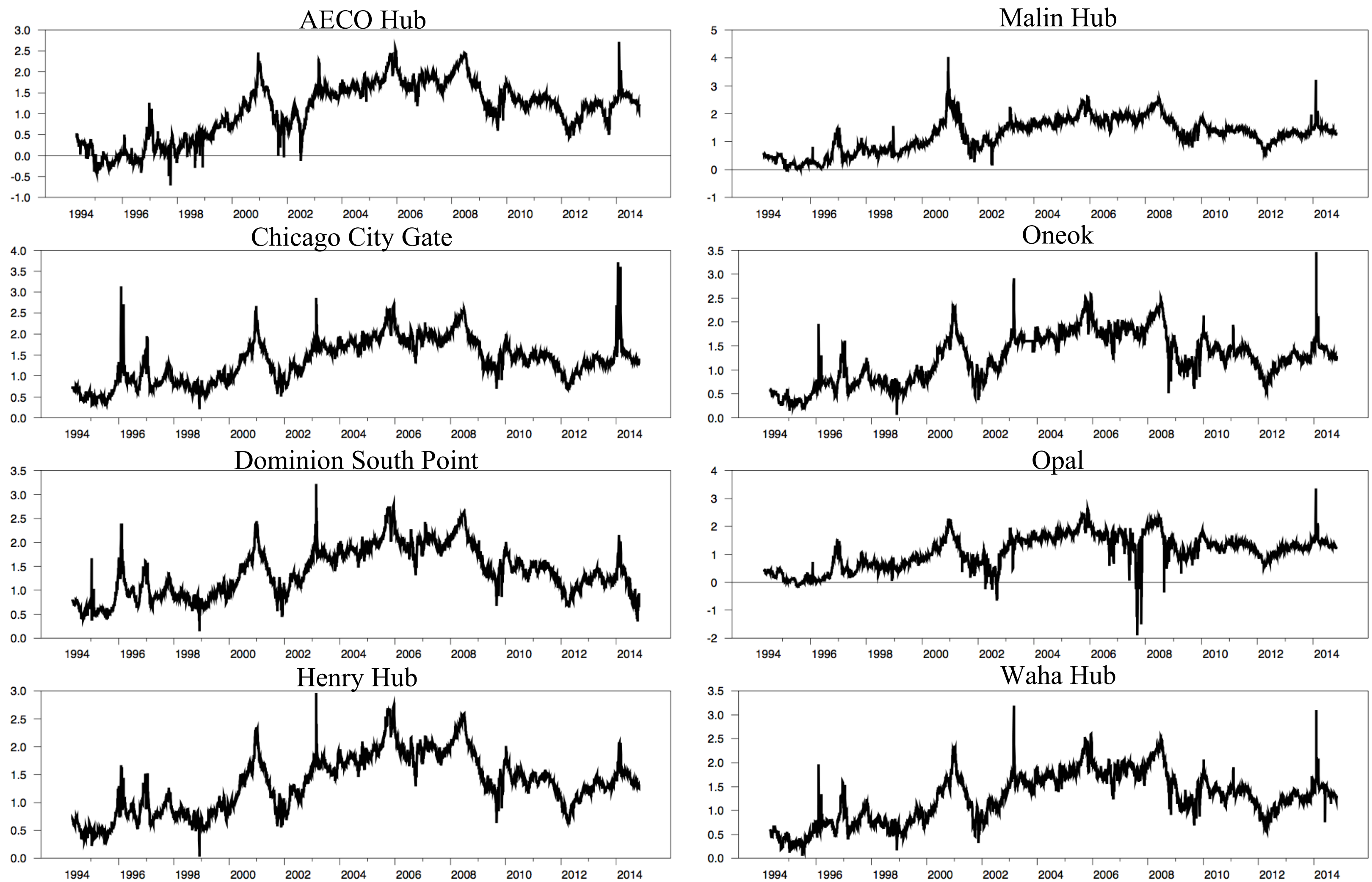

Figure 2.1. Plots of eight natural gas spot prices in the natural logarithms (May 3, 1994 - October 31, 2014)

Table 2.2. Augmented Dickey-Fuller (ADF) and Kwiatkowski-Philips-Schmidt-Shin (KPSS) Test[a] Statistics of Eight Natural Gas Spot Prices in the Natural Logarithms

	ADF		KPSS	
Price Series	t-Stat	Lag[b](k)	LM-Stat	Bandwidth[c]
		Test in Level		
AECO Hub	-2.5806	3	4.8239	56
Chicago	-2.9535	19	3.6618	56
Dominion South	-2.8588	19	2.9902	56
Henry Hub	-3.0193	2	3.7080	56
Malin	-2.7872	11	4.1550	56
Oneok	-3.1582	10	3.8187	56
Opal	-3.7137	7	4.2244	56
Waha Hub	-2.9942	15	3.9376	56
		Test in First Differences		
AECO Hub	-50.4972	2	0.0477	53
Chicago	-20.3818	18	0.0458	139
Dominion South	-18.1676	18	0.0640	33
Henry Hub	-59.9512	1	0.0418	24
Malin	-24.9024	10	0.0432	75
Oneok	-25.5139	9	0.0495	76
Opal	-36.2454	6	0.0243	63
Waha Hub	-21.0639	14	0.0513	82

Note: Under the null hypothesis of non-stationarity (unit root), the ADF test critical value at 1% level is -3.430; the null is rejected when t-Stat is less than the critical value (Said and Dickey 1984). Under the null hypothesis of stationarity, the KPSS test critical value at 1% level is 0.739; the null is rejected when LM-stat is greater than the critical value (Kwiatkowski et al. 1992).

[a] Only constant term is included in equations.

[b] Lag (k) is selected from 0 to 20 based on Schwarz information criteria.

[c] Bandwidth is estimated using the Newey-West (1994) method.

non-stationary at the 1% level. Both ADF and KPSS test statistics indicate that all price series are stationary after first differencing.

Empirical Results

Before conducting parameter instability tests of the cointegrated VAR model, Schwarz loss measures are used to determine the cointegrating rank and lag length simultaneously (Wang and Bessler 2005). The minimum Schwarz loss criterion suggests a rank of six cointegrating vectors with five lags[8] (table 2.3). Results suggest a potential weekday influences in the natural gas spot markets. A VECM model with five lags implies a VAR model with six lags.

Exclusion, Stationarity, and Weak Exogeneity in the Long-Run Relationship

Given six cointegrating vectors, tests of exclusion, stationarity, and weak exogeneity are performed (table 2.4). The null hypothesis of testing for variable exclusion is that an individual price series can be excluded from the long-run relationship. LR test statistics and corresponding *p*-values leads to rejecting the hypothesis of long-run exclusion, implying that no price series can be excluded.

Testing variable stationarity test is a multivariate version of the Dickey-Fuller test under the null hypothesis that each individual series is stationary given the cointegration rank. In table 2.4, results from tests of stationarity infer that conditional on the rank of cointegrating vector being equal to six, no price series can be considered stationary by itself.

[8] Results are consistent with the traditional two-step procedure (number of co-integrating vector is determined by the trace test after lag length is determined).

Table 2.3. Schwarz Loss Measures on One to Eight Co-Integrating Ranks and One to Six Lags on VAR model

Rank	One Lag	Two Lags	Three Lags	Four Lags	Five Lags	Six Lags
1	-47.7094	-48.0428	-48.2271	-48.2917	-48.3287	-48.2949
2	-47.8180	-48.1186	-48.2740	-48.3360	-48.3616	-48.3210
3	-47.8734	-48.1416	-48.2858	-48.3415	-48.3623	-48.3221
4	-47.9153	-48.1675	-48.2964	-48.3467	-48.3631	-48.3228
5	-47.9447	-48.1857	-48.3074	-48.3528	-48.3643	-48.3252
6	-47.9725	-48.2009	-48.3147	-48.3560	-48.3651*	-48.3254
7	-47.9785	-48.2039	-48.3158	-48.3543	-48.3620	-48.3221
8	-47.9770	-48.2024	-48.3139	-48.3524	-48.3600	-48.3201

Note: The asterisk '*' indicates minimum values of Schwarz loss measure.

Table 2.4. Results from Tests of Exclusion, Stationarity, and Weak Exogeneity Using the Entire Data

	Exclusion		Stationarity		Weak Exogeneity	
	LR-Test	*p*-Value	LR-Test	*p*-Value	LR-Test	*p*-Value
AECO Hub	70.7613	0.0000	50.496	0.0000	26.3842	0.000
Chicago	288.5096	0.0000	47.5830	0.0000	206.5310	0.000
Dominion South	31.0892	0.0000	48.1002	0.0000	21.4525	0.002
Henry Hub	232.97000	0.0000	47.1702	0.0000	24.2759	0.000
Malin	93.9012	0.0000	53.6027	0.0000	40.1149	0.000
Oneok	288.0352	0.0000	51.5380	0.0000	56.2269	0.000
Opal	69.8603	0.0000	52.9183	0.0000	68.7302	0.000
Waha Hub	316.9454	0.0000	51.6031	0.0000	111.2585	0.000

Note: Likelihood Ratio (LR) tests are executed conditional on six cointegrating vectors and five lags.

Variable exogeneity tests if any of the prices can be regarded as weakly exogenous when the parameter of interest is the vector β^*. Results from testing weak exogeneity for long-run parameters lead to rejecting the null hypothesis, implying that all eight prices respond to innovations in all six long-run equilibrium vectors.

Structural Changes in the Cointegrated VAR model

Tests for constancy of β^* in the cointegrated VAR model with the rank of six and five lags are performed. Test statistics (sup $Q_T^{(n)}$) based on equations (2.2) and (2.6) are consistent (figure 2.2). Testing the differences between $\hat{\beta}^{*(n)}$ and $\hat{\beta}^{*(T)}$, sup $Q_T^{(n)}$ approaches zero as $n \rightarrow T$ and eventually equal zero when $n = T$. The null hypothesis of parameter constancy is rejected when the test statistic is greater than the 5% critical value. In figure 2.2, the test statistics starting from the beginning of 1996 to approximately the end of 2010 are greater than the 5% critical value, implying the null hypothesis is rejected. The hypothesis of constancy of β^*, however, is only marginally rejected during the period 1996 to 2000; there appears to be a shift around the end of 2000. Instability of β^* implies that the long-run pricing relationships are not constant over the period 1994 to 2014.

Including exogenous variables, daily heating and cooling degree-days, in the cointegrating space, Park, Mjelde, and Bessler (2008) find seasonality in the long-run relationship among natural gas spot markets. In this study, Schwarz loss measure suggests a rank of six cointegrating vectors with four lags for the model in which daily heating and cooling degree-days are included. Test statistics for constancy of β^* in the

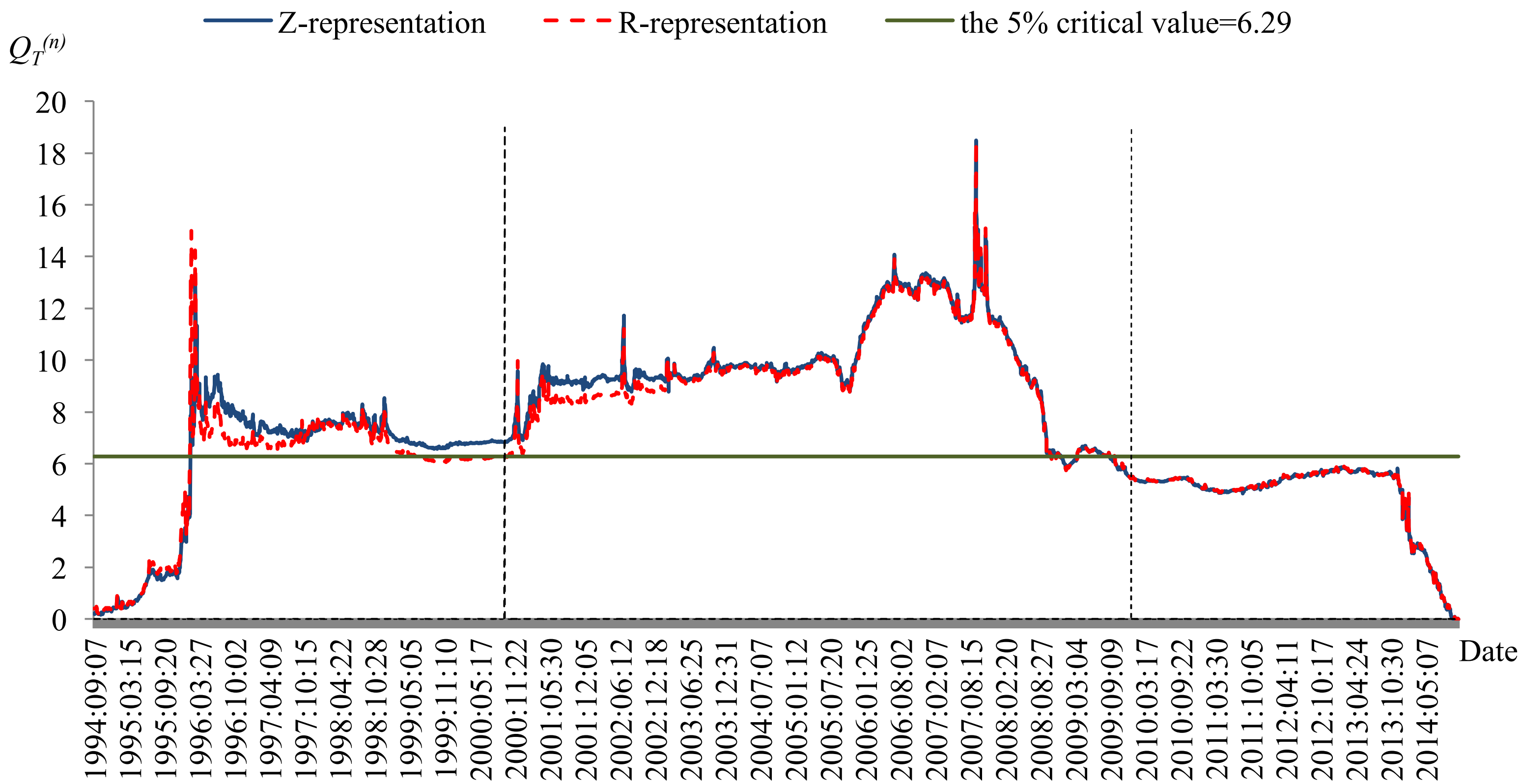

Figure 2.2. Plots of sup $Q_T^{(n)}$ for the entire data set (May 3, 1994 to October 31, 2014)
Note: The first vertical dash line indicates October 2, 2000. The second vertical dash line indicates January 1, 2010.

cointegrated system with exogenous variables (figure B.1. in Appendix B) have similar patterns to the model without exogenous variables. Test statistics of the model with daily heating and cooling degree-days during 2010-2014, however, are above the critical value line while those of the model without daily degree-days during 2010-2014 are below the line. In concern that the presence of parameter instability is due to seasonality, rather than the market structure changes, the model without daily degree-days is the focus of this study.

Structural Changes in the Cointegrated VAR model: Three Sub-Periods

Because of the β^* inconstancy, the data is divided into three sub-periods[9]: May 3, 1994 to September 29, 2000; October 2, 2000 to December 31, 2009; and January 1, 2010 to October 31, 2014. Each subsample[10] is tested for constancy of β^*. Schwarz loss measures indicate three lags are appropriate in each sub-period, but numbers of cointegrating vectors vary by subsamples. The ranks are three, seven, and four for the three sub-periods (table 2.5).

Conditional on the rank of cointegrating vectors from Schwarz loss criteria, tests of exclusion, stationarity, and weak exogeneity are executed. It appears that Malin can be excluded from the long-run relationship in the first sub-period; Opal can be excluded in the second sub-period; and AECO Hub can be excluded in the third sub-period, as LR test statistics and the corresponding *p*-values lead to rejecting the null hypothesis at the

[9] Using beak point a little bit earlier and later than October 2, 2000 and January 1, 2010 did not considerably affect results of parameter constancy inferences for each sub-period.

[10] Results of unit root and stationarity tests for each subsample are illustrated in Appendix B (table B.1).

Table 2.5. Schwarz Loss Measures on One to Eight Cointegrating Vectors (Rank) and One to Five Lags on VECM Model of Each Sub-Period

No. of Rank	One Lag	Two Lags	Three Lags	Four Lags	Five Lags
	First Sub-Period: May 3, 1994 - September 29, 2000				
1	-49.4419	-49.8651	-50.0501	-50.0061	-50.0002
2	-49.6631	-49.9702	-50.1221	-50.0480	-50.0223
3	-49.8090	-50.0581	-50.1754*	-50.0943	-50.0441
4	-49.8321	-50.0728	-50.1740	-50.0892	-50.0362
5	-49.8432	-50.0710	-50.1602	-50.0750	-50.0162
6	-49.8478	-50.0727	-50.1543	-50.0652	-50.0022
7	-49.8429	-50.0618	-50.1418	-50.0531	-49.9902
8	-49.8321	-50.0552	-50.1342	-50.0464	-49.9821
	Second Sub-Period: October 2, 2000 - December 31, 2009				
1	-49.9611	-50.2862	-50.3749	-50.3782	-50.3378
2	-50.0552	-50.3403	-50.3991	-50.3931	-50.3423
3	-50.1320	-50.3812	-50.4241	-50.3968	-50.3445
4	-50.2031	-50.4064	-50.4323	-50.3956	-50.3389
5	-50.2402	-50.4150	-50.4352	-50.3931	-50.3301
6	-50.2540	-50.4250	-50.4372	-50.3912	-50.3262
7	-50.2690	-50.4352	-50.4413*	-50.3924	-50.3251
8	-50.2657	-50.4321	-50.4371	-50.3878	-50.3213
	Third Sub-Period: January 1, 2010 - October 31, 2014				
1	-52.8178	-53.3655	-53.5012	-53.4867	-53.3622
2	-53.0956	-53.4778	-53.6021	-53.5656	-53.4014
3	-53.2156	-53.5282	-53.6089	-53.5663	-53.4003
4	-53.3002	-53.5610	-53.6124*	-53.5612	-53.3912
5	-53.3367	-53.5592	-53.5967	-53.5390	-53.3667
6	-53.3413	-53.5474	-53.5801	-53.5189	-53.3489
7	-53.3312	-53.5342	-53.5643	-53.5021	-53.3324
8	-53.3256	-53.5278	-53.5556	-53.4924	-53.3213

Note: The asterisk '*' indicates minimum values of Schwarz loss measure.

Table 2.6. Results from Tests of Exclusion, Stationarity, and Weak Exogeneity for Each Subsample

	Exclusion		Stationarity		Weak Exogeneity	
	LR-Test	*p*-Value	LR-Test	*p*-Value	LR-Test	*p*-Value
	First Sub-Period[a]: May 3, 1994 - September 29, 2000					
AECO Hub	13.6442	0.0034	173.5226	0.0000	7.5179	0.0571
Chicago	128.9945	0.0000	167.8125	0.0000	140.6442	0.0000
Dominion South	104.9531	0.0000	168.6998	0.0000	76.8253	0.0000
Henry Hub	141.4037	0.0000	170.7118	0.0000	17.5033	0.0006
Malin	2.2811	0.5161	172.7745	0.0000	5.9383	0.1147
Oneok	177.9637	0.0000	171.3246	0.0000	65.0674	0.0000
Opal	13.5674	0.0036	172.4573	0.0000	6.8963	0.0753
Waha Hub	161.1417	0.0000	171.6173	0.0000	62.0187	0.0000
	Second Sub-Period[b]: October 2, 2000 - December 31, 2009					
AECO Hub	79.4776	0.0000	38.8047	0.0000	7.7581	0.3544
Chicago	159.1988	0.0000	37.6941	0.0000	48.0765	0.0000
Dominion South	155.6357	0.0000	37.3962	0.0000	52.2563	0.0000
Henry Hub	135.5722	0.0000	36.9649	0.0000	22.9253	0.0018
Malin	77.0185	0.0000	38.8899	0.0000	30.2344	0.0000
Oneok	153.5586	0.0000	34.1045	0.0000	30.9838	0.0000
Opal	88.6046	0.0179	31.5392	0.0000	89.7231	0.0000
Waha Hub	180.2247	0.0000	37.3815	0.0000	65.0813	0.0000
	Third Sub-Period[c]: January 1, 2010 - October 31, 2014					
AECO Hub	9.3700	0.0525	66.5524	0.0000	36.5036	0.0000
Chicago	45.8918	0.0000	56.9517	0.0000	35.5045	0.0000
Dominion South	23.4924	0.0000	66.4445	0.0000	16.6796	0.0022
Henry Hub	221.7409	0.0000	68.1102	0.0000	26.8799	0.0000
Malin	79.3343	0.0000	67.8710	0.0000	60.5215	0.0000
Oneok	185.1969	0.0000	67.8716	0.0000	55.6605	0.0000
Opal	64.3699	0.0000	67.2952	0.0000	45.5946	0.0000
Waha Hub	266.4992	0.0000	67.8909	0.0000	138.4444	0.0000

[a] Likelihood Ratio (LR) tests are executed conditional on three cointegrating vectors and three lags.
[b] Likelihood Ratio (LR) tests are executed conditional on seven cointegrating vectors and three lags.
[c] Likelihood Ratio (LR) tests are executed conditional on four cointegrating vectors and three lags.

1% level (table 2.6). Regardless of the sub-period, no price series can be considered stationary by itself. AECO Hub, Malin, and Opal are weakly exogenous to the long-run relationship in the first sub-period; only AECO Hub are considered weakly exogenous in the second sub-period; and no price series are regarded as weakly exogenous in the third sub-period. For the sake of comparison and consistency, the eight prices are included in the VECM for every sub-period.

The long-run relationships in the first sub-period appear to be generally constant as most test statistics of both Z- and R-representations are below the 5% critical line (figure 2.3). The test statistics, however, spike during the period from the end of 1995 to the beginning of 1996. This spike is consistent with the test statistics for the entire sample (figure 2.2). The spike signals that something unusual might occur at the end of 1995. High natural gas prices because of cold weather that caused very rapid decline in natural gas stocks, which were already low because of irregularly cold weather in November and December 1995 (U.S. EIA 1996), may be a possible cause.

Test statistics are generally less than the 5% critical value for the second subsample except begining at the end of 2005 extending into 2007 (figure 2.4). Such inconsistency is also detected when testing constancy of β^* using the entire data set. Hurricanes Katrina and Rita in 2005 are possibly behind the instability. These hurricanes caused damage to the U.S. natural gas and petroleum infrastructure; many Gulf of Mexico wells, processing plants, and pipelines were closed (U.S. EIA 2010). In addition to the hurricane season, the increase in domestic production associated with shale gas are likely behind this inconstancy.

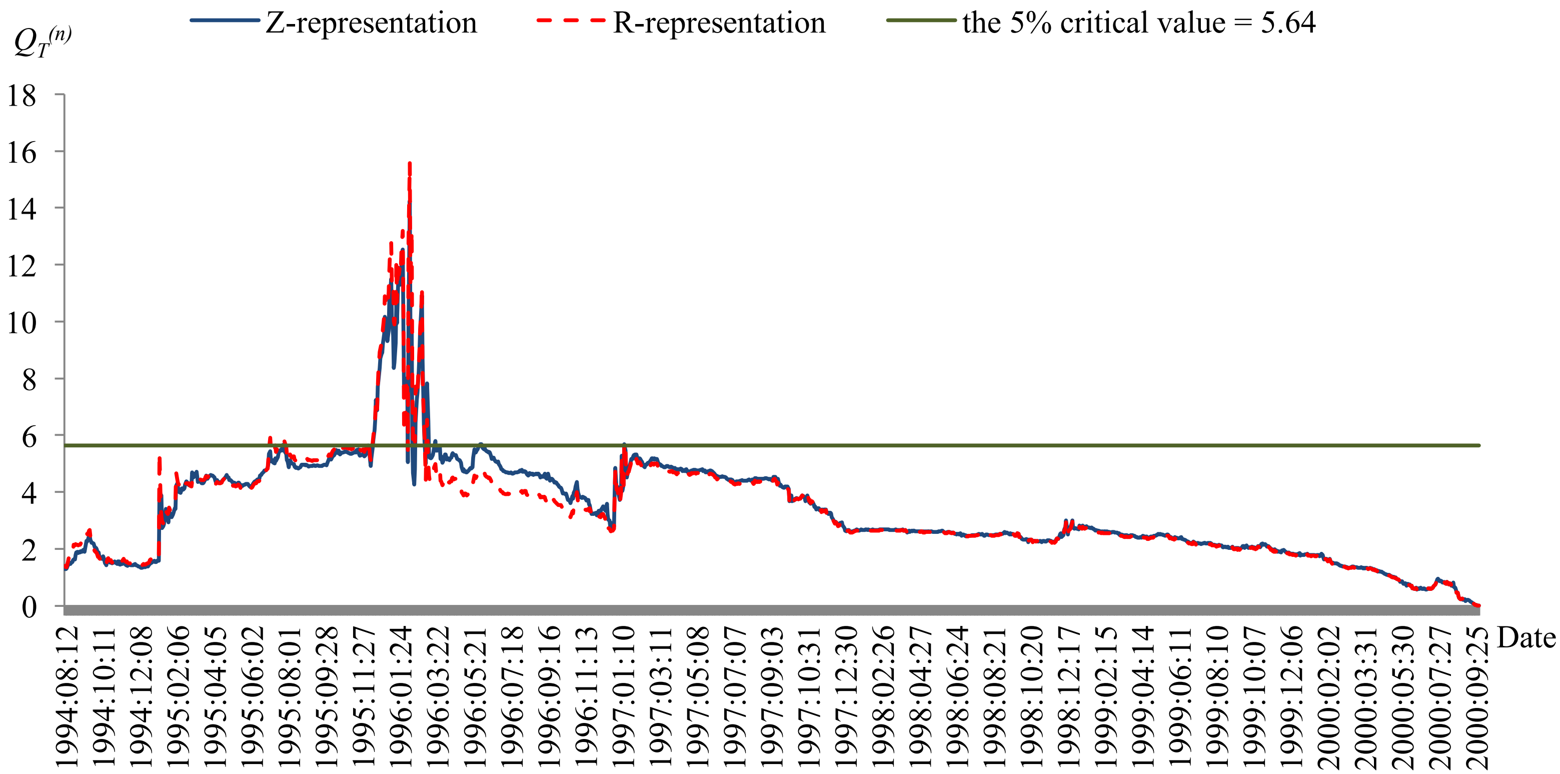

Figure 2.3. Plots of sup $Q_T^{(n)}$ for the first sub-period (May 3, 1994 - September 29, 2000)

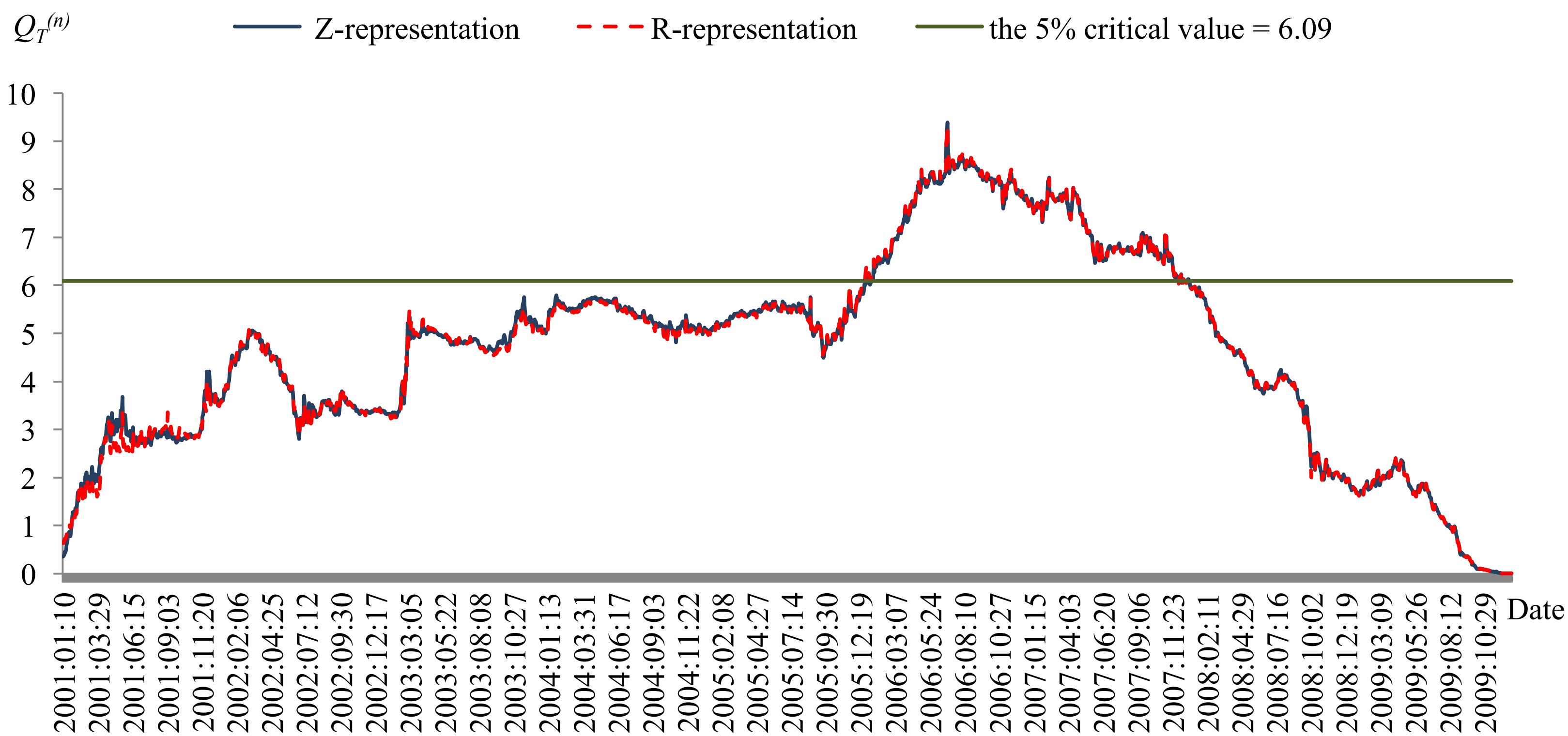

Figure 2.4. Plots of sup $Q_T^{(n)}$ for the second sub-period (October 2, 2000 - December 31, 2009)

U.S. natural gas gross withdrawals have increased; horizontal drilling and hydraulic fracturing are behind the increase in natural gas production (U.S. EIA 2011, 2014d). These techniques have allowed access to large volume of both oil and natural gas that were previously unprofitable to produce (U.S. EIA 2011). Large-scale natural gas production from shale started around 2000 as Mitchell Energy and Development Corporation developed a hydraulic fracturing technique that could economically produce commercial volumes of shale gas in the Barnett Shale (located in north-central Texas) (U.S. EIA 2011). Because of the profitability of the Barnett shale, other companies started applying the technique to the shale formations; as such by 2005 the Barnett Shale was producing almost half a trillion cubic feet (Tcf) of natural gas per year (U.S. EIA 2011).

Both the Z- and R-represention test statistics using the third subsample are around the borderline during 2012 but spike at the beginning of 2014 (figure 2.5). This spike is not seen when using the entire data set. The spike is most likely associated with the North Polar Vortex, which led to unusual extremely cold weather affecting a large part of Canada and the U.S. during the winter of 2013-2014, resulting in increased natural gas spot prices (U.S. EIA 2014e).

Transitory rejecting the null hypothesis of constancy of β^* in each sub-period should not be considered structural changes. When using the entire data set, the evidence of β^* inconstancy suggests that the potential presence of structural changes in pricing relationships among North America natural gas spot markets might occur during

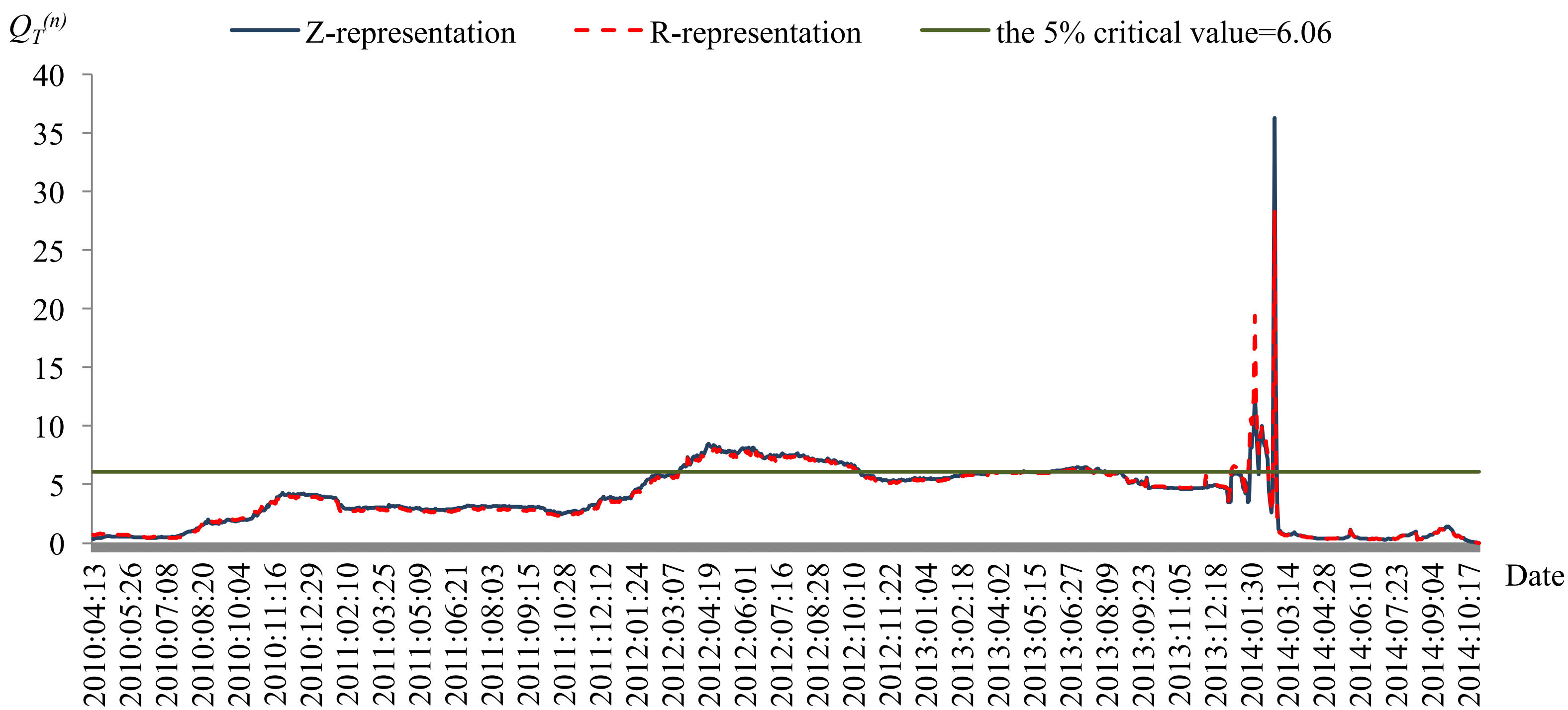

Figure 2.5. Plots of sup $Q_T^{(n)}$ for the third sub-period (January 1, 2010 - October 31, 2014)

2000 and again during 2009. The shift during 2000 to 2001 may be related to unexpectedly high and volatile natural gas prices (Alterman 2012; Joskow 2013). Henry Hub and the NYMEX futures prices clearly show a period of increased prices and volatility around this time period. In addition, the U.S. Federal Energy Regulatory Commission (FERC) Order No. 637, which involves removing some pipeline price ceilings, was enacted in 2000 (U.S. FERC 2000). Alterman (2012) suggests natural gas price volatility at the end of 2000 was due to the second coldest November on record since 1895. Joskow (2013, p. 340) notes, "…there had been a gas supply overhang during the 1990s and that as demand caught up with supply more expensive gas production sources would have to be relied upon to balance supply and demand, including more imports from Canada…" U.S. natural gas imports had been increasing (U.S. EIA 2015h). Ratios of U.S. natural gas imports to U.S. dry natural gas production are high in 2000 relative to during the 1990s and peak during 2005-2007 (figure 2.6). The increases in imports might be a sign of market instability, as the U.S. natural gas industry become more critically dependent on imports. In the entire sample (figure 2.2), the test statistics are a borderline case around 2009 and below the 5% critical value after 2009. Inference is that the long-run relationships changed after 2009. The U.S. EIA (2011) claims that shale resource is a "game changer" for the U.S. natural gas market. Because of the increased domestic natural gas production, the U.S. becomes less import-reliance and is expected to become a net exporter in natural gas. Ratios of U.S. natural gas imports to U.S. dry natural gas production have been decreasing since 2009 (figure 2.6).

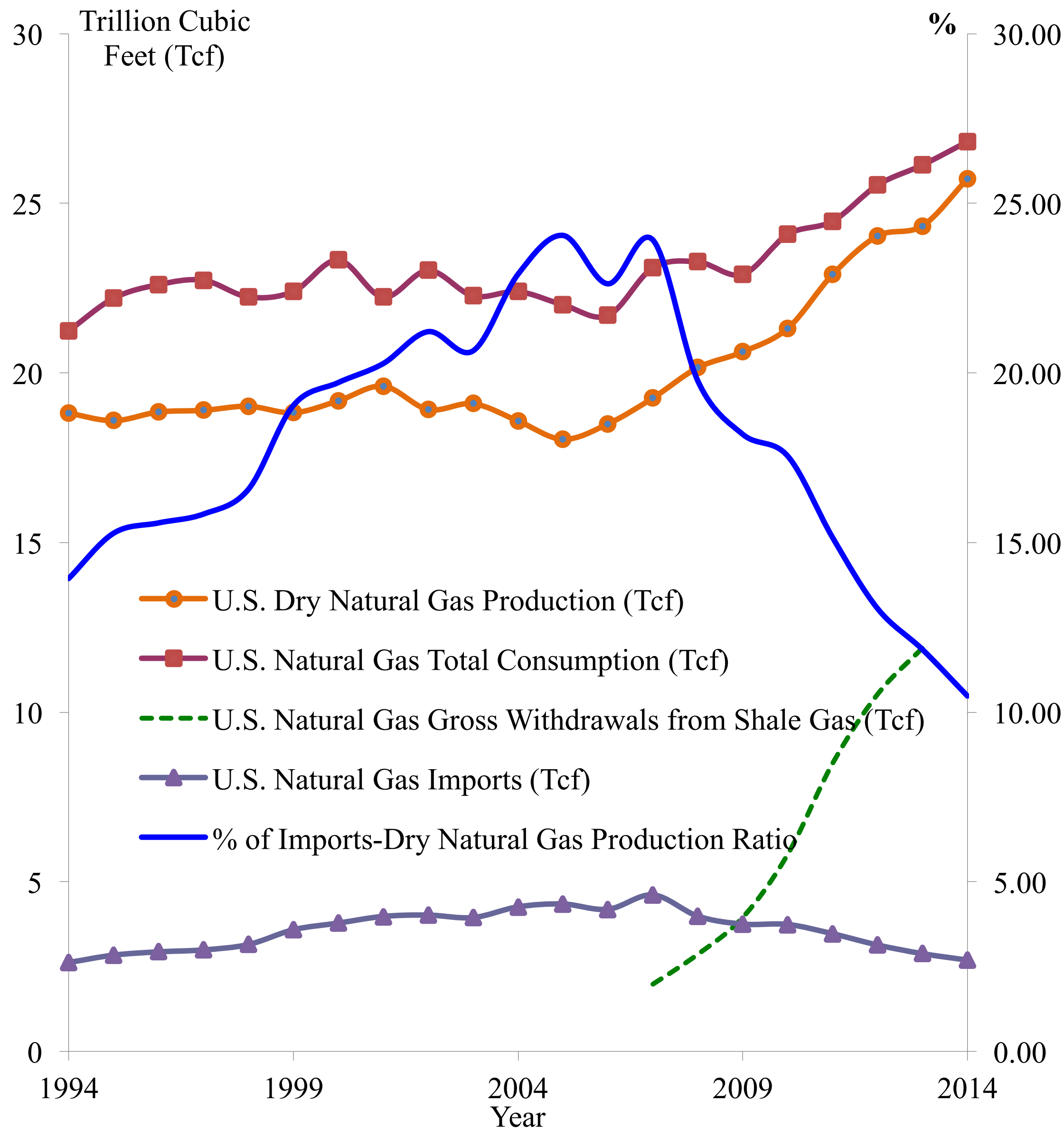

Figure 2.6. U.S. annual dry natural gas production, natural gas total consumption, gross withdrawals from shale gas, imports, and percentages of natural gas imports to dry natural gas production ratio (1994-2014) (U.S. EIA 2015c, 2015d, 2015h)
Note: The U.S. EIA started reporting U.S. natural gas gross withdrawals from shale gas in 2007

Evidences of Structural Changes from Other Tests

Even though it may not be perfectly comparable, the estimation of structural break dates proposed by Lüthepohl, Saikkonen, and Trenkler (2004) yields similar results; the estimation suggests the potential break dates on August 26, 1998, August 26, 2009, and December 12, 2013. The structural shift in 2009 may be a result of growing domestic natural gas production. The North Polar Vortex seems to be behind the shift at the end 2013.

Applying structural break tests suggested by Bai and Perron (1998, 2003) on each price series reveals three sequentially determined breaks for all price series except Opal which has four breaks. Possible break dates vary, but are similar across the eight series. The first break dates of the eight price series are generally around the beginning of 2000; the second break dates are during 2003 to 2004; and the last break dates are around the beginning of 2009. The first and the third break dates are roughly close to those found in the previous section.

Contemporaneous Causal Flows

Eight residual series from the estimated VECM are found to have non-Gaussian distribution, as the null hypothesis of Jarque-Berra normality test on each residual series is rejected and their histograms with overlaid Gaussian distributions in figure 2.7 reveal kurtosis. Based on the LinGAM algorithm executed in Tetrad version five with one prune factor,[11] DAGs are employed to identify restrictions for generating orthogonal innovations. As there exist large deviations from the null of parameter constancy when

[11] Prune factor is the threshold of pruning edges.

using the entire data,[12] innovation accounting analysis is performed for each sub-period. DAGs for the three subsamples are given in figures 2.8, 2.9, and 2.10.

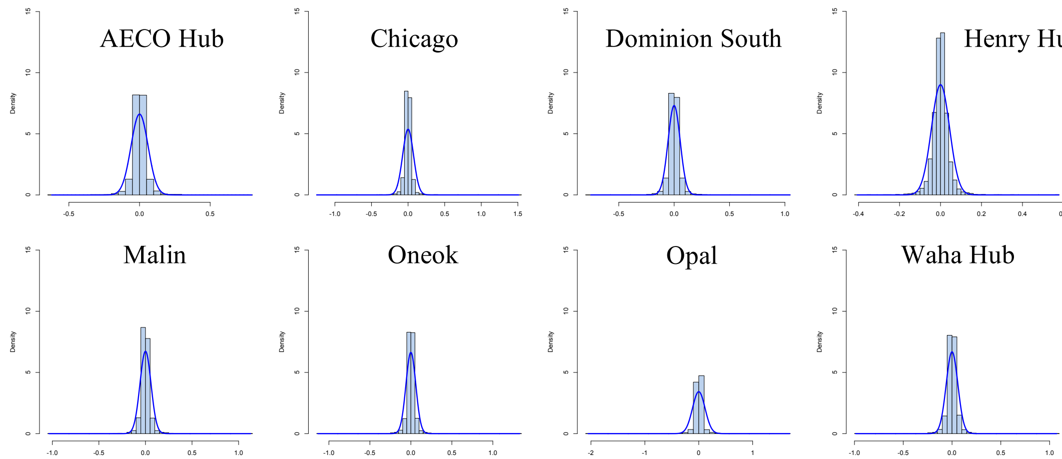

Figure 2.7. Histograms of eight residual series from the estimated VECM using the entire data

The eight markets are linked together by 11 or 12 contemporaneous causal flows regardless of the sub-period. Of the twelve causal flows in the first and the last sub-periods, seven are the same. The third sub-period added causal flows between Malin and AECO Hub and Malin and Henry Hub but removed a causal flow between Chicago and Dominion South. Causal flows in the middle sub-period are generally different than the other two sub-periods. The only causal flow that is the same in all periods is from Henry Hub to Chicago.

[12] DAG using the entire data set are shown in Appendix B (figure B.2).

Figure 2.8. Contemporaneous casual flows for the first sub-period (May 3, 1994 - September 29, 2000)

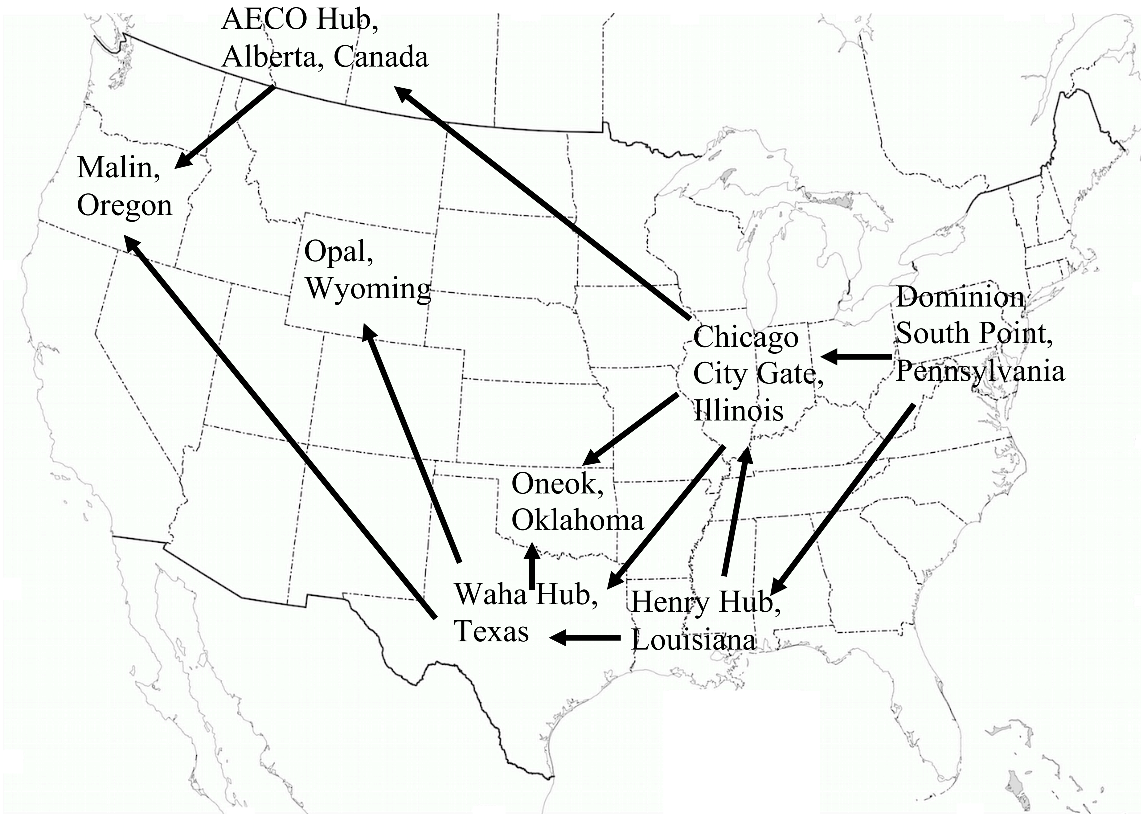

Figure 2.9. Contemporaneous casual flows for the second sub-period (October 2, 2000 - December 31, 2009)

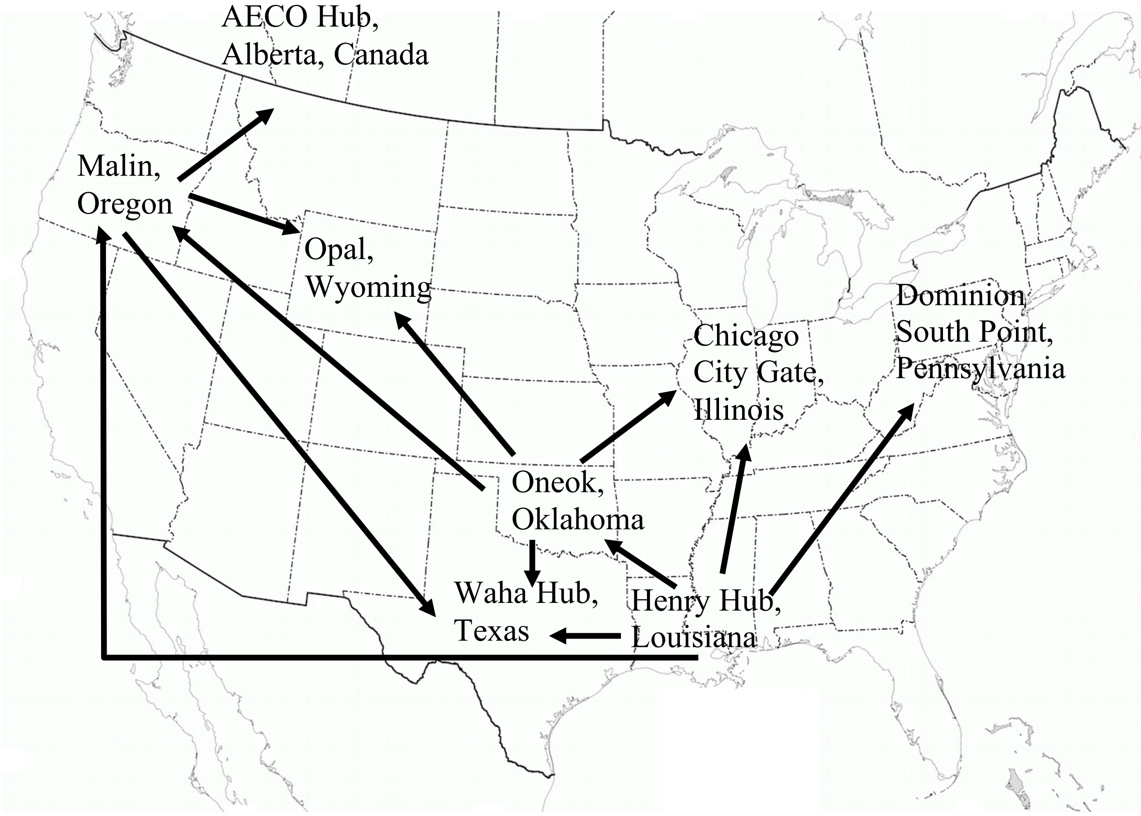

Figure 2.10. Contemporaneous casual flows for the third sub-period (January 1, 2010 - October 31, 2014)

Consistent with results from testing weak exogeneity for long-run parameter in the first and second sub-periods, innovations of AECO Hub do not contemporaneously cause any other markets and do not respond to any other markets' shock in the first sub-period; AECO Hub is exogenous in contemporaneous time. AECO Hub contemporaneously receives information from Chicago and sends to Malin in the second sub-period. The causal flow between AECO Hub and Malin switches direction in the third sub-period and the contemporaneous link to Chicago disappears.

Information flows from Henry Hub to Chicago in every sub-period. Chicago contemporaneously responds to information from Oneok in the first and the third sub-periods, but the information flow changes direction in the second sub-period. In the second sub-period, information flow from Chicago influences not only Oneok but also AECO Hub and Waha Hub. In the first sub-period, Chicago transmits information to Dominion South, whereas, in the second sub-period, Dominion South transmits information to Chicago. Nonetheless, there is no information transmission between Chicago and Dominion South in the third sub-period. Dominion South exchanges contemporaneous information with Henry Hub in every sub-period; Dominion South acts as a recipient in the first and the third sub-period, but becomes a contributor in the second sub-period.

Henry Hub behaves as both receiver and provider of information in the first and the second sub-periods. In the first sub-period, Henry Hub receives information from Oneok and Waha Hub and sends information to Chicago and Dominion South. In the second sub-period, Henry Hub obtains information from only Dominion South and transmits it to Chicago and Waha Hub. Henry Hub behaves solely as a sender, conveying information to Chicago, Dominion South, Malin, Oneok, and Waha Hub in the third sub-period.

Malin gathers information from Oneok and Opal and transmits to Waha Hub in the first sub-period. In the second sub-period, Malin, however, is influenced by information from Waha Hub as well as AECO Hub. Malin receives information from

Henry Hub and Oneok and provides information to AECO Hub, Opal, and Waha Hub in the third sub-period.

Oneok behaves solely as a sender of information in the first sub-period; information from Oneok influences Chicago, Henry Hub, Malin, Opal, and Waha Hub. Conversely, Oneok behaves solely as a receiver (endogenous) in the second sub-period; it receives information from Chicago and Waha Hub. In the third sub-period, Oneok is both a sender and receiver; supplying information to Chicago, Malin, Opal, and Waha Hub and obtaining information from Henry Hub.

Information flows from Oneok to Opal in the first and the third sub-periods; whereas, contemporaneous interaction between these two markets disappears in the second sub-period. Similarly, Opal and Malin have connections in the first and third sub-periods; Opal acts as a sender in the first sub-period while becomes a receiver in the third sub-period. Opal affects Waha Hub in the first sub-period but the information flow switches direction in the second sub-period; no communication between these two markets exists in the third sub-period.

Waha Hub obtains information from Malin, Oneok, and Opal and provides information to Henry Hub in the first sub-period. These information flows change direction in the second sub-period. In the third sub-period, Waha Hub is influenced by innovations from Henry Hub, Malin, and Oneok.

Impulse Response Functions

Impulse response functions provide the dynamic responses of each series to a one-time shock in each series. For comparision purposes, the responses are normalized such that each response is divided by the standard error of its innovations. Each sub-graph provides the response of the market given by the row heading to a one-time shock in the series listed in the column heading.

The dynamic price system is stable in each sub-period. Irrespective of the sub-period, the eight markets positively respond to a shock of its own price. Innovations of Oneok appears to be the most important in the first sub-period; Dominion South appears to be the most significant market in the sub-period; and Henry Hub appear to be the most essential market in the third sub-period,[13] as shocks in these markets create relatively large impacts on other markets.

In the first sub-period (figure 2.11), AECO Hub responses to shocks in the other markets are small; similarly, other markets barely respond to a shock in AECO Hub. Responses of Malin, Oneok, Opal, and Waha Hub to a shock in Chicago are initially postitive and then go negative a few days after the shock. A shock of Dominion South largely impacts itself, Chicago, and Henry Hub. Chicago and Dominion South react positively to a shock in Henry Hub, while AECO Hub, Malin, and Opal responses to such a shock are small. Except itself, other markets have almost no responses to a shock in Malin; this is consistent with results indicating that Malin can be excluded and considered weakly exogenous. All markets except AECO Hub have relatively large responses to a shock in Oneok. A shock in Opal positively affects other markets, especially Malin, Oneok, and Waha Hub. Among others, Henry Hub responds to a

[13] Impulse response functions using the entire data set suggest Henry Hub is the most important market (figure B.3. in Appendix B).

Figure 2.11. Impulse response functions of eight natural gas spot prices for the first sub-period (May 3, 1994 - September 29, 2000)

Figure 2.12. Impulse response functions of eight natural gas spot prices for the second sub-period (October 2, 2000 - December 31, 2009)

Figure 2.13. Impulse response functions of eight natural gas spot prices for the third sub-period (January 1, 2010 - October 31, 2014)

shock in Waha Hub the most. Chicago, Dominion South, and Oneok initially respond positively to a shock in Waha Hub.

In the second sub-period (figure 2.12), after the first day of the shock, other markets slightly respond positively to a shock in AECO Hub. Unlike the first sub-period, all markets have positive responses to a shock in Chicago. Relative to shocks in other markets, a shock in Dominion South causes the largest responses in all markets. A shock in Henry Hub initially affects Dominion South negatively, but a couple days after the shock the response is positive while the effects on the other markets are always positive. Similar to the first sub-period, all markets but itself responses to a shock in Malin are small. Oneok, Opal, and Waha Hub importance seems to have decreased as market responses to shocks in these markets have decreased.

In the third sub-period, all markets react to a shock in Henry Hub (figure 2.13). Responses to a shock in Henry Hub are larger than those to shocks in other markets. There appears to be an increasing importance of Malin as the market responses are generally larger in the third period than in either of the other periods. Responses of other markets to shocks in Chicago, Dominion South, Opal, and Waha Hub are small. In all markets, there appears to be non-lasting impact of a shock in Oneok.

Forecast Error Variance Decompositions

To determine how the forecast error variance of each price series depends on its own innovations and other price series' innovations, forecast error variance decompositions at horizons of one, five, and 10 trading days ahead are provided. Values in each row indicate, at a specific time horizon, how much variation in each price series is due to itself and the other price series; the sum of the values in each row must be 100.

Consistent with evidence from impulse response functions, Oneok appears to be the most important in the first sub-period; Dominion South appears to be the most significant market in the sub-period; and Henry Hub appears to be the most essential in the third sub-period[14] as innovations from these markets are the main factor inducing price variation in other markets. In the second sub-period, the importance of Oneok, Opal, and Waha Hub in explaining price uncertainty in other markets is generally small relative to the first sub-period; Oneok, Opal, and Waha Hub become more influential in the third sub-period relative to the second sub-period.

In the first sub-period, at any time horizon, price variation in AECO Hub is predominantly due to itself (table 2.7). At one day ahead, variation of prices in Chicago largely comes from its own shock (49%) and Oneok's shock (47%). Similarly, at five and 10 days ahead, variation in Chicago is primarily because of itself and Oneok; Oneok provides a larger influence than Chicago for five and 10 days ahead. The uncertainty of prices in Dominion South at one day ahead is primarily due to the innovations of itself (64%), Oneok (19%), and Henry Hub (11%). At five and 10 days ahead, the innovation of Henry Hub becomes primary source of price uncertainty in Dominion South. Oneok is the main cause of the price uncertainty in Henry Hub at any time horizon. For Malin, its own shock, Opal's, and Oneok's mostly explain the deviation in prices at any time horizon. At one day ahead, variation of Oneok natural gas spot prices is dependent only

[14] Using the entire data set, it appears that Henry Hub is the most important market influencing price variation in other markets (table B.2 in Appendix B).

Table 2.7. Forecast Error Variance Decompositions of Eight Natural Gas Spot Prices for the First Sub-Period (May 3, 1994 - September 29, 2000)

Horizon	AECO Hub	Chicago	Dominion South	Henry Hub	Malin	Oneok	Opal	Waha Hub
				AECO Hub				
1	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	97.68	0.18	0.23	0.42	0.02	0.06	0.87	0.53
10	97.02	0.37	0.25	0.62	0.02	0.06	0.99	0.68
				Chicago				
1	0.00	49.39	0.00	2.59	0.01	47.21	0.08	0.73
5	0.45	27.20	3.15	13.61	0.02	45.63	5.70	4.25
10	0.48	18.11	5.52	22.54	0.01	42.97	6.00	4.38
				Dominion South				
1	0.00	1.63	64.12	11.47	0.04	19.17	0.34	3.23
5	0.37	2.10	33.26	29.49	0.01	25.36	4.66	4.76
10	0.42	1.15	25.07	40.14	0.01	24.55	4.38	4.30
				Henry Hub				
1	0.00	0.00	0.00	42.76	0.16	43.79	1.27	12.04
5	0.15	0.22	2.00	39.80	0.06	39.14	6.75	11.88
10	0.11	0.35	3.76	37.63	0.04	41.84	6.45	9.82
				Malin				
1	0.00	0.00	0.00	0.00	57.50	15.57	26.93	0.00
5	0.01	0.87	0.01	0.07	45.12	15.23	38.66	0.04
10	0.01	1.43	0.04	0.04	44.15	14.97	39.25	0.11
				Oneok				
1	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
5	0.20	1.71	1.11	1.42	0.14	76.77	12.87	5.78
10	0.19	3.98	1.77	2.10	0.15	72.42	13.73	5.67
				Opal				
1	0.00	0.00	0.00	0.00	0.00	20.73	79.27	0.00
5	0.01	0.74	0.03	0.03	0.19	19.42	79.47	0.12
10	0.02	1.25	0.06	0.01	0.15	19.72	78.45	0.34
				Waha Hub				
1	0.00	0.00	0.00	0.00	0.46	59.96	3.77	35.81
5	0.14	1.49	0.78	1.33	0.59	56.96	16.14	22.58
10	0.08	3.80	1.51	1.59	0.58	60.93	14.88	16.63

on its own innovation (100%). At five and 10 days ahead, Oneok still explains the majority of its own variation, but Opal and Waha Hub increase in importance. The variation of prices in Opal is primarily explained by itself with Oneok explaining approximately 20% at any time horizon. For Waha Hub, forecast error variance is explained by Oneok, Opal, and itself. Oneok explains more variance in Waha Hub than any other market including Waha Hub itself.

In the second sub-period, AECO price variance is explained by itself, Dominion South, Chicago, and Henry Hub at all time periods; Malin, Oneok, Opal, and Waha Hub explain very little (table 2.8). At one day ahead, innovations in Chicago itself, Dominion South, and Henry Hub influence the uncertainty of prices in Chicago. The most significant source of the uncertain of prices in Chicago at any time horizon is Dominion South. In addition, at any time horizon, Dominion South is the leading source of the price variance in Henry Hub, Oneok, and Waha Hub. Price variation in Dominion South is totally the result of its own innovation at one day ahead; at five and 10 days ahead Dominion South remains largely exogenous. Price variation in Malin is primarily because of itself and Dominion South at any time horizon; innovations in Oneok and Opal cause very little on price variation in Malin. Price uncertainty in Opal is largely due to itself, Dominion South, and Waha. Other than Malin, all markets contribute to uncertainty in Oneok and Waha Hub price variations.

In the third sub-period, at one day ahead, the uncertainty of AECO prices is due to its own innovation and innovations of Henry Hub, Malin, and Oneok; at five and 10 days ahead, the uncertainty is also primarily affected by its own shock and shocks in

Table 2.8. Forecast Error Variance Decompositions of Eight Natural Gas Spot Prices for the Second Sub-Period (October 2, 2000 - December 31, 2009)

Horizon	AECO Hub	Chicago	Dominion South	Henry Hub	Malin	Oneok	Opal	Waha Hub
				AECO Hub				
1	54.02	7.47	35.24	3.27	0.00	0.00	0.00	0.00
5	47.25	10.29	37.43	3.68	0.88	0.45	0.02	0.01
10	46.58	9.10	37.66	4.76	1.26	0.62	0.01	0.01
				Chicago				
1	0.00	16.26	76.64	7.10	0.00	0.00	0.00	0.00
5	3.83	17.05	70.53	6.60	0.38	1.54	0.04	0.02
10	6.41	13.45	69.09	8.59	0.46	1.90	0.07	0.03
				Dominion South				
1	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
5	2.72	1.99	93.25	0.76	0.03	0.99	0.07	0.20
10	4.77	2.15	88.74	2.71	0.04	1.33	0.09	0.17
				Henry Hub				
1	0.00	0.00	78.66	21.34	0.00	0.00	0.00	0.00
5	2.69	1.41	80.26	14.22	0.03	1.23	0.08	0.08
10	4.86	1.44	77.21	14.86	0.05	1.39	0.13	0.06
				Malin				
1	5.94	2.76	26.06	3.53	56.86	0.00	0.00	4.86
5	12.58	5.08	23.24	3.37	52.17	0.09	0.06	3.41
10	14.74	5.03	24.58	4.30	47.61	0.16	0.05	3.52
				Oneok				
1	0.00	4.80	51.31	7.30	0.00	24.03	0.00	12.55
5	2.04	6.89	49.65	6.52	0.09	26.05	0.05	8.70
10	3.81	6.26	48.93	7.54	0.21	25.47	0.05	7.73
				Opal				
1	0.00	0.31	5.30	0.88	0.00	0.00	90.89	2.62
5	1.74	1.46	9.00	1.73	0.43	2.31	80.01	3.32
10	4.35	1.45	9.97	2.30	0.64	3.73	75.01	2.55
				Waha Hub				
1	0.00	3.38	58.21	9.67	0.00	0.00	0.00	28.74
5	2.41	7.00	60.77	7.67	0.09	4.57	0.10	17.40
10	4.45	6.32	59.96	8.73	0.21	6.60	0.15	13.59

Table 2.9. Forecast Error Variance Decompositions of Eight Natural Gas Spot Prices for the Third Sub-Period (January 1, 2010 - October 31, 2014)

Horizon	AECO Hub	Chicago	Dominion South	Henry Hub	Malin	Oneok	Opal	Waha Hub
				AECO Hub				
1	44.44	0.00	0.00	24.54	13.57	17.45	0.00	0.00
5	59.85	0.02	0.11	14.72	14.28	9.09	0.15	1.79
10	62.15	0.02	0.14	14.73	15.43	5.24	0.26	2.04
				Chicago				
1	0.00	72.10	0.00	19.37	0.00	8.53	0.00	0.00
5	5.68	60.49	0.49	18.40	3.98	9.51	0.55	0.92
10	5.78	52.13	0.37	24.51	7.05	8.14	0.45	1.57
				Dominion South				
1	0.00	0.00	58.13	41.87	0.00	0.00	0.00	0.00
5	1.76	0.39	47.11	46.27	0.43	0.80	1.27	1.96
10	2.04	0.23	45.54	45.99	0.69	1.33	1.51	2.67
				Henry Hub				
1	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
5	5.99	0.26	0.10	87.49	0.77	1.85	2.21	1.34
10	6.92	0.18	0.06	83.90	1.39	2.78	2.93	1.85
				Malin				
1	0.00	0.00	0.00	44.17	24.42	31.40	0.00	0.00
5	9.79	0.03	0.11	39.92	25.52	21.40	0.62	2.62
10	10.19	0.17	0.09	45.52	25.42	14.23	1.16	3.22
				Oneok				
1	0.00	0.00	0.00	51.26	0.00	48.75	0.00	0.00
5	6.66	0.08	0.17	54.11	1.22	35.28	0.20	2.30
10	6.89	0.28	0.15	62.24	1.88	25.52	0.30	2.74
				Opal				
1	0.00	0.00	0.00	46.26	2.16	40.61	10.98	0.00
5	7.05	0.16	0.16	45.72	6.20	29.80	8.41	2.50
10	7.43	0.61	0.12	49.49	9.46	21.58	8.09	3.22
				Waha Hub				
1	0.00	0.00	0.00	51.02	1.42	23.48	0.00	24.08
5	6.21	0.02	0.09	58.92	2.02	21.37	0.38	10.99
10	6.61	0.11	0.06	66.01	2.63	15.84	0.51	8.23

these three markets (table 2.9). The primary causes of price uncertainty in Chicago are its own innovations, Henry Hub, and Oneok at any time horizon. Different than the second sub-period, at one, five, and 10 days ahead, innovations of Dominion South and Henry Hub explain price uncertainty in Dominion South. Unlike the second sub-period, Henry Hub appears to primarily be exogenous to the system. Price uncertainty in Malin is generated mostly by innovations in Henry Hub, Oneok, and Malin itself. Uncertainty in Oneok prices is triggered by shocks in Henry Hub and itself. Along with its own and Malin's, innovations of Henry Hub and Oneok play an important role in explaining the uncertainty of Opal prices. Price variation in Waha Hub is largely due to innovations in Henry Hub, Oneok, itself, and Malin.

Discussion

Tests for constancy of β^*, which are the long-run relationship parameters, are used to discover the possible existence of structural changes in pricing relationships among the eight North America natural gas spot markets during 1994 to 2014. Instability of β^* indicates that the long-run pricing relationships among natural gas spot markets in North America change around 2000 and 2009. The data is split into three sub-periods to investigate price dynamics.

Regardless of the sub-period, consistent with findings in the literature, adjacent markets appear to provide more price information to each other than to markets that are located far apart. AECO Hub has provided less information to other markets; including other markets in Canada may provide different inferences. Evidence of information flows corresponding to trading hours beginning in the eastern markets and moving to the western markets is found in the second sub-period. Such casual flows are less pronounced in the first and third sub-periods.

Termination of the wellhead natural gas price regulation occurred by the end of 1992, therefore, the first sub-period (May 1994 – September 2000) is the phase that the natural gas industry was maturing and becoming competitive as a result of the development of natural gas trading hub and natural gas spot, term, and derivatives markets (Joskow 2013). Innovations in Oneok, Oklahoma generally influence price dynamics in most of the eight markets in the first sub-period. This may be because Oklahoma has been one of the largest natural gas producing states in the U.S. and 17 of the 100 largest natural gas reserves in the U.S. are located in Oklahoma (U.S. EIA 2014i). Moreover, among other natural gas producing states, in the first sub-period, Oklahoma was nearest to Illinois, which is a key transportation hub for natural gas with more than 12 interstate natural gas pipelines and two natural gas market centers (U.S. EIA 2015f).

Natural gas prices were expensive and volatile during 2000. The U.S. natural gas market was more import-intensive in the second sub-period (October 2000 – December 2009), as ratios of U.S. natural gas imports to U.S. dry natural gas production were high relative to other period. The study of price dynamics indicates that Dominion South plays an important role in the second sub-period and becomes more independent in the third sub-period. Pennsylvania is mostly likely an excess demand area in the second sub-period, but is an excess supply in the third sub-period. This switch is because natural gas production in Pennsylvania has dramatically increased with the development of the Marcellus shale formation. The growth in Marcellus shale gas production has changed U.S. natural gas transportation patterns east of the Mississippi River, where great volumes of natural gas produced in Texas, Louisiana, and Oklahoma were historically transported to (U.S. EIA 2014b). The change, furthermore, is resulting in the bidirectional natural gas pipeline project in the Northeast; the plan is to expand existing systems and construct new systems to transport natural gas produced in the Northeast to consuming markets outside the region (U.S. EIA 2014a).

Tests for parameter constancy reveals the constancy of β^* after 2009. Shale gas production is possibly behind this change as technological advancement are leading to accessing large-scale natural gas which has augmented domestic natural gas production. Larger stable supplies encourage market stability, as the natural gas industry becomes less import-reliance. Henry Hub is a dominant market as its innovation causes the price dynamics in most natural gas spot prices in the third sub-period (January 2010 – October 2014). This is not surprising because Henry Hub is noted in the literature as an important market for pricing of the North America natural gas spot and futures markets[15]. Malin comes to be more dependent in the third sub-period. This may be because natural gas is delivered to California, which is one of the top five natural gas consuming states, through Malin (U.S. EIA 2014g). Most electric power generating plants in California are natural gas-fired while California's natural gas production has gradually declined (U.S. EIA 2014g).

[15] The first natural gas futures contract was issued by the New York Mercantile Exchange Market (NYMEX) in 1990 (U.S. EIA 2010).

CHAPTER III

PREQUENTIAL FORECASTING ANALYSIS OF RETURNS IN NORTH AMERICA NATURAL GAS SPOT MARKETS

Prequential data analysis introduced by Dawid (1984) is applied to evaluate predictive distributions for out-of-sample data of returns in North America natural gas spot markets. Because of the potential presence of structural changes in pricing relationships among North America natural gas spot markets found in Chapter II, vector autoregressive (VAR) models are estimated for three in-sample periods: (1) May 4, 1994 to October 31, 2014; (2) October 2, 2000 to October 31, 2014; and (3) January 1, 2010 to October 31, 2014. The objective is to determine whether and how the potential presence of structural breaks affects out-of-sample probability forecasting performance. To address the objective, calibration measures (calibration plots and chi-squared goodness-of-fit test statistics), root mean-squared error, the Brier score and its decompositions, and the ranked probability score are applied for model assessments.

Prequential Analysis Studies

Dawid (1984) introduced the prequential approach under the aims of generating forecasts, proposing appropriate measures of the uncertainty related to unknown events or quantities, and exploiting the sequential nature in forecasting. Because the uncertainty of forecasts can be expressed as probabilities, forecasts are given as probability distributions over unknown or uncertain events (probability forecasting).

Dawid (1984) suggests that a forecast for the next value should be based on an analysis of earlier values and calls this prequential (predictive sequential) forecasting.

The adequacy of prequential probabilities can be assessed by using probability calibration (Dawid 1984). Calibration is the ability of a model's forecasted probability distribution to correspond to the *ex post* relative frequency of all events. A forecasting model is said to be well calibrated when the *ex post* relative frequency of all events, whose probability is assigned a probability of P^*, is P^*. Kling and Bessler (1989) employ probability forecasting to interest rates, money stock, consumer prices, and industrial production. They test for calibration and develop a procedure for recalibrating distributions, based on the bias estimated in previous distributions. Recalibration of the forecasts provides improved results. Bessler and Kling (1990) investigate prequential relationships between cash prices and futures prices for cattle and find that daily futures prices help to forecast daily cash prices. Estimating both univariate and multivariate (bivariate) models, they find that the multivariate model provides more information than the univariate model on the predictive distribution of cash prices; while the multivariate model does not provide additional information in forecasting futures prices. Standard mean-squared error and probability calibration measures (calibration plots and chi-squared goodness-of-fit test statistics) are used to measure performance of probability forecasts.

An alternative to the mean-squared error test for evaluating probabilistic forecasts is the mean probability score, known as the Brier score (Brier 1950). The Brier score is a quadratic scoring measure, which can be partitioned into components that indicate calibration and resolution (Bessler and Ruffley 2004). Resolution is the ability of a model in sorting or partitioning uncertain events into disjointed subgroups that have probability measures differing from long-run relative frequencies. Calibration measures cannot capture this sorting ability. The Brier score, therefore, provides more information of predictive performance than do calibration metrics.

Zellner, Hong, and Min (1991) use the Brier score to rank probability forecasts of turning points in the growth rates of 18 countries from various fixed and time-varying-parameter models. They find that time-varying parameter models perform marginally better than the fixed parameter models. In Bessler and Ruffley (2004), an ordinary least squares model and a random walk model are used to forecast the U.S. stock market returns. Results from calibration measures and the Brier score and its partition reveal that the OLS model tends to perform better than the random walk model. Studying probability forecasts of inflation and GDP, Casillas-Olvera and Bessler (2006) evaluate the probability forecasts of the Monetary Policy Committee and those of the group of undisclosed external forecasters using the Brier score and its partition. It appears that both the Monetary Policy Committee and the other forecasters respond to information not related to the forecasted variable.

Methodology

The VAR model is used to generate forecasts of natural gas returns. The general VAR model is

(3.1) $\Phi_{\mathrm{t}}(B)X_t = \epsilon_t,$

where $\Phi_t(B)$ is the autoregressive parameter matrix, X_t is a vector of considered series, and ϵ_t is a vector of innovations that are uncorrelated over time, but may be contemporaneously correlated (Hamilton 1994).

Equation (3.1) is first estimated for each in-sample data period using OLS. Then one-step-ahead out-of-sample forecasts are generated. At each time period, the model's parameters are updated before generating the next forecasts. Instead of point forecasts, probabilistic forecasts are generated because of the uncertainty associated with the parameters and error term. To deal with uncertainty in $\Phi_t(B)$ and ϵ_t, the procedure suggested by Fair (1986) is implemented. At each time *t* the elements of $\Phi_t(B)$ are assumed to be normally distributed with mean $\Phi_t(B)$ and variance-covariance matrix $V_t = PP'$. Updating equation (3.1) with the Kalman filter at each *t* after the initial in-sample estimation, which allows for a small degree of time variation in the parameters, yields the estimated parameter matrix $\widehat{\Phi}_t(B)$. A particular draw $\Phi_t^*(B)$ is given as

(3.2) $\Phi_t^*(B) = \widehat{\Phi}_t(B) + P_t e,$

where e is a vector of standard normal draws. Uncertainty in innovations is modeled by drawing from the normal distribution with mean zero and variance-covariance matrix equal to the estimated variance-covariance matrix for one-step forecast errors. A one-step-ahead forecast vector is given as

(3.3) $\hat{X}_{t+1} = \Phi^*(B)X_t + \epsilon_{t+1}^*.$

One thousand point forecasts of $\hat{X}_{t+1}$ are obtained by drawing e and ϵ_{t+1}^{*} 1000 times at each *t*. At the next *t*, the model is updated using the Kalman filter estimator to generate a new set of probability forecasts.

Probability Forecasting

Let $x_{i,t}, i = 1, \ldots, m, t = 1, \ldots, T$, where T is the number of in-sample data be an observed element of the *m* x 1 vector of time series X_t. At time *T*, values for $X_t, t = 1, \ldots, T$, are known or observed. A set of probability distributions P_{T+k} for the unknown values X_{T+k} can be generated. A prequential forecasting system is defined by a rule which associates a choice of $P_{n+j}, n = T+1, \ldots, T+K$, where *K* is the number of out-of-sample data points, *j* indicates forecast horizon, for each *n* with any possible set of outcomes X_{n+j} (Dawid 1984; Kling and Bessler 1989).

If the $x_{i,n+j}$ (time series *i*, forecast horizon *j*) are continuous random variables with continuous distribution function, $F_{i,n+j}$, the random fractiles, $u_{i,n+j} = F_{i,n+j}(x_{i,n+j})$, are independent uniform *U*[0, 1] random variables (Dawid 1984). If the $x_{i,n+j}$ are discrete with cumulative distribution functions, $F_{i,n+j}$, then the random fractiles, $u_{i,n+j}$, have distribution functions of the form $G(u_{i,n+j}) = u_{i,n+j}$, even though the functions are not continuous. In either case, the assessment of the prequential forecasting system reduces to a test of the hypothesis that the observed sequence $u_{i,n+j} = F_{i,n+j}(x_{i,n+j})$ is from a probability distribution with the cumulative distribution $G(u_{i,n+j}) = u_{i,n+j}$. The prequential forecasting system is considered well calibrated when this hypothesis cannot be rejected.

An estimated cumulative distribution function, $\hat{G}(u_{i,n+j})$, for $u_{i,n+j}$, is obtained by taking the observed sequence $u_{i,n+j}$=$F_{i,n+j}(x_{i,n+j})$, sorting the sequence of $u_{i,n+j}(1), \ldots, u_{i,n+j}(K)$ in ascending order, and calculating

(3.4) $\hat{G}(u_{i,n+j}(k)) = (k/K); \qquad k = 1, \ldots, K,$

where *K* is the number of out-of-sample observations. Equation (3.4) is referred to as the calibration function (Bunn 1984).

Probability Forecasting Assessment

Calibration measures (graphical representation and chi-squared goodness-of-fit), the quadratic loss measure, the Brier score and its partition, and the rank probability score are used to evaluate prequential forecasts in this study.

Calibration Measures

Testing the observed fractiles obtained from the sequence of estimated probability forecasts is a test of calibration (Dawid 1984). Graphical representation and a goodness-of-fit test statistic are commonly used.

The plot of relative frequency (y-axis) against the realized fractiles (x-axis) illustrates calibration performance. For a well-calibrated prequential forecasting system, the plot should approach a 45-degree line. Whether a particular plot deviates from the 45-degree line enough to reject calibration, however, is left to the analyst to decide. Graphical representations do not provide a statistic test.

If there is a sequence of *K* such forecasts, under the null hypothesis of well calibration, the observed fractiles are expected to follow the uniform distribution, such that any subinterval of the line (0, 1) of length *L* ($0 < L < 1$) has *L* x *K* observed fractiles.

If there are Q mutually exclusive and exhaustive subintervals, a chi-squared goodness-of-fit test statistic is

(3.5) $\chi^2 = \sum_{q=1}^{Q}[(a_q - L_q K)^2 / L_q K] \quad \sim\chi^2(Q-1)$,

where a_q is the actual number of observed fractiles in the interval q and L_q is the length of interval q. Under the hypothesis of well calibration and under the weak conditions that the independence of the distributions underlying the forecasts is not required, the test statistic is distributed as chi-squared with Q -1 degrees of freedom (Dawid 1984).

Quadratic Loss and the Brier Score

The quadratic loss function is the most popular criterion for evaluating predictive distributions. Similar to point forecast, mean-squared error (MSE) criterion can also be applied to probability forecasts. MSE for probability forecasts is

(3.6) $MSE(P; x_{i,n+j}; n = T+1, \ldots, T+K) = 1/K \sum_{n=T+1}^{T+K} (x_{i,n+j} - \bar{P}_{i,n+j})^2$,

where $\bar{P}_{i,n+j}$ represent the expected value from the distribution $P_{i,n+j}$.

An alternative test for assessing probability forecasts, which is similar to the quadratic loss function, was introduced by Brier (1950). The Brier score is a probability score that encompasses both calibration and resolution. As previously noted, the latter is an ability of a model in sorting or partitioning uncertain events into mutually exclusive and exhaustive subgroups/bins that have probability measures differing from long-run relative frequencies.

The Brier score (PS) for a single event is

(3.7) $PS(f, d) = (f - d)^2$,

where f is the probabilistic forecast for an event and d is outcome index. If the event occurs, $d = 1$; otherwise, $d = 0$. Over N occasions, the mean of PS is

(3.8) $\overline{PS}(f, d) = 1/N \sum_{i=1}^{N}(f_i - d_i)^2$,

where $i = 1, \ldots, N$ indicates each occasion.

Yates (1988) proposes a covariance decomposition to partition the Brier score into forecast components. This partition is

(3.9) $\overline{PS}(f, d) = Var(d) + MinVar(f) + Scat(f) + Bias^2 - 2Cov(f, d)$.

The variance of the observed outcomes, *Var(d)*, is

(3.10) $Var(d) = \bar{d}(1 - \bar{d})$,

where $\bar{d} = (\frac{1}{N}) \sum_{i=1}^{N} d_i$. Because *Var(d)* captures out-of-model factors affecting forecasts, *Var(d) is* out of a forecaster's control. The remaining components, however, are partially under a forecaster's control. The smaller the Brier score, the better predictive performance. One, therefore, strives to obtain small values for *MinVar(f), Scat(f),* and $Bias^2$ but a large value for *Cov(f,d).*

$Bias^2$ is

(3.11) $Bias^2 = (\bar{f} - \bar{d})^2$,

where $\bar{f} = (\frac{1}{N}) \sum_{i=1}^{N} f_i$. *Bias* is referred to as the mean probability judgment because it indicates the overall miscalibration of the forecasts, i.e. how much the forecast is under- or overestimated. $Bias^2$ indicates the calibration error regardless of the direction (positive or negative) of the error.

The covariance term, *Cov(f,d)*, reveals the ability of a model in distinguishing between individual occurrence whether the event occurs or does not occur. It is defined as

(3.12) $Cov(f, d) = [slope][Var(d)]$,

where $slope = \overline{f}_1 - \overline{f}_0$ and $\overline{f}_k = \left(\frac{1}{N_k}\right) \sum_{j=1}^{N_k} (f_{kj})$ for $k = 0, 1$. $\overline{f}_1$ is the conditional mean of probability forecasts over N_1 occasions that the event actually occurs. In contrast, $\overline{f}_0$ is the conditional mean of probability forecasts over N_0 occasions that the event does not occur.

The scatter term, *Scat(f)*, is

(3.13) $Scat(f) = \left(\frac{1}{N}\right) [N_1 Var(f_1) + N_0 Var(f_0)]$,

where $Var(f_k) = \left(\frac{1}{N_k}\right) \sum_{j=1}^{N_k} (f_{kj} - \overline{f}_k)^2$ for $k = 0, 1$. $Var(f_1)$ is the conditional variance of the probability forecast for an event that actually occurs N_1 times and $Var(f_0)$ is the conditional variance of the probability forecast for an event that does not occur N_0 times. *Scat(f)* is the weighted average of the conditional variances $Var(f_1)$ and $Var(f_0)$. It appears that the scatter captures the conditional dispersion of probability forecasts.

The *MinVar(f)* is the minimum forecast variance of the probability forecast defined as

(3.14) $MinVar(f) = Var(f) - Scat(f)$,

where *Var(f)* is the overall variance of probability forecasts. *MinVar(f)* measures the dispersion of probability forecasts, which cannot be explained by the conditional dispersion. *MinVar(f)* is exactly *Var(f)* when *Scat(f)* = 0.

In this study, the Brier score is calculated for the multiple event case. For a *K*-event (where *K* > *2*) case, the multiple probability score for the *k*th event (Murphy 1970) is

(3.15) $PSM(\mathbf{f}, \mathbf{d}) = (\mathbf{f} - \mathbf{d})'(\mathbf{f} - \mathbf{d})$,

where $\mathbf{f} = (f_1, \ldots, f_k)'$ and $\mathbf{d} = (d_1, \ldots, d_k)'$. f_k and d_k are the probability forecast and the outcome index for an event *k*. Over *N* occurrences, the mean *PSM* is

(3.16) $\overline{PSM} = \sum_{k=1}^{K} \overline{PS}_k$,

where $\overline{PS}_k = 1/N \sum_{i=1}^{N} (f_{ki} - d_{ki})^2$, the probability mean score for the *k*th event.

The covariance decomposition for the multiple event forecast is

(3.17) $$\overline{PSM}(\mathbf{f}, \mathbf{d}) = \sum_{k=1}^{K} Var(d_k) + \sum_{k=1}^{K} MinVar(f_k) + \sum_{k=1}^{K} Scat(f) + \sum_{k=1}^{K} (Bias)^2 - 2 \sum_{k=1}^{K} Cov(f_k, d_k).$$

The interpretation of each term in the multiple event case is similar to that in the single event case.

Rank Probability Score

Unlike the Brier score, the ranked probability score proposed by Epstein (1969) involves using cumulative distribution functions instead of probability density functions. The ranked probability score (RPS) is

(3.18) $RPS = \frac{1}{K-1} \sum_{k=1}^{K} \left(\sum_{j=1}^{k} f_k - \sum_{j=1}^{k} d_k \right)^2 = \frac{1}{K-1} \sum_{k=1}^{K} (F_k - D_k)^2$,

where F_k and D_k are the cumulative distribution of forecasts and outcomes (observations). The *RPS* is equivalent to the Brier score when *K*= 2. Over *N* occasions, the mean *RPS* is

$$(3.19)\ \ \overline{RPS} = \frac{1}{N}\sum_{i=1}^{N} RPS_i.$$

Similar to the Brier score, the lower the *RPS*, the better performance of probability forecast.

The *RPS* assesses how close the distribution is to the observed value (Murphy 1970). The idea of "closer" (distance) does not appear in the Brier score (Epstein 1969). To illustrate, consider two probability forecasts for four categories. Let the two different probability forecasts be: *P* = (0.5, 0.3, 0.1, 0.1) and *P'* = (0.1, 0.3, 0.5, 0.1). Further, assume the observed event occurs in the last category. The Brier score on *P* = $[(0.5-0)^2+(0.3-0)^2+(0.1-0)^2+(0.1-1)^2]=1.16$. Similarly, the Brier score on *P'* = $[(0.1-0)^2+(0.3-0)^2+(0.5-0)^2+(0.1-1)^2]=1.16$. The *RPS* on *P* = $[(0.5-0)^2+(0.8-0)^2+(0.9-0)^2+(1-1)^2]=1.7$. Similarly, the *RPS* on *P'* = $[(0.1-0)^2+(0.4-0)^2+(0.9-0)^2+(1-1)^2]$ =0.98. The Brier scores of the two forecasts are equal; whereas, the *RPS* of the latter is smaller. The *RPS* penalizes forecasts less severely when probabilities are closer to the actual outcomes, and more severely when probabilities are further from the actual outcome (Murphy 1970).

Data

Data used in Chapter II are used for the in-sample data. This data consists of eight natural gas spot prices in Canada and United States: AECO Hub, Alberta, Canada; Chicago City Gate, Illinois; Dominion South Point, Pennsylvania; Henry Hub, Louisiana; Malin, Oregon; Oneok, Oklahoma; Opal, Wyoming; and Waha Hub, Texas. Weekday nominal prices of natural gas from May 3, 1994 to April 30, 2015 are obtained from Bloomberg L.P. (2015). A missing value is replaced by a prior day's price. Each price is the closing price for a specific location for natural gas to be delivered on the next day. All prices are in U.S. dollars per MMBtu (a unit of heat equal to one million British thermal units).

ADF and KPSS tests (in Chapter II) indicate that all prices series in natural logarithms are stationary after first differencing. All price series used in forecasting are first differences of natural logarithms. This implies that returns of natural gas spot markets are forecasted. The in-sample data (May 4, 1994 to October 31, 2014) is augmented with out-of-sample data (November 3, 2014 to April 30, 2015) for forecasting purposes.

Empirical Results

Because of the potential existence of structural breaks around 2000 and 2009 (Chapter II), unrestricted VAR models of the eight series are estimated for three periods with the Schwarz loss criteria used to determine the appropriate number of lags for each period. The full model is fitted over the period of May 4, 1994 to October 31, 2014 with five lags minimizing the Schwarz loss criteria. The second model, two-period model, is fitted over the period of October 2, 2000 to October 31, 2014 with three lags found to be appropriate. The recent model is fitted over the period of January 1, 2010 to October 31, 2014 with three lags minimizing the Schwarz loss measures. As previously discussed, all three models are used to estimate out-of-sample values for one step-ahead horizon covering the period of November 3, 2014 to April 30, 2015 using the Kalman filter to update the model parameters.

Calibration Measures

A graphical representation and a goodness-of-fit test statistic are used to evaluate calibration performance. In figures 3.1-3.8, all calibration plots between the relative frequency and realized fractiles of the forecasts for the eight markets from the three models are close to the 45-degree line, suggesting that the forecasts are well calibrated.

Twenty non-overlapping subintervals of observed fractiles are used to compute chi-squared test statistics. Consistent with the plots, chi-squared goodness-of-fit test statistics on all forecasts are less than the 5% critical value of $\chi^2(19)$, implying that the null hypothesis of well calibration cannot be rejected for any model and market (table 3.1).

Table 3.1. Chi-Squared Goodness-of-Fit Test Statistics on Probability Forecasts of Returns in Eight Natural Gas Spot Markets

Markets	Full[a]	Two-Period[b]	Recent[c]
AECO Hub	21.1890	17.3622	18.7619
Chicago	9.1951	29.4000	19.3968
Dominion South	4.8110	15.2047	17.4094
Henry Hub	13.0476	13.0476	18.7619
Malin	11.1463	16.5397	19.4800
Oneok	25.5984	16.7795	26.0635
Opal	14.3175	18.7619	13.0476
Waha Hub	13.9449	25.9134	13.9449

Note: The null hypothesis of well calibration cannot be rejected if the chi-squared test statistic is less than the 5% critical value of χ^2(19)=30.144.

[a] The full model is initially fitted over the period of May 4, 1994 to October 31, 2014.

[b] The two-period model is initially fitted over the period of October 2, 2000 to October 31, 2014.

[c] The recent model is initially fitted over the period of January 1, 2010 to October 31, 2014.

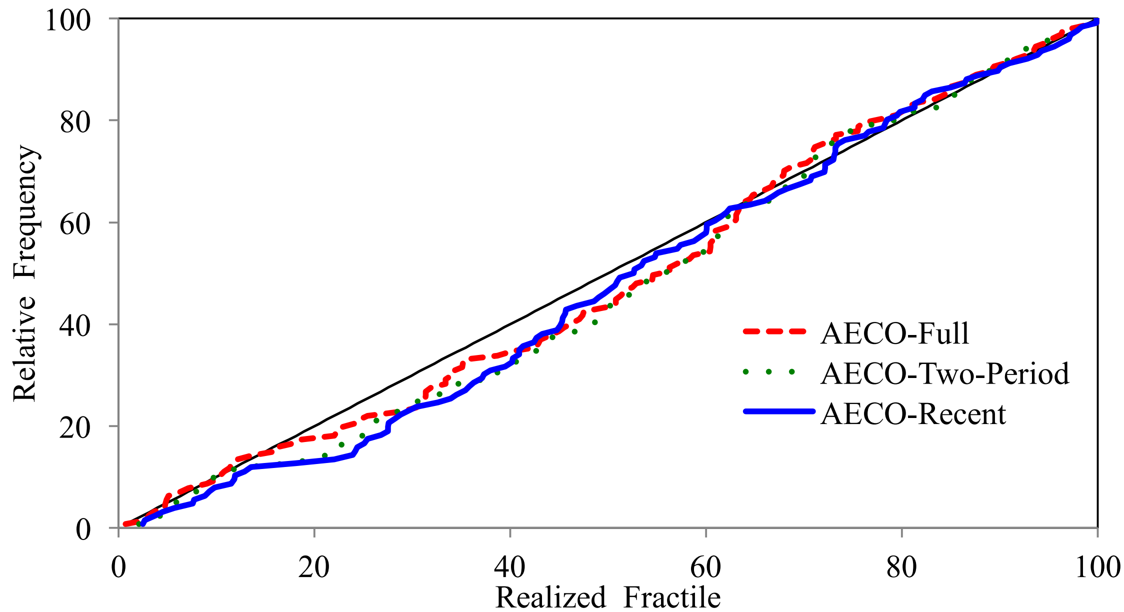

Figure 3.1. Calibration plots for the AECO forecasts from the three models

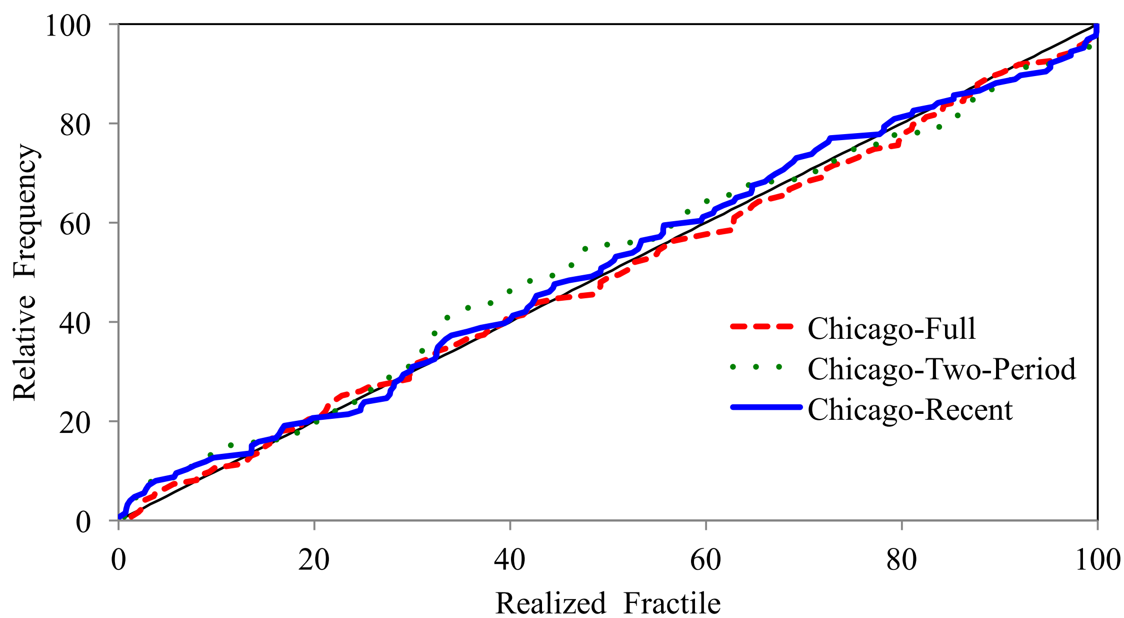

Figure 3.2. Calibration plots for the Chicago forecasts from the three models

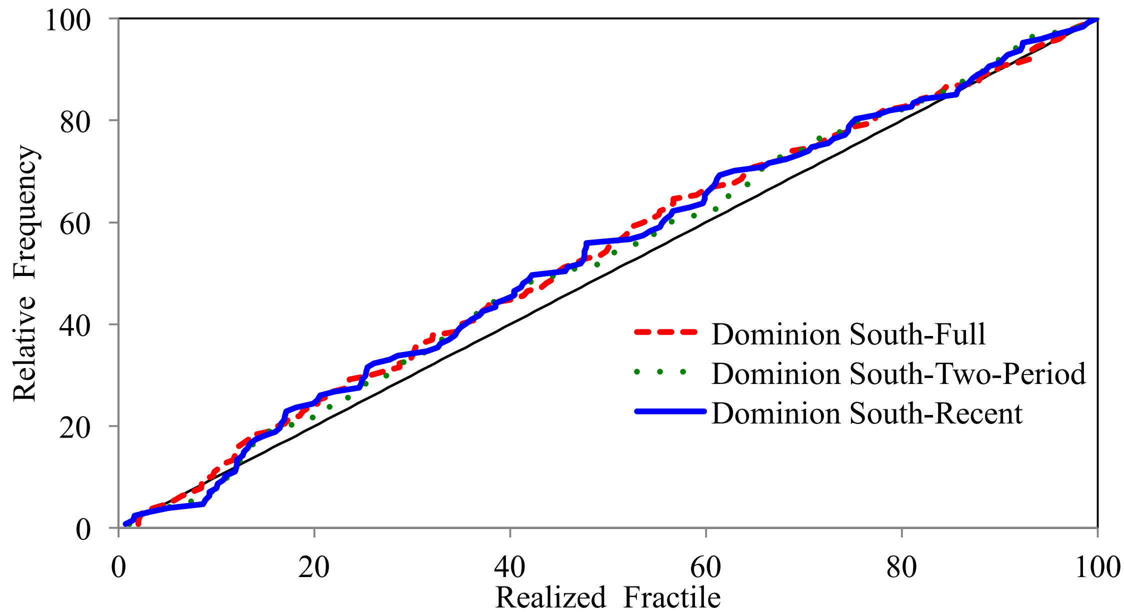

Figure 3.3. Calibration plots for the Dominion South forecasts from the three models

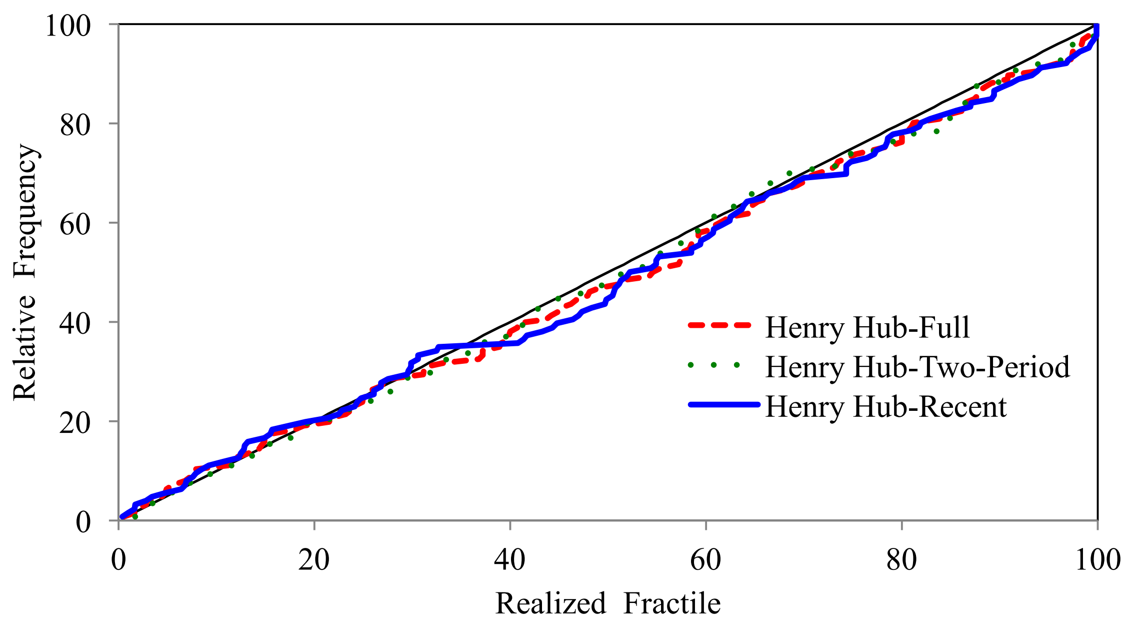

Figure 3.4. Calibration plots for the Henry Hub forecasts from the three models

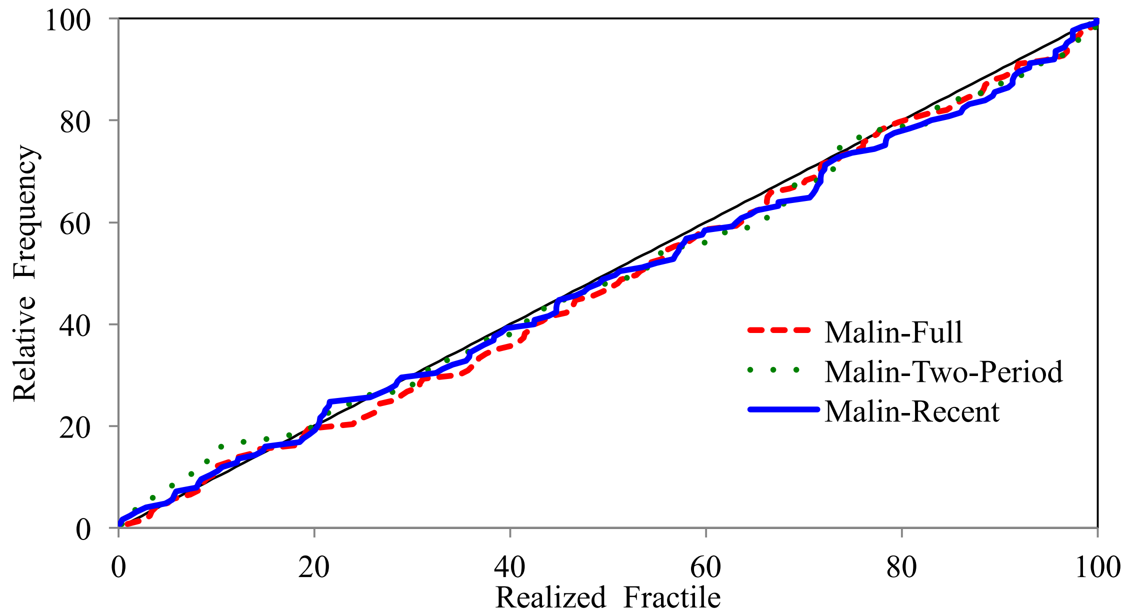

Figure 3.5. Calibration plots for the Malin forecasts from the three models

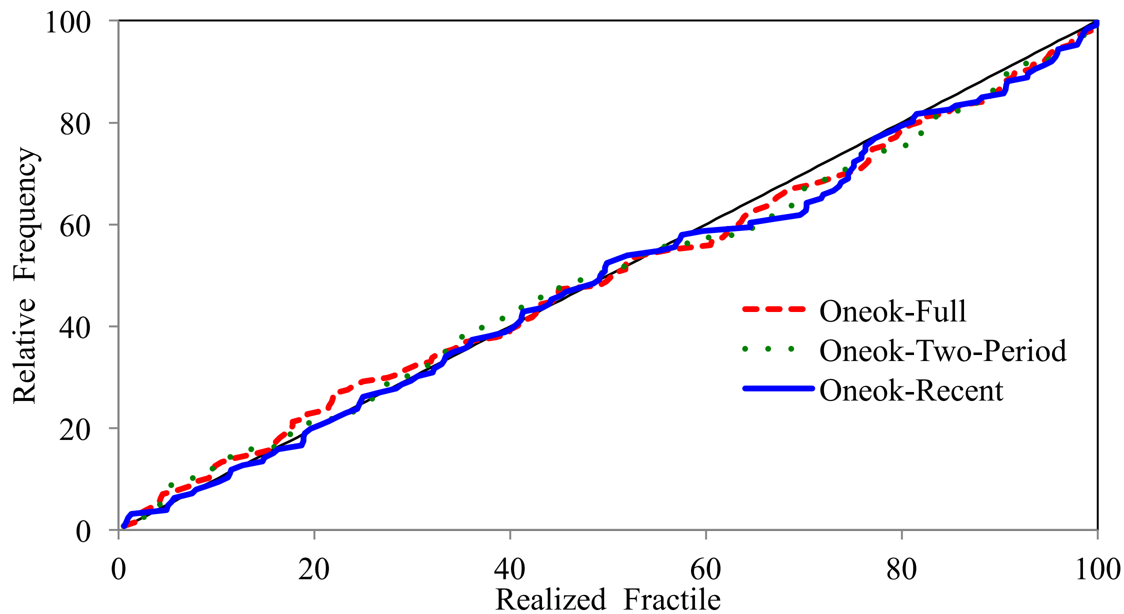

Figure 3.6. Calibration plots for the Oneok forecasts from the three models

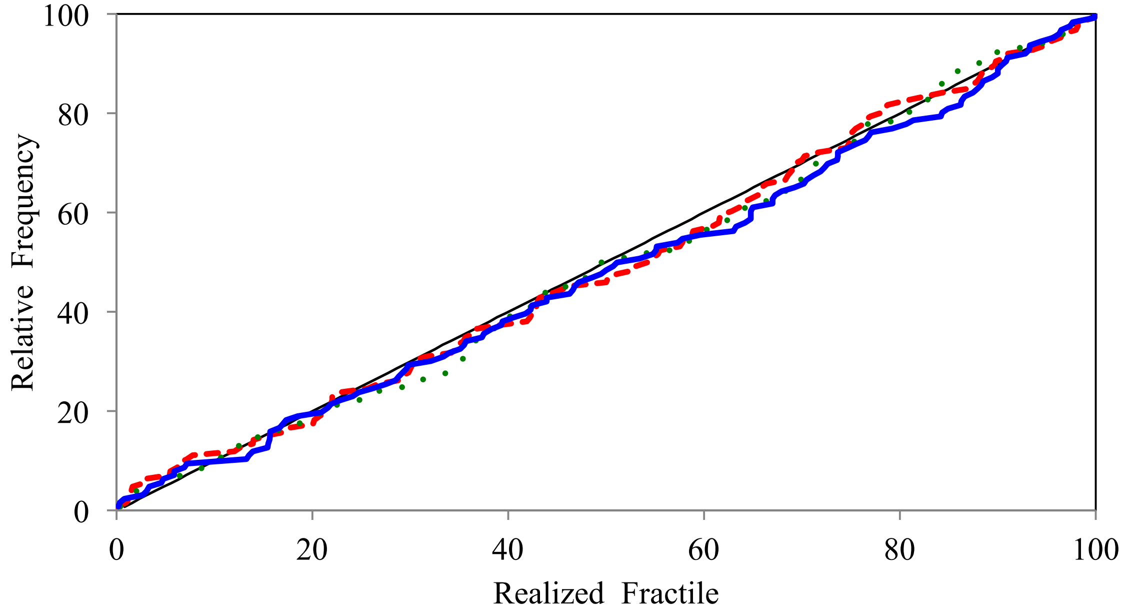

Figure 3.7. Calibration plots for the Opal forecasts from the three models

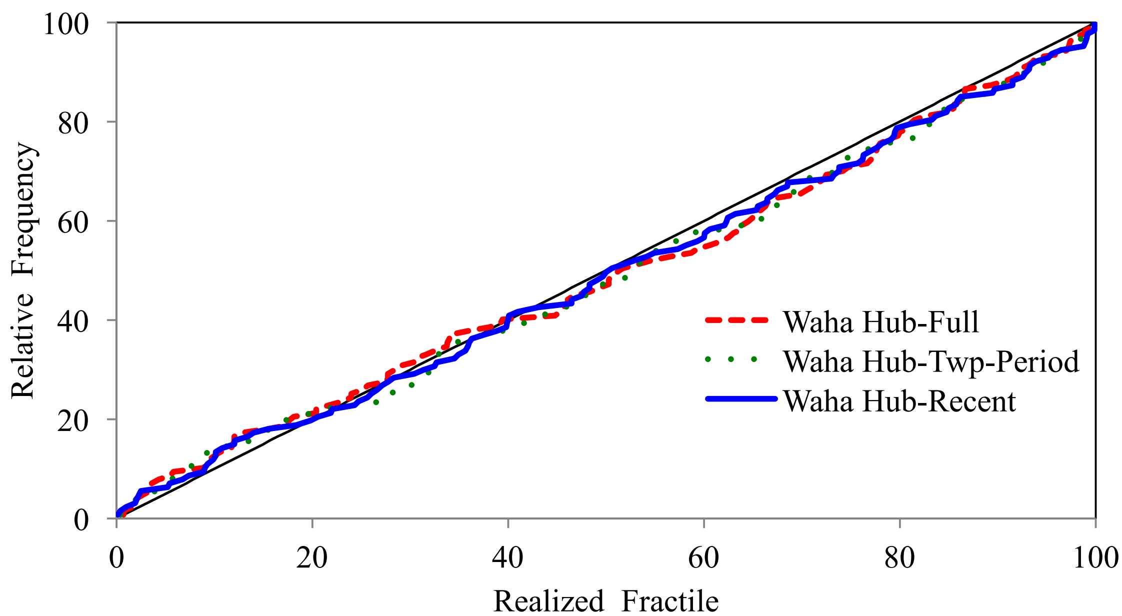

Figure 3.8. Calibration plots for the Waha Hub forecasts from the three models

Table 3.2. Root Mean-Squared Error (RMSE), the Brier Score, and the Ranked Probability Score (RPS) on the Probabilistic Forecast of Returns in Eight Natural Gas Spot Markets

Markets	RMSE	Brier Score	RPS
	The Full Model[a]		
AECO Hub	0.0435	0.8241	0.0997
Chicago	0.1364	0.8587	0.1378
Dominion South	0.1451	0.8881	0.2003
Henry Hub	0.0397	0.8061	0.0978
Malin	0.0494	0.8330	0.1091
Oneok	0.0501	0.8202	0.1091
Opal	0.0609	0.8441	0.1232
Waha Hub	0.0548	0.8368	0.1167
System	0.0828	0.8389	0.1242
	The Two-Period Model[b]		
AECO Hub	0.0435	0.8020	0.0961
Chicago	0.1381	0.8547	0.1435
Dominion South	0.1476	0.8950	0.1982
Henry Hub	0.0407	0.8040	0.0987
Malin	0.0489	0.8344	0.1098
Oneok	0.0496	0.8348	0.1104
Opal	0.0584	0.8381	0.1184
Waha Hub	0.0540	0.8332	0.1164
System	0.0834	0.8370	0.1239
	The Recent Model[c]		
AECO Hub	0.0431	0.8120	0.0983
Chicago	0.1455	0.9245	0.1693
Dominion South	0.1451	0.8901	0.1969
Henry Hub	0.0394	0.8088	0.0976
Malin	0.0506	0.8536	0.1179
Oneok	0.0526	0.8554	0.1208
Opal	0.0582	0.8560	0.1245
Waha Hub	0.0549	0.8471	0.1231
System	0.0847	0.8559	0.1310

[a] The full model is initially fitted over the period of May 4, 1994 to October 31, 2014.
[b] The two-period model is initially fitted over the period of October 2, 2000 to October 31, 2014.
[c] The recent model is initially fitted over the period of January 1, 2010 to October 31, 2014.

Root Mean-Squared Error (RMSE)

To calculate the RMSE for probability forecast, the means of the probability distributions are used as point forecast. Forecasts with lower RMSE are considered better.

Relative to the other markets, the two markets that always have the smallest RMSE are Henry Hub and AECO Hub, with Malin a close third, regardless of the model (table 3.2). The two markets with the largest RMSE are Dominion South and Chicago. The Dominion South forecasts have the largest RMSE in the full and two-period models, while the Chicago has the largest RMSE in the recent model.

Plots of observed returns and means of forecasted values for each market from the three models are illustrated in figure 3.9. Based on the plots, it is difficult to determine which model performs better in forecasting returns in each market. Note, the vertical scales for each panel are different.

The Brier Score

Similar to RMSE, relative to the other markets, the Henry Hub and AECO Hub usually have the smaller Brier scores (table 3.2). The Brier score for the Henry Hub forecast is the smallest in the full and recent models and is the second smallest in the two-period model. The Brier score for the AECO forecast is the smallest in the two-period model, second smallest in the recent model, and the third smallest in the full model. The second smallest of the Brier score in the full model belongs to the Oneok forecast.

Irrespective the model, two markets that usually have the two largest Brier scores are Dominion South and Chicago. Within a given sample, the Brier score for the

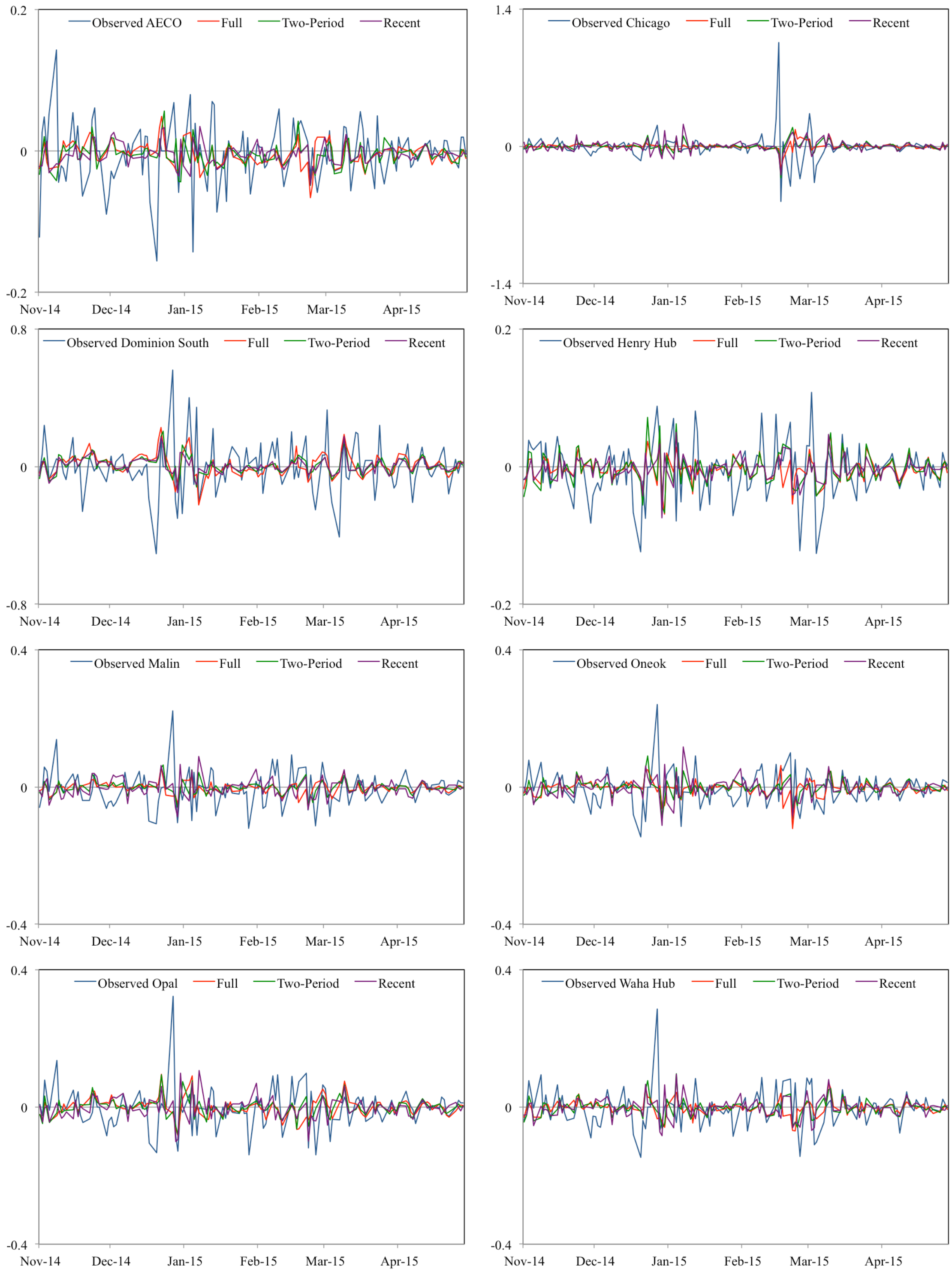

Figure 3.9. Plots of observed natural gas returns and means of forecasts for each market from the three models

Note: vertical scales for each panel are different.

Dominion South forecast is the largest of the eight markets in the full and two-period models and the second largest in the recent model. The Brier score for the Chicago forecast is the largest in the recent model and the second largest in the full and two-period models.

Yates' Covariance Decomposition

Results of Yates' composition are reported in table 3.3. *Var(d)* reflects the underlying variance of observed outcome; *Var(d)* for each market are the same for each of the three models because they are calculated over the same out-of-sample data. Two markets that have the lowest *Var(d)* are Henry Hub and AECO Hub and two markets that have the largest *Var(d)* are Dominion South and Chicago. These size differences in *Var(d)* are one of the reasons why the forecasts of Henry Hub and AECO Hub have smaller Brier scores and the forecasts of Dominion South and Chicago have larger Brier scores.

In the full and recent models, the *MinVar* of the AECO forecasts are smallest, relative to other forecasts; the *MinVar* of the Malin forecast is smallest in the two-period model. Two markets that have largest *MinVar* in the full model are Dominion South and Chicago. The Chicago and Henry Hub forecasts have the largest *MinVar* in the two-period model; the Opal and Chicago forecasts have the largest *MinVar* in the recent model.

Scatter reflects the amount of extra variability over and above the minimum variance and is, sometimes, called the overall noise of the forecasts. As given earlier, scatter is given by *Var(f) – MinVar(f)*. The AECO forecast does not have as much of this additional noise as the other markets because its scatter value is always the smallest

Table 3.3. The Brier Score and Yates' Decomposition on the Probabilistic Forecast of Returns in Eight Natural Gas Spot Markets

Markets	Brier Score	*Var(d)*	*MinVar*	*Scat(f)*	*Bias2*	*Cov(f, d)*
			The Full Model[a]			
AECO Hub	0.8241	0.8001	0.0001	0.0168	0.0115	0.0022
Chicago	0.8587	0.8199	0.0009	0.0594	0.0082	0.0148
Dominion South	0.8881	0.8753	0.0021	0.0485	0.0108	0.0243
Henry Hub	0.8061	0.7909	0.0006	0.0371	0.0041	0.0133
Malin	0.8330	0.8100	0.0002	0.0225	0.0092	0.0045
Oneok	0.8202	0.8075	0.0006	0.0439	0.0025	0.0171
Opal	0.8441	0.8145	0.0004	0.0372	0.0142	0.0111
Waha Hub	0.8368	0.8158	0.0004	0.0405	0.0063	0.0131
System	0.8389	0.8298	0.0016	0.0523	0.0049	0.0249
			The Two-Period Model[b]			
AECO Hub	0.8020	0.8001	0.0005	0.0199	0.0098	0.0142
Chicago	0.8547	0.8199	0.0026	0.0850	0.0169	0.0349
Dominion South	0.8950	0.8753	0.0004	0.0331	0.0067	0.0103
Henry Hub	0.8040	0.7909	0.0013	0.0514	0.0056	0.0226
Malin	0.8344	0.8100	0.0001	0.0291	0.0084	0.0066
Oneok	0.8348	0.8075	0.0002	0.0419	0.0061	0.0104
Opal	0.8381	0.8145	0.0005	0.0311	0.0153	0.0116
Waha Hub	0.8332	0.8158	0.0007	0.0443	0.0102	0.0189
System	0.8370	0.8298	0.0014	0.0540	0.0062	0.0272
			The Recent Model[c]			
AECO Hub	0.8120	0.8001	0.0001	0.0136	0.0092	0.0055
Chicago	0.9245	0.8199	0.0008	0.0724	0.0533	0.0109
Dominion South	0.8901	0.8753	0.0005	0.0312	0.0059	0.0114
Henry Hub	0.8088	0.7909	0.0003	0.0385	0.0019	0.0114
Malin	0.8536	0.8100	0.0003	0.0364	0.0157	0.0044
Oneok	0.8554	0.8075	0.0003	0.0389	0.0172	0.0042
Opal	0.8560	0.8145	0.0010	0.0395	0.0202	0.0096
Waha Hub	0.8471	0.8158	0.0007	0.0450	0.0136	0.0140
System	0.8559	0.8298	0.0011	0.0526	0.0105	0.0190

Note: Brier Score = $Var(d) + MinVar(f) + Scat(f) + Bias^2 - 2Cov(f, d)$.

[a] The full model is initially fitted over the period of May 4, 1994 to October 31, 2014.

[b] The two-period model is initially fitted over the period of October 2, 2000 to October 31, 2014.

[c] The recent model is initially fitted over the period of January 1, 2010 to October 31, 2014.

regardless of the model. Events that eventually occur have the same forecasted variance as events that eventually do not occur, suggesting that the AECO is not adjusting the variance of its forecasts in anticipation of occurrences and non-occurrences. The Chicago forecast, however, has much additional noise because its scatter value is always the largest regardless of the model. The forecast on Chicago does show this differential adjustment of variance on events that occur versus events that do not occur.

Without considering direction, the $Bias^2$ is a measure of miscalibration. In the full model, two markets that have the smallest $Bias^2$ are Oneok and Henry Hub; two markets that have the largest $Bias^2$ are Opal and AECO Hub. In the two-period and recent models, the $Bias^2$ of the Henry Hub forecast is the smallest, while the $Bias^2$ of the Chicago and Opal forecasts are the largest.

The covariance term of the partition is the essence of the forecasting exercise (Yates 1988; Casillas-Olvera and Bessler 2006). Larger covariances are associated with better forecasts. Among the eight markets, Dominion South, Chicago, and Waha Hub are the markets that have the largest covariance between the forecasts and the observed outcomes in the full, two-period, and recent models.

In the full model, even though the Dominion South forecast has the largest covariance, its Brier score is largest because it has the largest *Var(d)* and the largest *MinVar*. Similarly, the Chicago forecast in the recent model has the largest covariance but still has a large Brier score as the Chicago's *Var(d)* is large, as well as the *MinVar,* scatter, and the $Bias^2$. In the recent model, although the covariance on the Waha Hub forecast is the highest, its Brier score is not the smallest as the Waha Hub forecast has quite large *Var(d)*, *MinVar,* and scatter.

Focusing on the overall Brier score is a bit misleading. Of course, *Var(d)* is not under control of the modeler, but the random variable being forecasted (or the partition of outcomes into bins) is. Stated alternatively, given the bin width selection, the covariance between the forecasts and the observed outcomes is a helpful guide in indicating which markets the model is doing a better job of forecasting and which markets the model is doing less well in forecasting. Markets having a high covariance, and thus which the full, two-period, and recent models are forecasting better, are Dominion South, Chicago, and Waha Hub. The Malin forecast has a consistently low covariance term, indicating the three models are not discriminating well between events which occur and those that do not occur, ex ant.

The Ranked Probability Score (RPS)

In addition to the location concept, the (closer) distance concept is taken into account in the *RPS.* The closer the forecast is to the actual outcome, the lower the *RPS.* The lower the *RPS*, the better the probability forecast.

The two market forecasts that always have small *RPS* are Henry Hub and AECO Hub, regardless of the model (table 3.3). In the full and recent models, the *RPS* on the Henry Hub forecast is the smallest; in the two-period model, the *RPS* on the AECO Hub is the smallest. Dominion South and Chicago forecasts usually have the largest *RPS* regardless the model.

Comparison of Each Series across Models

To evaluate the predictive performance of each model, the RMSE, the Brier score and its Yates' partitions, and the RPS on each forecast are compared across the three models. AECO, Dominion South, Henry Hub, and Opal forecasts have the smallest RMSE in the recent model, while Chicago, Malin, Oneok, and Waha Hub forecasts have the largest RMSE in the recent model.

Compared with the other two models, the recent model seems to provide poorer forecasts, as the forecasts of all markets except AECO and Henry Hub have the largest Brier score in the recent model. Similarly, all forecasts except AECO, Dominion South, and Henry Hub have the largest RPS in the recent model.

The variance of the observed outcomes (*Var(d)*) on each series is the same for the three models because it is evaluated over the same out-of-sample data. Compared across the three models, the AECO, Opal, and Waha Hub forecasts have the minimum *MinVar* in the full model; the Dominion South, Malin, and Oneok have the minimum *MinVar* in the two-period model; and the Chicago and Henry Hub forecasts have the minimum *MinVar* in the recent model. Among the three models, the minimum scatter of Chicago, Henry Hub, Malin, and Waha Hub is found in the full model. The minimum scatter of the Opal forecast is found in the two-period model; the minimum scatter of AECO Hub, Dominion South, and Oneok forecasts is found in the recent model. Five of eight forecasts maximum $Bias^2$ occurs in the recent model; the markets are Chicago, Malin, Oneok, Opal, and Waha Hub. Similarly, five forecasts minimum covariance occurs in the recent model; the markets are Chicago, Henry Hub, Malin, Onoke, and Opal.

Comparison of the Three Systems

RMSE, the Brier score and its Yates' partitions, and the RPS of the system of eight series are estimated for the three models (table 3.2). RMSE of each system is the square root of the average MSE of all forecasts or variables in the system. The Brier score and its partitions of each system are calculated by treating all variables in the system as a single variable. The RPS on each system is an average RPS for all forecasts in the system.

RMSE for the system is minimized using the full sample, whereas, the Brier Score and RPS are minimized for the system using the two-period sample. The recent sample always produces the largest RMSE, Brier Score, and RPS for the system between models. This result is surprising given the potential structural breaks found in Chapter II.

The full system has the greatest minimum variance of forecasting, but the smallest scatter and the smallest Bias^2. The ordering of Bias^2 of the three systems is consistent with the ordering of RMSE. The maximum covariance term on the two-period system is possibly behind the minimum Brier score, although the two-period system has the largest scatter compared to the other two systems. The recent system has the smallest minimum variance, but with the highest Bias^2 and the smallest covariance, the Brier score of the recent system is the largest.

Encompassing tests (Harvey and Newbold 2000) on probability forecasts from the three systems are performed. In encompassing regression, the forecast error (the difference between observed return and mean of forecasted value) from each system is regressed on the difference between itself and the other systems' forecast errors[16]. If one encompasses the others, it means that the others contain no useful information not present in the encompassing forecast. Encompassing results suggest that the full system encompasses the two-period and recent systems; the two-period system encompasses the recent system; and the recent system encompasses the two-period system. Based on these results, it appears that the full system is the best, and the two-period and recent systems are not statistically different.

Discussion

Regardless of the model (or the data used), Henry Hub and AECO Hub are either the first or second easiest market to forecast; whereas, Dominion South and Chicago are either the hardest or second hardest market to forecast in terms of RMSE and scoring rules. Several different aspects may help in explain these results.

First, Henry Hub may be easier to predict because it is the important market for pricing of the North American natural gas spot and futures markets (Serletis and Rangel-Ruiz 2004; U.S. EIA 2014f), while AECO Hub may be simpler to predict because it may not play a significant role in price discovery (Working Group of Commercial Energy Firms 2009). Along these lines, Olsen, Mjelde, and Bessler (2014) find that AECO, Alberta, is less important for price discovery than other Canadian markets. Results from

[16] See Harvey and Newbold (2000) and Bessler and Wang (2012) for more details on encompassing tests.

the study of price dynamics in Chapter II suggest AECO is exogenous in contemporaneous time and AECO Hub provides less information to the other markets in the system.

Second, there may exist the production and consumption differences in terms of forecasting. Henry Hub and AECO Hub, which are located in the production zones, are easier to forecast than Chicago, Illinois, which has fewer producing natural gas wells in the area and Illinois is one of the top natural gas consuming states in the U.S. (U.S. EIA 2015f). Production is easier to control relative to consumption, as the latter is crucially dependent on a large unknown in weather. Third, the difficulty in predicting Dominion South returns may be because of the alteration of the market's role. Depending on interstate pipelines to supply natural gas, Dominion South, Pennsylvania had been recognized as an excess demand zone until 2009 (U.S. EIA 2015g). Because of the development of the Marcellus shale, natural gas production in Pennsylvania has increased considerably since 2010; this area has become the second largest U.S. natural gas producing area (U.S. EIA 2015g). The Northeast region previously known as the excess demand area has become an excess supply area; pipelines are being reformed to transport natural gas from the Marcellus area to the Midwest and the Gulf Coast (U.S. EIA 2015g). This alteration may also possibly affect Chicago.

In term of the Brier score's partition, the covariance between the forecasts and the observed outcomes is the best indicator of forecasting ability, given bin/subgroup assignment; Dominion South, Chicago, and Waha Hub returns appear to be easier to predict when using the full, two-period, and recent models, respectively. Returns in Malin appear to be more difficult to predict, irrespective the model.

The objective of this study is to determine whether and how the presence of structural breaks affects performance of out-of-sample probability forecasting. The difference among the three models is the in-sample data used. Based on the findings in Chapter II, there appears to be two possible structural shifts (around 2000 and 2009) during the period of the full sample, one possible structural break (around 2009) during the period of data used in the two-period model, and no structural changes occurring during the period of the recent sample. Different in-sample data yields different probability forecasts. With the minimum probability scores and the maximum covariance term from Yates' decomposition, it appears that the two-period model is more preferable to the other two models. With no structural shifts during the period of data used, it was expected that the recent model would yield the best probability forecasts among the three models; both the RMSE and the scoring rules do not suggest this is true. This may be partially because of the time-varying parameters associated with the use of the Kalman filter in the procedure to estimate the probability forecasts. Additionally, based on the RMSE and scoring rules on the system of eight series, it appears that number of observations somehow matters in forecasting performance; as they suggest that the recent model in which the smallest data set is consider has poorer performance relative to the other two models. The RMSE is smallest in the full model, which has the largest number of in-sample observations, while both the Brier score and the RPS suggest the two-period model in which the number of observations is between the full and recent models. This issue is left for the further research.

CHAPTER IV

EFFECTS OF THE STRUCTURAL CHANGE ON TRANSACTION COSTS BETWEEN NORTH AMERICA NATURAL GAS SPOT MARKETS

The cointegration model, introduced by Granger (1981), has been employed to capture long-run equilibrium relationships among non-stationary economic variables. The idea of cointegration is that two or more non-stationary (unit-root) economic variables have a propensity to move toward equilibrium in the long run; an error correction model (ECM) (Granger 1981; Engle and Granger 1987) can explain this movement. Cointegration and the ECM implicitly assume that such movements occur every period (Balke and Fomby 1997). Concerned that fixed costs may prevent continuous correction toward the long-run equilibrium, Balke and Fomby (1997) introduce threshold cointegration models. In these models, two series are cointegrated when the series are far (outside the threshold) from the equilibrium, but are not cointegrated when they are close to the equilibrium (within the threshold). Balke and Fomby (1997) employ a threshold autoregressive (TAR) model to describe this nonlinear adjustment process.

Departures from equilibrium may be because of the presence of transaction costs; transaction costs including transportation costs and/or arbitrage costs may induce price differences between two markets in which returns are free to diverge and in which an arbitrage opportunity exists (Balke and Fomby 1997). As long as the price difference is greater than transaction costs, traders profit from purchasing a commodity in the lower priced market and selling it in the higher priced market; trade continues until the price gap is equal to the transaction costs. This phenomenon is known as the law of one price; after considering transaction costs, prices of a given commodity in two markets converge to a single price (Ardeni 1989; Yang, Bessler, and Leatham 2000). The existence of threshold cointegration implies that when the price difference is within the threshold bands, arbitrage opportunities do not exist; the prices in two markets are not cointegrated. In contrast, when the price difference is outside the threshold bands, arbitrage will drive the price disparity towards the threshold bands (transaction costs); the prices in two markets are cointegrated.

Under the law of one price, threshold cointegration models involve two components; one is the difference (or interval) between the upper and the lower threshold values at a point time and another is the average of the upper and the lower threshold values at a point time (Park, Mjelde, and Bessler 2007). In most previous threshold cointegration studies (Tsay 1998; Goodwin and Piggott 2001; Lo and Zivot 2001; O'Connell and Wei 2002), both components are constant over time (figure 4.1b). It is straightforward to show that the threshold cointegration model is the traditional cointegration if the difference of threshold values equals zero (figure 4.1a). It, however, may not be realistic to assume invariant threshold values (Park, Mjelde, and Bessler 2007).

Threshold cointegration models can be developed into three additional alternative scenarios (Park, Mjelde, and Bessler 2007). First, the intervals of the upper and lower threshold bands are allowed to vary over time, but the average of the threshold values is

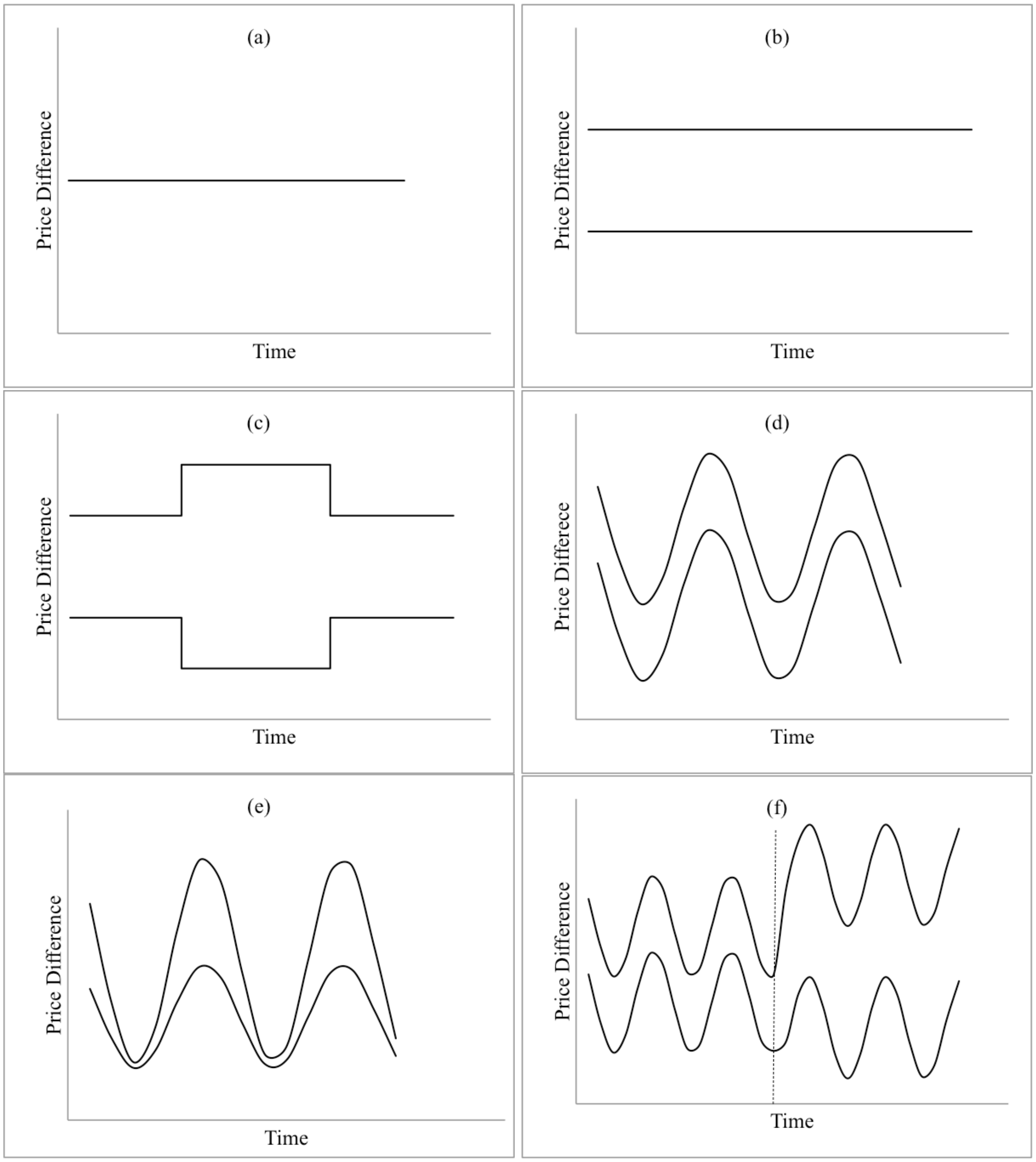

Figure 4.1. Threshold cointegration models under the law of one price in diverse scenarios

Note: Figures in panels (b) to (e) are adopted from Park, Mjelde, and Bessler (2007).

fixed (figure 4.1c). Second, the interval is fixed, but the averages are variable over time (figure 4.1d). Third, the two components are both time-dependent (figure 4.1e).

Park, Mjelde, and Bessler (2007) model threshold cointegration for the scenario given in figure 4.1d using seasonality to generate time-varying threshold bands. Bekkerman, Goodwin, and Piggott (2013) propose the scenario given in figure 4.1e, in which time-dependent, conditional threshold bands are estimated to investigate market linkages. Nonetheless, their model may be mis-specified, as the model is not consistent with either Balke and Fomby (1997) or Lo and Zivot (2001). Market linkages, thereby, may not exist. In this study, given the potential presence of structural breaks defined in Chapter II, the data are divided into two subsamples. Threshold cointegration for each subsample is estimated and time-varying threshold values are obtained using daily degree-days; results are similar to scenario in figure 4.1f.

The objective of this study is to examine the presence of threshold cointegration between market pairs before and after the potential break associated with the shale gas revolution in the long-term pricing relationship among North America natural gas spot markets presented in Chapter II. Differences in transaction costs before and after the potential structure change are analyzed.

Literature on Threshold Cointegration and the Law of One Price

The law of one price notes that transaction costs (including transportation costs) can influence arbitrage in spatially separated markets. As previously noted, this influence can be explained by the existence of threshold cointegration (Tsay 1998; Goodwin and

Piggott 2001; Lo and Zivot 2001; O'Connell and Wei 2002; Park, Mjelde, and Bessler 2007; Bekkerman, Goodwin, and Piggott 2013).

The existence of threshold cointegration is found in goods that are tradable and relatively homogenous (Lo and Zivot 2001). Spatial market linkages with non-linear adjustment have been found in both commodity and financial markets (Goodwin and Piggott 2001; Tsay 1998; Park, Mjelde, and Bessler 2007). With constant threshold values, such non-linear adjustment can be explained by transaction costs (Goodwin and Piggott 2001; Tsay 1998). In some markets including financial markets, estimated threshold values are not solely determined by transaction costs but also interest rates, economic risks, and financial purpose of a trade, as it is impossible to identify if a trade is strictly for arbitrage purposes (Tsay 1998). Even though the threshold remains fixed, time-varying threshold bands can be estimated to capture the effect of seasonality (Park, Mjelde, and Bessler 2007).

Methodology

The threshold cointegration model introduced by Balke and Fomby (1997) is a combination of cointegration and non-linearity (Hansen and Seo 2002). Two non-stationary series are cointegrated when a linear combination of the series is stationary (Granger 1981). The essence of cointegration is that there exists a long run equilibrium relationship that causes the series to have a tendency to move together in the long run.

The idea of cointegration is illustrated by

(4.1) $x_{1t} + \alpha x_{2t} = z_t,$

where x_{1t} and x_{2t} are two nonstationary time series and α is a parameter (Balke and Fomby 1997). If these two series are cointegrated, equation (4.1) represents the equilibrium relationship between x_{1t} and x_{2t}, where z_t is the deviation from the equilibrium; the cointegrating vector is given by $(1, \alpha)$ (Balke and Fomby 1997). Engle and Granger's (1987) requirements for x_{1t} and x_{2t} to be cointegrated are that the deviation, z_t, is stationary and follows a linear autoregressive model

(4.2) $z_t = \rho z_{t-d} + \varepsilon_t,$

where ρ is a parameter and ε_t is a random variable with zero mean and constant variance.

In the threshold cointegration model, the long-run relationship is inactive inside an interval but becomes active once the deviations are outside the interval. To describe such a nonlinear adjustment process, Balke and Fomby (1997) assume that the deviation, z_t, follows a threshold autoregressive (TAR) model in which ρ depends on the past realization of z_t. In particular,

(4.3) $$\begin{aligned} \rho &= 1 && if |z_{t-d}| \leq \theta \\ &= \rho^*,\ with\ |\rho^*| < 1 && if |z_{t-d}| > \theta, \end{aligned}$$

where d is a positive integer, indicating the delay parameter in the error correction process, and θ is a threshold value (Balke and Fomby 1997). Deviations from equilibrium are described by

(4.4) $$\begin{aligned} z_t &= z_{t-d} + \varepsilon_t && if\ |z_{t-d}| \leq \theta \\ &= (1-\rho^*)\theta + \rho^* z_{t-d} + \varepsilon_t && if\ |z_{t-d}| > \theta. \end{aligned}$$

z_t is a random walk when $|z_{t-d}| \leq \theta$; z_t is stationary when $|z_{t-d}| > \theta$ (Balke and Fomby 1997). That is, x_{1t} and x_{2t} are not cointegrated if z_{t-d} is in the interval $[-\theta, \theta]$; x_{1t} and x_{2t} are cointegrated if z_{t-d} is outside the interval (Balke and Fomby 1997).

Exploiting the full structure of the model, multivariate techniques of testing threshold cointegration have higher power than the univariate techniques because the univariate techniques neglect the restrictions imposed by the multivariate structure (Lo and Zivot 2001). The multivariate threshold cointegration model can be characterized by the threshold vector autoregressive (TVAR) model (Lo and Zivot 2001)

(4.5) $$X_t = \mu^{(j)} + \Phi_1^{(j)} X_{t-1} + \Phi_2^{(j)} X_{t-2} + \cdots + \Phi_k^{(j)} X_{t-k} + \epsilon_t^{(j)}$$

$$if\ c^{j-1} \leq z_{t-d} \leq c^j.$$

Equation (4.5) represents a general J-regime bivariate TVAR model for X_t, where:

X_t is a vector of two series, $X_t = (x_{1t}, x_{2t})'$;

t = 1, …, T indicates time;

j = 1, …, J indicate regimes;

k indicates lag length;

z_{t-d} represents a threshold variable where d (<k) is a positive integer indicating a delay parameter in the adjustment process;

c represents threshold values, where $-\infty = c^{(0)} < c^{(1)} < c^{(2)} \ldots < c^{(J)} = \infty$; and

$\epsilon_t^{(j)}$ is a vector of residuals with mean zero and variance-covariance matrix $\Sigma^{(j)}$ which are assumed to be serially uncorrelated (Lo and Zivot 2001).

The TVAR model can be rearranged as

(4.6) $$\Delta X_t = \mu^{(j)} + \Pi^{(j)} X_{t-1} + \sum_{i=1}^{k-1} \Psi_i^{(j)} \Delta X_{t-i} + \epsilon_t^{(j)} \qquad if\ c^{j-1} \leq z_{t-d} \leq c^j,$$

where $\Pi^{(j)} = \sum_{i=1}^{k} \Phi_i^{(j)} - I_2$ and $\Psi_i^{(j)} = -\sum_{l=i+1}^{k} \Phi_l^{(j)}$ (Lo and Zivot 2001). Within each regime, if X_t is I(1) and x_{1t} and x_{2t} are cointegrated with a cointegrating vector $\beta' = (1, -\beta_2)$, then the rank of $\Pi^{(j)} = 1$ and $\Pi^{(j)} = \gamma^{(j)}\beta' = \begin{pmatrix} \gamma_1^{(j)} \\ \gamma_2^{(j)} \end{pmatrix}(1, \quad -\beta_2)$.

The threshold vector error-correction model (TVECM) can be expressed as

$$(4.7) \quad \Delta X_t = \mu^{(j)} + \gamma^{(j)}\beta' X_{t-1} + \sum_{i=1}^{k-1} \Psi_i^{(j)} \Delta X_{t-i} + \epsilon_t^{(j)} \qquad if\ c^{j-1} \leq z_{t-d} \leq c^j,$$

where ΔX_t is the first difference of X_t, $(X_t - X_{t-1})$, and $\mu^{(j)}$ is a vector of constant terms. $\gamma^{(j)}\beta'$ is a matrix of coefficients of lagged levels and $\Psi_i^{(j)}$ is a matrix of coefficients (Park, Mjelde, and Bessler 2007). Superscript *j* indicates a regime-specific. $\beta' X_{t-1}$ represents a nonlinear error correcting process (Lo and Zivot 2001). The cointegrating vector β is assumed to be identical in all regimes; this assumption, however, is not restrictive (Lo and Zivot 2001).

To obtain time-varying threshold values, Park, Mjelde, and Bessler (2007) first estimate constant threshold values and then modify the time-invariant threshold values for the effect of seasonality using cooling and heating degree-days. Even though threshold values are variable, the difference between the upper and lower threshold values (transaction costs) remain constant overtime. In Bekkerman, Goodwin, and Piggott (2013), time-dependent transaction costs are estimated as functions of fuel costs and seasonality. It, however, is likely that their model is mis-specified. Let z_t be the difference between two series markets, $z_t = x_{1t} - x_{2t}$, in the TAR model, Bekkerman, Goodwin, and Piggott (2013) designate differentials of the series difference at time t and t-1, $\Delta z_t$and Δz_{t-1}, as dependent and independent variables. Unfortunately, this contradicts the TAR models, presented by Balke and Fomby (1997) and Lo and Zivot (2001). Balke and Fomby (1997) assign the differences of two series at time t and t-1, z_t and z_{t-1}, as dependent and independent variables. In Lo and Zivot's (2001), the differential of the difference at time t, Δz_t, is on the left hand size of the equal sign; whereas, the difference of two series at time t-1, z_{t-1}, is on the right hand size. Using the differential of the differences, Δz_{t-1}, which is a I(0) process, does not guarantee the difference, z_{t-1}, is stationary. Consequently, the two series may not be cointegrated.[17]

Estimation

Testing for threshold cointegration involves two steps[18] (Balke and Fomby 1997; Park, Mjelde, and Bessler 2007). The first step is to test whether cointegration exists. If cointegration is found, one proceeds to the second step to test whether the transition of the cointegrating relationship is linear or nonlinear.

Testing for Cointegration

In line with Park, Mjelde, and Bessler (2007, 2008), cointegrating rank and lag length are determined simultaneously using Schwarz loss measure (the formula is given in Chapter II). This method provides better large sample results in Monte Carlo simulations than the trace test, which determines the cointegrating rank given the lag order (Wang and Bessler 2005).

[17] See Lo and Zivot (2001) for more details.

[18] Balke and Fomby also propose two steps of the threshold cointegration test. The difference between tests of Balke and Fomby (1997) and Park, Mjelde, and Bessler (2007) is that the first step of Balke and Fomby's (1997) test is under the univariate setting while Park, Mjelde, and Bessler's (2007) is under the multivariate setting.

Testing for Nonlinearity

The nonlinearity test proposed by Balke and Fomby (1997) is employed to determine whether the univariate cointegrating residual (the threshold variable) is linear by testing for structural breaks in a rearranged autoregressive model. In the rearranged model, the data is ordered based on the value of the threshold variable instead of time. Data reordering does not change the dynamics of the cointegrating relationship but is beneficial for identifying nonlinearity as the presence of a threshold in the time-ordered data translates into a structural change in the rearranged data (Lo and Zivot 2001).

Based on the supremum-Wald statistic, Hansen (1996, 1999) provides a method for testing nonlinearity under the null hypothesis of a TAR model with one regime against the alternative of a TAR model with m regimes, where m is a positive integer. Lo and Zivot (2001) extend Hansen's method for testing nonlinearity in univariate TAR models to test nonlinearity in a multivariate TVECM. Under the null hypothesis of a linear VECM against the alternative of a TVECM(m) for m > 1, the supremum-Likelihood Ratio (sup-LR) statistic, which is equivalent to the sup-Wald, is used. The sup-LR statistic is

(4.8) $$sup - LR = T\left\{\ln\left(\det\left(\hat{\Sigma}\right)\right) - \ln\left(\det\left(\hat{\Sigma}_m\left(\hat{c}^{(j)}, \hat{d}\right)\right)\right)\right\},$$

where $\hat{\Sigma}$, and $\hat{\Sigma}_m\left(\hat{c}^{(j)}, \hat{d}\right)$ are the variance-covariance matrices of the estimated residual from the linear VECM and m-regime TVECM, $\hat{c}^{(j)}$ are the estimated threshold values, $\hat{d}$ is the estimated delay parameter, and det is the matrix determinant operator (Lo and Zivot 2001). Under the null hypothesis of linear cointegration, $c^{(j)}$ are unknown and unidentified; the bootstrap procedure proposed by Hansen (1999) and modified by Lo and Zivot (2001) is used to compute *p*-values for the test.

Estimating Three-Regime TVECM

Similar to Tsay (1998), Goodwin and Piggott (2001), Lo and Zivot (2001), Park, Mjelde, and Bessler (2007), and Bekkerman, Goodwin, and Piggott (2013), this study examines the existence of three-regime threshold cointegration. An unrestricted bivariate three regimes TVECM (equation (4.7)) is

$$(4.9)\quad \Delta X_t = \begin{cases} \eta_1' F_{t-1} + \epsilon_t^{(1)} & if -\infty = c^{(0)} \leq z_{t-d} \leq c^{(1)} \\ \eta_2' F_{t-1} + \epsilon_t^{(2)} & if\ \ c^{(1)} \leq z_{t-d} \leq c^{(2)} \\ \eta_3' F_{t-1} + \epsilon_t^{(3)} & if\ \ c^{(2)} \leq z_{t-d} \leq c^{(3)} = \ \infty, \end{cases}$$

where $F_{t-1} = (1, z_{t-d}, \Delta X_{t-1}, \dots, \Delta X_{t-k+1})'$, η_j' is a matrix of coefficients, and $z_{t-d} = \beta' X_{t-1}$ is the threshold variable classifying observations into three regimes (Lo and Zivot 2001). The cointegrating vector β' is assumed to be a known vector of $(1, -1)'$ and is common for all regimes; these assumptions are applicable under the law of one price (Balke and Fomby 1997; Lo and Zivot 2001). The variance of the error term in each regime is assumed to be identical such that var($\epsilon_t^{(1)}$) = var($\epsilon_t^{(2)}$) = var($\epsilon_t^{(3)}$) (Enders 2004).

To estimate the multivariate TVECM, sequential conditional least squares are performed. Equation (4.9) is expressed as

$$(4.10)\quad \Delta X_t = \eta_1' F_{t-1} I_t^{(1)}(c,d) + \eta_2' F_{t-1} I_t^{(2)}(c,d) + \eta_3' F_{t-1} I_t^{(3)}(c,d) + \epsilon_t,$$

where $I_t^{(j)}(c,d) = I_t^{(j)}\left(c^{(j-1)} \leq z_{t-d} \leq c^{(j)}\right)$ denotes an indicator function, taking on the value of 1 if $c^{(j-1)} \leq z_{t-d} \leq c^{(j)}$ and 0 otherwise (Lo and Zivot 2001). When the threshold values ($c^{(1)}$ and $c^{(2)}$) are known, equation (4.10) is a multivariate regression model with dummy variables (Lo and Zivot 2001). In general $c^{(1)}$ and $c^{(2)}$ are unknown. $c^{(1)}$ and $c^{(2)}$ are estimated along with the other parameters under the assumption that $c^{(1)}$ and $c^{(2)}$ are between the minimum and maximum values of the data series (Enders and Chumrusphonlert 2004). To constrain the threshold values, at least 10% of data are required to be contained to be in each regime; initial candidates for $c^{(1)}$ and $c^{(2)}$ are selected from samples such that the initial middle interval contains 80% of data (Hansen 1999).

The sequential conditional least squares regression involves two steps (Lo and Zivot 2001). In the first step, potential candidates for the threshold values and delay parameter $(c^{(1)}, c^{(2)}, d)$ are selected as starting values to estimate $(\eta_1', \eta_2', \eta_3')$ by multivariate least squares. In this study, the delay parameter is assumed to be one, which is consistent with Balke and Fomby (1997), Lo and Zivot (2001), and Park, Mjelde, and Bessler (2008). The estimation in the first step yields the residual sum of squares, $RSS_3(c^{(1)}, c^{(2)}, 1)$ for all possible combination of $(c^{(1)}, c^{(2)}, 1)$. In the second step, a three-dimensional grid search is used to find the threshold values that minimize the residual sum of square, $RSS_3\left(c^{(1)^*}, c^{(2)^*}, 1\right)$. $c^{(1)^*}$ and $c^{(2)^*}$ are applied to reestimate the parameters $(\eta_1', \eta_2', \eta_3')$ of the TVECM. Because of computational issues associated with the three-dimensional grid search method, Hansen (1999) suggest using the sequential estimation of multiple break points proposed by Bai (1997) to estimate the three-regime TVECM.

Obtaining Time-Varying Threshold Values

Park, Mjelde, and Bessler's (2007) procedure to obtain time-varying threshold values by using U.S. aggregate cooling and heating degree-days (CDD and HDD) is employed. "Heating degree-days are summations of negative differences between the mean daily temperature and the 65 degrees' Fahrenheit base; cooling degree days are summations of positive differences from the same base" (National Oceanic and Atmospheric Administration 2014). For example, if the average temperature for a given day is 85 degrees then the CDD for that day equals 20 and HDD equals zero. Similarly, if the average temperature for a given day is 50 degrees, then the HDD for that day equals 15 and CDD equals zero.

To filter the daily impact of seasonality from the data, the Frisch-Waugh theorem is used (Park, Mjelde, and Bessler 2007). Under the Frisch-Waugh theorem, the partial regression coefficients are estimated by a simple regression (Baltagi 2011). Following Park, Mjelde, and Bessler (2007), the ordinary least squares are applied to regress each data series (x_{it}) separately on lagged CDD and HDD.

(4.11) $x_{it} = \delta_i + \omega_i CDD_{t-1} + \pi_i HDD_{t-1} + e_{it}$,

where δ_i is a constant term, $\omega_i \; and \; \pi_i$ are coefficients of lagged CDD and HDD, and subscript *i* indicates each data series, *t* is time (day), and *e* is the error term. In this study the residual, $\hat{e}_{it}$, from the filtering regressions are the filtered data for the *i*th series.

The use of the filtered data, $\hat{e}_{1t}$ and $\hat{e}_{2t}$ provides estimates of the constant lower and upper bound values, $c^{(1)}$ and $c^{(2)}$. The relationship of the middle regime thereby follows

(4.12) $c^{(1)} \leq \hat{e}_{1t} - \hat{e}_{2t} \leq c^{(2)}$,

where $\hat{e}_{it} = x_{it} - \hat{\delta}_i - \widehat{\omega}_i CDD_{t-1} - \hat{\pi}_i HDD_{t-1}$, i = 1 and 2 (Park, Mjelde, and Bessler 2007). Following Park, Mjelde, and Bessler (2007), dynamic (daily) threshold values are

(4.13) $c_t^{(1)} = c^{(1)} + (\hat{\delta}_1 + \widehat{\omega}_1 CDD_{t-1} + \hat{\pi}_1 HDD_{t-1} - \hat{\delta}_2 - \widehat{\omega}_2 CDD_{t-1} - \hat{\pi}_2 HDD_{t-1})$,

and

(4.14) $c_t^{(2)} = c^{(2)} + (\hat{\delta}_1 + \widehat{\omega}_1 CDD_{t-1} + \hat{\pi}_1 HDD_{t-1} - \hat{\delta}_2 - \widehat{\omega}_2 CDD_{t-1} - \hat{\pi}_2 HDD_{t-1})$.

The time-varying thresholds are recovered from equations (4.13) and (4.14).

Data

Eight natural gas spot prices in Canada and United States are considered: AECO Hub, Alberta, Canada; Chicago City Gate, Illinois; Dominion South Point, Pennsylvania; Henry Hub, Louisiana; Malin, Oregon; Oneok, Oklahoma; Opal, Wyoming; and Waha Hub, Texas. Weekday nominal prices of natural gas from October 2, 2000 to October 31, 2014 are obtained from Bloomberg L.P. (2015). A missing value is replaced by the prior day's price. Each price is the closing price for a specific location for natural gas to be delivered on the next day. All prices are in U.S. dollars per MMBtu (a unit of heat equal to one million British thermal units). U.S. daily degree-days are from the National Oceanic and Atmospheric Administration (2014).

Because of the possible existence of structural changes occurring around 2000 and 2009 (see Chapter II), the data are divided into two subsamples. The first subsample contains 2,414 observations from October 2, 2000 to December 31, 2009. The second subsample contains 1,261 observations from January 1, 2010 to October 31, 2014.

Table 4.1. Augmented Dickey-Fuller (ADF) and Kwiatkowski-Philips-Schmidt-Shin (KPSS) Test[a] Statistics of Eight Natural Gas Spot Prices and Daily Degree-Days for Each Subsample

	ADF		KPSS		ADF		KPSS	
Price Series	t-Stat	Lag (k)	LM-Stat	Band width	t-Stat	Lag (k)	LM-Stat	Band width
	First Subsample[b]: October 2, 2000 - December 31, 2009				Second Subsample[c]: January 1, 2010 - October 31, 2014			
	Test in Level				Test in Level			
AECO Hub	-2.8842	4	1.3685	39	-2.7746	15	0.5794	28
Chicago	-3.5625	3	1.2629	39	-3.8886	14	0.4074	27
Dominion South	-3.5713	6	1.2694	39	-2.9281	15	1.3039	29
Henry Hub	-3.1836	3	1.3229	39	-2.9068	16	0.6716	29
Malin	-4.9451	6	0.4972	38	-3.4403	11	0.5958	28
Oneok	-4.0686	2	0.9425	39	-5.7145	6	0.5859	27
Opal	-3.6516	5	0.8156	39	-5.3457	6	0.5200	28
Waha Hub	-3.7079	6	1.0696	39	-4.0220	8	0.6250	28
HDD	-3.3042	5	0.0425	39	-2.8720	6	0.0804	29
CDD	-3.4709	5	0.0257	39	-2.7993	4	0.0516	29
	Test in First Difference				Test in First Difference			
AECO Hub	-25.6692	3	0.0262	5	-13.6688	14	0.1287	85
Chicago	-44.9292	1	0.0297	40	-16.2620	13	0.0142	17
Dominion South	-21.5989	5	0.0226	26	-8.5709	14	0.0749	32
Henry Hub	-31.3169	2	0.0286	26	-9.8420	15	0.1366	51
Malin	-27.8001	5	0.0348	163	-17.3823	10	0.1415	174
Oneok	-42.0385	1	0.0293	49	-19.1227	7	0.0986	129
Opal	-22.1266	6	0.0304	56	-18.0357	8	0.0932	117
Waha Hub	-24.0828	5	0.0309	59	-19.0725	7	0.1194	114
HDD	-30.5740	4	0.0304	23	-19.7109	5	0.0517	92
CDD	-28.3909	4	0.0148	56	-24.2223	3	0.0355	70

Note: Under the null hypothesis of non-stationarity (unit root), the ADF test critical value at 1%, and 5% levels are -3.430 and -2.860; the null is rejected when t-Stat < the critical value (Said and Dickey 1984). Under the null hypothesis of stationarity, the KPSS test critical value at 1% and 5% levels are 0.739 and 0.463; the null is rejected when LM-stat > the critical value (Kwiatkowski et al. 1992).

[a] Only constant term is included in equations.

[b] Lag (k) is selected from 0 to 20 based on Schwarz information criteria.

[c] Bandwidth is estimated using the Newey-West (1994) method.

Empirical Results

Augmented Dickey-Fuller (ADF) tests (Said and Dickey 1984) are employed under the null hypothesis that each price series has a unit root. Under the null hypothesis of unit root, the ADF test may have lower power against the alternative hypothesis of stationarity (DeJong et al. 1992). The Kwiatkowski-Philips-Schmidt-Shin (KPSS) test (Kwiatkowski et al. 1992) under the null hypothesis of stationarity is also employed. ADF and KPSS tests give somewhat contradicting results (table 4.1).

In the first subsample, ADF test statistics reveal that all price series except AECO Hub and Henry Hub are stationary at the 1% level as the null hypothesis that price series has unit root is rejected at the 1% level. KPSS test statistics, however, indicates that all prices but Malin have a unit root as the null hypothesis of stationarity is rejected at 1% level. At the 5% level, it appears that all prices are stationary when based on the ADF test, but all prices are non-stationary when based on the KPSS test. When based on the ADF test, HDD is stationary at the 5% level and CDD is stationary at the 1% level; whereas, when based on the KPSS test, the null hypothesis of stationarity of both HDD and CDD cannot be rejected at either the 1% or 5% levels.

In the second subsample, ADF test statistics of all price series except AECO Hub, Dominion South, and Henry Hub suggest the null hypothesis of unit root is rejected at the 1% level, implying Chicago, Malin, Oneok, Opal, and Waha Hub prices are stationary at the 1% level. KPSS test statistics of all prices except Dominion South suggest the null hypothesis that price series are stationary cannot be rejected at the 1% level, implying all prices but Dominion South price are stationary at the 1% level. At the 5% level, ADF test statistics indicate that all prices are stationary. In contrast, KPSS test statistics indicate that all prices but Chicago have a unit root at the 5% level. ADF test statistics of HDD and CDD suggest that HDD and CDD are non-stationary at the 1% level because the null hypothesis of unit root cannot be reject at the 1% level. Nevertheless, KPSS test statistics of HDD and CDD suggest that HDD and CDD are stationary because the null hypothesis of stationarity cannot be rejected. Regardless of the test and the subsample, all price series are stationary after first differencing.

Obtaining Filtered Data

The ordinary least squares regression of each price series on the lagged HDD and CDD is implemented to obtain filtered data. Results of the filtering regression for each subsample are presented in table 4.2. In the first subsample, all estimated coefficients of the lagged HDD are significant at the 1% level; coefficients of the lagged CDD are significant at the 1% level in explaining all prices but AECO Hub, Henry Hub, and Malin. Coefficients of the lagged CDD in Henry Hub and Malin are significant at the 5% and 10% levels. In the second subsample, all estimated coefficients of the lagged HDD and CDD are significant at the 1% level. The positivity of all parameter estimates reveals that an increase of either HDD or CDD leads to a rise in natural gas prices. Since HDD and CDD are different each day, the residuals from filtering regressions capture the seasonality in natural gas prices. Filtered prices and original prices are plotted in figure 4.2. As expected the filter prices are smaller than the original, but the general patterns of price movement are the same.

Table 4.2. Results of the Filtering Regression

Price Series	Constant	$HDD_{(t-1)}$	$CDD_{(t-1)}$
		First Subsample (October 2, 2000 - December 31, 2009)	
AECO Hub	4.8129	0.0309	-0.0038
	(0.0000)	(0.0000)	(0.8124)
Chicago	5.2804	0.0439	0.0479
	(0.0000)	(0.0000)	(0.0074)
Dominion South	5.6393	0.0474	0.0527
	(0.0000)	(0.0000)	(0.0074)
Henry Hub	5.4836	0.0343	0.0380
	(0.0000)	(0.0000)	(0.0417)
Malin	4.8425	0.0659	0.0410
	(0.0000)	(0.0000)	(0.0529)
Oneok	4.5806	0.0455	0.0736
	(0.0000)	(0.0000)	(0.0000)
Opal	3.8209	0.0572	0.0468
	(0.0000)	(0.0000)	(0.0038)
Waha Hub	4.6817	0.0436	0.0752
	(0.0000)	(0.0000)	(0.0000)
		Second Subsample (January 1, 2010 - October 31, 2014)	
AECO Hub	2.9250	0.0343	0.0287
	(0.0000)	(0.0000)	(0.0008)
Chicago	2.6645	0.0917	0.1210
	(0.0000)	(0.0000)	(0.0000)
Dominion South	2.8604	0.0502	0.0697
	(0.0000)	(0.0000)	(0.0000)
Henry Hub	3.1687	0.0368	0.0668
	(0.0000)	(0.0000)	(0.0000)
Malin	3.1594	0.0406	0.0497
	(0.0000)	(0.0000)	(0.0000)
Oneok	2.8472	0.0507	0.0806
	(0.0000)	(0.0000)	(0.0000)
Opal	2.9468	0.0445	0.0577
	(0.0000)	(0.0000)	(0.0000)
Waha Hub	2.9992	0.0428	0.0717
	(0.0000)	(0.0000)	(0.0000)

Note: *p*-values are in parenthesis

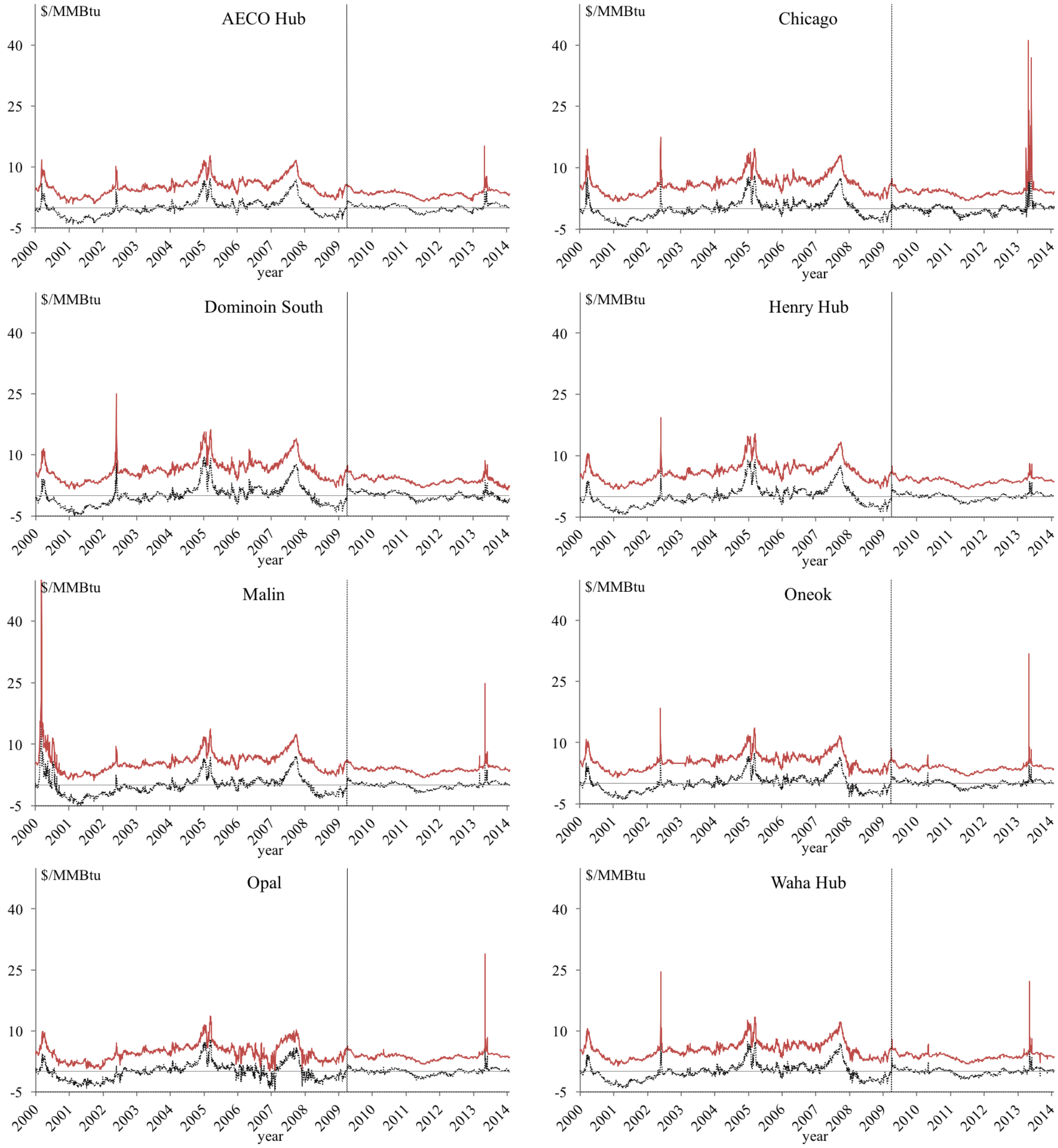

Figure 4.2. Original daily natural gas spot prices (red solid line) and filtered data (black dotted line) (October 2, 2000 to October 31, 2014)

Note: The vertical (dashed) line indicates January 1, 2010.

Table 4.3. Augmented Dickey-Fuller (ADF) and Kwiatkowski-Philips-Schmidt-Shin (KPSS) Test[a] Statistics of Filtered Data for Each Subsample

	ADF		KPSS		ADF		KPSS	
Price Series	t-Stat	Lag (k)	LM-Stat	Band width	t-Stat	Lag (k)	LM-Stat	Band width
	First Subsample: October 2, 2000 - December 31, 2009				Second Subsample: January 1, 2010 - October 31, 2014			
	Test in Level				Test in Level			
AECO Hub	-2.7026	3	1.4532	39	-2.7096	11	0.6121	28
Chicago	-3.4521	2	1.3323	39	-4.1770	14	0.5245	26
Dominion South	-3.3446	6	1.3367	39	-2.8092	15	1.3546	29
Henry Hub	-2.9895	2	1.3703	39	-2.9858	5	0.6899	29
Malin	-5.0823	6	0.5586	38	-3.2580	11	0.6589	28
Oneok	-3.8805	3	1.0003	39	-4.6162	9	0.6355	27
Opal	-3.1772	7	0.9111	39	-5.6505	6	0.6296	27
Waha Hub	-3.7656	4	1.1269	39	-3.6905	9	0.6651	28
	Test in First Difference				Test in First Difference			
AECO Hub	-34.7748	2	0.0300	14	-17.4256	10	0.1294	123
Chicago	-46.2531	1	0.0369	49	-16.1962	13	0.0113	16
Dominion South	-22.7601	5	0.0281	29	-8.9452	14	0.0921	48
Henry Hub	-46.7860	1	0.0341	30	-23.4845	4	0.1475	62
Malin	-27.7416	5	0.0374	143	-18.5145	10	0.1789	283
Oneok	-35.1983	2	0.0347	58	-18.3075	9	0.1320	195
Opal	-23.5370	6	0.0405	70	-18.0875	9	0.1020	153
Waha Hub	-32.1484	3	0.0367	68	-18.9187	8	0.1571	188

Note: Under the null hypothesis of non-stationarity (unit root), the ADF test critical value at 1%, and 5% levels are -3.430 and -2.860; the null is rejected when t-Stat < the critical value (Said and Dickey 1984). Under the null hypothesis of stationarity, the KPSS test critical value at 1% and 5% levels are 0.739 and 0.463; the null is rejected when LM-stat > the critical value (Kwiatkowski et al. 1992).

[a] Only constant term is included in equations.

[b] Lag (k) is selected from 0 to 20 based on Schwarz information criteria.

[c] Bandwidth is estimated using the Newey-West (1994) method.

ADF and KPSS test statistics of filtered data are shown in table 4.3. ADF test statistics of Chicago, Malin, Oneok, and Waha Hub filtered prices in the first subsample suggest the unit root hypothesis is rejected at the 1% level, indicating that these filtered prices are stationary at the 1% level. KPSS test statistics of all prices except Malin suggest the stationarity hypothesis is rejected at the 1% level, indicating that all filtered prices except Malin have a unit root at the 1% level. At the 5% level, all filtered prices except AECO Hub are stationary based the ADF test but all filtered prices are non-stationary based the KPSS test.

In the second subsample, the ADF test indicates that filtered prices of Chicago, Oneok, Opal, and Waha Hub are stationary at the 1% level. The KPSS test indicates that all filtered prices except Dominion South are stationary as test statistics of all filtered prices but Dominion South suggest the null hypothesis of stationarity cannot be rejected at the 1% level. At the 5% level, the ADF test indicates that all filtered prices except AECO Hub are stationary whereas the KPSS test indicates that all filtered prices have unit root at the 5% level. Regardless of the test and the subsample, all filtered prices are stationary after first differencing. It appears that the KPSS test gives more consistent results between original prices and filtered prices than does the ADF test. As the ADF and KPSS tests yield contradicting results, a test of variable stationarity, which is a multivariate version of the Dickey-Fuller test under the null hypothesis that each individual series is stationary given the cointegration space, is also executed.

Table 4.4. Schwarz Loss Measures on One to Two Cointegrating Vectors (Rank) and One to Five Lags on VECM Model of Seven Market-Pairs Using Filtered Data

Market Pairs	Rank	One Lag	Two Lags	Three Lags	Four Lags	Five Lags
			First Subsample (October 2, 2000 - December 31, 2009)			
AECO-	1	-4.8801	-4.9801	-5.0401	-5.0455*	-5.0345
Henry Hub	2	-4.8789	-4.9789	-5.0373	-5.0411	-5.0312
Chicago-	1	-4.9082	-5.0112	-5.1023	-5.1378	-5.1463*
Henry Hub	2	-4.9081	-5.0110	-5.1001	-5.1356	-5.1434
Dominion South-	1	-4.6378	-4.8823	-4.9454	-5.0571	-5.0744*
Henry Hub	2	-4.6390	-4.8789	-4.9423	-5.0552	-5.0712
Malin-	1	-2.1602	-2.1788	-2.2591*	-2.2510	-2.2445
Henry Hub	2	-2.1600	-2.1786	-2.2562	-2.2478	-2.2431
Oneok-	1	-4.8234	-4.9302	-5.0039	-5.0156	-5.0202*
Henry Hub	2	-4.8253	-4.9287	-5.0001	-5.0143	-5.0178
Opal-	1	-3.4045	-3.4342	-3.5314	-3.5390	-3.5485*
Henry Hub	2	-3.4051	-3.4343	-3.5278	-3.5372	-3.5454
Waha-	1	-4.7682	-4.7788	-4.8582	-4.8610	-4.8741*
Henry Hub	2	-4.7660	-4.7783	-4.8554	-4.8582	-4.8713
			Second Subsample (January 1, 2010 - October 31, 2014)			
AECO-	1	-5.8102	-5.9094	-6.0148	-6.0344	-6.0789*
Henry Hub	2	-5.8204	-5.9173	-6.0143	-6.0321	-6.0755
Chicago-	1	-2.4578	-2.5664	-2.6834	-2.7019	-2.7071*
Henry Hub	2	-2.4664	-2.5751	-2.6830	-2.7010	-2.7042
Dominion South-	1	-7.2632	-7.3032	-7.5401	-7.5878	-7.6524*
Henry Hub	2	-7.2634	-7.3012	-7.5342	-7.5789	-7.6421
Malin-	1	-4.6521	-4.6850	-4.8478	-4.8445	-4.9113*
Henry Hub	2	-4.6589	-4.6932	-4.8490	-4.8421	-4.9067
Oneok-	1	-4.1103	-4.1531	-4.3092	-4.3043	-4.3921*
Henry Hub	2	-4.1162	-4.1589	-4.3101	-4.3032	-4.3873
Opal-	1	-4.2920	-4.3520	-4.5074	-4.4989	-4.5942*
Henry Hub	2	-4.2978	-4.3589	-4.5083	-4.4984	-4.5891
Waha-	1	-5.0590	-5.1023	-5.2554	-5.2563	-5.3382*
Henry Hub	2	-5.0663	-5.1083	-5.2567	-5.2540	-5.3341

Note: The asterisk '*' indicates minimum values of Schwarz loss measure.

Testing Threshold Cointegration

As noted in Chapter II, Henry Hub is an important market for pricing of the North America natural gas spot and future markets; it is used as the benchmark market in both the linear VECM and the three-regime TVECM. As such, seven market pairs of AECO Hub-Henry Hub, Chicago-Henry Hub, Dominion South-Henry Hub, Malin-Henry Hub, Oneok-Henry Hub, Opal-Henry Hub, and Waha Hub-Henry Hub are modeled.

The first step of testing threshold cointegration is to test if there exist pair-wise cointegration. Cointegrating vector and lag length are simultaneously determined by the Schwarz loss metric. The minimum Schwarz loss measure suggests one cointegrating vector for all market pairs in the two filtered subsamples (table 4.4). In the first subsample, the minimum Schwarz loss metric suggest three lags for Malin-Henry Hub, four lags for AECO Hub-Henry Hub, and five lags for the other five market-pairs. The minimum SL is at five lags for all market pairs in the second subsample. Five lags may indicate a day of the week effect.

Conditional on one cointegrating vector, results from tests of exclusion, stationarity, and weak exogeneity are reported in table 4.5. In the first subsample, likelihood ratio test statistics and corresponding *p*-values indicate that no price series can be excluded from the pair-wise long-run relationship, irrespective of the market pair, as the null hypothesis of exclusion is rejected at the 1% level. Tests of variable stationarity reveal that in the pair-wise models no price series can be considered stationary by itself when the cointegration rank equals one. Variable exogeneity is tested if any of prices in each market pair can be regarded as weakly exogenous when the parameter of interest is

β. A price series does not respond to perturbations in the long-run equilibrium when it is considered weakly exogenous. Henry Hub is considered weakly exogenous when it is paired with all other markets except AECO Hub. Paired with Henry Hub, AECO Hub is observed weakly exogenous.

In the second subsample, the null hypotheses of exclusion and stationarity are rejected for all prices in every market pair except Dominion South-Henry Hub. This means that Dominion South and Henry Hub prices can be excluded and considered stationary by themselves when the cointegration rank equals one. Conditional on one cointegrating vector, Henry Hub is regarded as weakly exogenous when it is paired with AECO Hub, Chicago, Malin, Oneok, and Opal.

Even though Schwarz loss measures suggest one cointegrating vector for all market pairs, irrespective of the subsample, the Engle-Granger approach is also used to confirm whether pair-wise cointegration exists. Estimated residuals from regressing individual price series on Henry Hub price are tested for stationarity. Stationarity of estimated residual infers the pair-wise cointegration. Following Balke and Fomby (1997), the ADF test is used. In table 4.6, estimated residuals from all market-pair regressions are stationary in the first subsample, implying the presence of pair-wise cointegration. In the second subsample, only estimated residuals obtained by regressing Dominion South on Henry Hub have unit root, implying that there exists pair-wise long-run relationships in all market pairs except Dominion South-Henry Hub.

Table 4.5. Results from Tests of Exclusion, Stationarity, and Weak Exogeneity for Seven Market-Pairs Using Filtered Data

Price Series in Each Market Pair	Exclusion LR-Test	Exclusion *p*-Value	Stationarity LR-Test	Stationarity *p*-Value	Weak Exogeneity LR-Test	Weak Exogeneity *p*-Value
			First Subsample (October 2, 2000 - December 31, 2009)			
AECO Hub	111.0187	0.0000	112.9263	0.0000	3.4590	0.0629
Henry Hub	112.9263	0.0000	111.0187	0.0000	43.3190	0.0000
Chicago	104.7309	0.0000	102.6047	0.0000	11.8626	0.0006
Henry Hub	102.6047	0.0000	104.7309	0.0000	2.6446	0.1039
Dominion South	72.1203	0.0000	70.0889	0.0000	18.0742	0.0000
Henry Hub	70.0889	0.0000	72.1203	0.0000	0.0079	0.9291
Malin	121.3449	0.0000	72.8529	0.0000	99.7822	0.0000
Henry Hub	72.8529	0.0000	121.3449	0.0000	3.1324	0.0768
Oneok	33.4556	0.0000	28.7997	0.0000	6.0849	0.0136
Henry Hub	28.7997	0.0000	33.4556	0.0000	0.7918	0.3736
Opal	28.9196	0.0000	25.8468	0.0000	12.4999	0.0004
Henry Hub	25.8468	0.0000	28.9196	0.0000	3.8624	0.0494
Waha Hub	53.9987	0.0000	48.0298	0.0000	21.9198	0.0000
Henry Hub	48.0298	0.0000	53.9987	0.0000	2.0918	0.1481
			Second Subsample (January 1, 2010 - October 31, 2014)			
AECO Hub	58.9369	0.0000	49.1118	0.0000	34.7695	0.0000
Henry Hub	49.1118	0.0000	58.9369	0.0000	0.0750	0.7842
Chicago	88.1067	0.0000	34.5516	0.0000	80.2483	0.0000
Henry Hub	34.5516	0.0000	88.1067	0.0000	0.3451	0.5569
Dominion South	0.5344	0.4648	1.3851	0.2392	8.4986	0.0036
Henry Hub	1.3851	0.2392	0.5344	0.4648	8.8127	0.0030
Malin	170.6979	0.0000	143.3857	0.0000	125.7307	0.0000
Henry Hub	143.3857	0.0000	170.6979	0.0000	1.4846	0.2231
Oneok	208.6700	0.0000	166.4654	0.0000	167.6977	0.0000
Henry Hub	166.4654	0.0000	208.6700	0.0000	5.8819	0.0153
Opal	162.6634	0.0000	124.8073	0.0000	127.4584	0.0000
Henry Hub	124.8073	0.0000	162.6634	0.0000	3.5216	0.0606
Waha Hub	212.4244	0.0000	189.5004	0.0000	160.3772	0.0000
Henry Hub	189.5004	0.0000	212.4244	0.0000	9.0664	0.0026

Note: All pair models are conditional on one cointegrating vector.

Table 4.6. Augmented Dickey-Fuller (ADF) Test[a] Statistics of Estimated Residuals Obtained by Regressing Price Series on Henry Hub

	First Subsample (October 2, 2000 - December 31, 2009)		Second Subsample (January 1, 2010 - October 31, 2014)	
	Test in Level		Test in Level	
Estimated Residual from	t-Stat	Lag[b](k)	t-Stat	Lag[b](k)
AECO-Henry Hub	-7.9011	6	-3.9891	18
Chicago-Henry Hub	-13.2329	1	-5.7469	14
Dominion South-Henry Hub	-8.9409	9	-0.9250	10
Malin-Henry Hub	-7.9004	6	-18.8784	1
Oneok-Henry Hub	-6.4350	4	-28.8135	0
Opal-Henry Hub	-6.1791	4	-19.0176	1
Waha-Henry Hub	-5.2752	15	-28.5193	0

Note: The ADF critical value at 1 % level is -3.430.
[a] Only constant term is included in equations.
[b] Lag (k) is selected from 0 to 20 based on Schwarz information.

Table 4.7. Bootstrap *p*-values for Testing VECM against Three-Regime TVECM Using Filtered Data

	Bootstrap *p*-Values	
Market Pairs	First Subsample (October 2, 2000 - December 31, 2009)	Second Subsample (January 1, 2010 - October 31, 2014)
AECO-Henry Hub	0.0000	0.0000
Chicago-Henry Hub	0.0000	0.0000
Dominion South-Henry Hub	0.0000	0.0000
Malin-Henry Hub	0.0000	0.0000
Oneok-Henry Hub	0.0000	0.0300
Opal- Henry Hub	0.0000	0.0200
Waha-Henry Hub	0.0000	0.0100

Note: Less than 0.05 *p*-values indicate that three-regime TVECM is significantly better than VECM at the 5% level.

The second step of testing threshold cointegration is to test whether the deviation process of the cointegrating relationship is linear or not. Even though the existence of cointegration in some market pairs is ambiguous, the seven market-pairs are tested for non-linear cointegration with cautious interpretation. Under the null hypothesis of a linear VECM against the alternative of a three-regime TVECM, the sup-LR test statistic is obtained by estimating VECM and three-regime TVECM assuming that a cointegrating vector is known as (1, -1). Lag length in each model varies across pair-wise models based on the results in the first step. Bootstrap p-values are computed as the percentage of bootstrapped LR statistics, which are greater than the observed LR statistics (Hansen 1999). In table 4.7, bootstrap p-values are less than 0.05, indicating that three-regime TVECM is significantly better than VECM at the 5% level in all seven market-pairs.

In table 4.8, $c^{(1)}$ and $c^{(2)}$, the lower and upper threshold bounds, obtained from the sequential conditional least squares regression, are presented. The difference between the upper and lower threshold values ($c^{(2)} - c^{(1)}$) and average of the threshold values are also presented (table 4.8). When the price difference is less (greater) than the lower (upper) threshold value, the system is in regime one (three). The system is in regime two, when the price difference is between the lower and upper bounds. Regime two is the arbitrage-free range of price differences; one would not benefit from arbitrage trading between any two markets. Numbers of observations and percentages of observations in each regime are given in table 4.9.

Table 4.8. Estimated Threshold Values for Seven Market-Pairs

Market Pairs	Lower Bound $c^{(1)}$	Upper Bound $c^{(2)}$	$c^{(2)}-c^{(1)}$	Average $(c^{(1)}, c^{(2)})$
	First Subsample (October 2, 2000 - December 31, 2009)			
AECO-Henry Hub	-0.5016	0.5348	1.0364	0.0166
Chicago-Henry Hub	-0.1758	0.2344	0.4102	0.0293
Dominion South-Henry Hub	-0.1112	0.2488	0.3600	0.0688
Malin-Henry Hub	-0.8595	0.5705	1.4300	-0.1445
Oneok-Henry Hub	-0.3740	-0.0787	0.2953	-0.2264
Opal- Henry Hub	-1.4951	0.7859	2.2810	-0.3546
Waha-Henry Hub	0.1390	0.5153	0.3763	0.3272
	Second Subsample (January 1, 2010 - October 31, 2014)			
AECO-Henry Hub	-0.2826	0.2851	0.5677	0.0013
Chicago-Henry Hub	-0.5669	0.2030	0.7699	-0.1820
Dominion South-Henry Hub	-0.6673	-0.0056	0.6617	-0.3365
Malin-Henry Hub	-0.1721	0.1687	0.3408	-0.0017
Oneok-Henry Hub	-0.1366	0.1116	0.2482	-0.0125
Opal- Henry Hub	-0.1454	0.1925	0.3379	0.0236
Waha-Henry Hub	-0.1202	0.0894	0.2096	-0.0154

Table 4.9. Numbers of Observations and Percentages of Observations in Three Regimes

Market Pairs	No. of Obs in Regime One	No. of Obs in Regime Two	No. of Obs in Regime Three	% of Obs in Regime One	% of Obs in Regime Two	% of Obs in Regime Three
	First Subsample (October 2, 2000 - December 31, 2009)					
AECO-Henry Hub	262	1893	253	10.88	78.61	10.51
Chicago-Henry Hub	385	1763	259	16.00	73.24	10.76
Dominion South-Henry Hub	808	1352	247	33.57	56.17	10.26
Malin-Henry Hub	284	1881	244	11.79	78.08	10.13
Oneok-Henry Hub	460	254	1693	19.11	10.55	70.34
Opal- Henry Hub	259	1517	631	10.76	63.02	26.22
Waha-Henry Hub	1114	1041	252	46.28	43.25	10.47
	Second Subsample (January 1, 2010 - October 31, 2014)					
AECO-Henry Hub	154	968	132	12.28	77.19	10.53
Chicago-Henry Hub	218	910	126	17.38	72.57	10.05
Dominion South-Henry Hub	154	206	894	12.28	16.43	71.29
Malin-Henry Hub	162	966	126	12.92	77.03	10.05
Oneok-Henry Hub	216	912	126	17.22	72.73	10.05
Opal- Henry Hub	245	881	128	19.54	70.26	10.21
Waha-Henry Hub	146	978	130	11.64	77.99	10.37

Obtaining Time-Varying Threshold Values

Time-varying threshold values recovered using equations (4.13) and (4.14) are illustrated along with (original) price differences in figures 4.3-4.9. How frequent price differences are in each regime appears to be consistent with results in table 4.9. In the first subsample, price differences of all market-pairs except Oneok-Henry Hub and Waha Hub-Henry Hub are most often observed in regime two with the percentages being 63% or larger. Oneok-Henry Hub price differences are usually in regime three. Price differences between Waha Hub-Henry Hub are almost evenly split between regimes one and two. In the second subsample, over 70% of the time price differences of all market pairs except Dominion South-Henry Hub are in regime two; price differences between Dominion South and Henry Hub are most often observed in regime three.

Between the two subsamples, percentages of AECO-Henry Hub, Chicago-Henry Hub, and Malin-Henry Hub price differences in the three regimes are approximately the same. In the second subsample, percentages of Dominion South-Henry Hub price differences observed in regimes one and two are less than those in the first subsample. As previously noted, the percentage of observed Dominion South-Henry Hub price differences increases for regime three. The percentage of Oneok-Henry Hub price differences in regime one is approximately the same in the two subsamples, while the percentage in regime two (three) increases (decreases) in the second subsample. In the second subsample, percentages of Opal-Henry Hub price differences in regimes one and two are greater than those in the first subsample, while the percentage in regime three in the second subsample are lower than that in the first subsample. Compared to the first subsample, percentages of Waha Hub-Henry Hub observations in regimes one and three decrease in the second subsample relative to the first; regime two percentage increases in the second subsample.

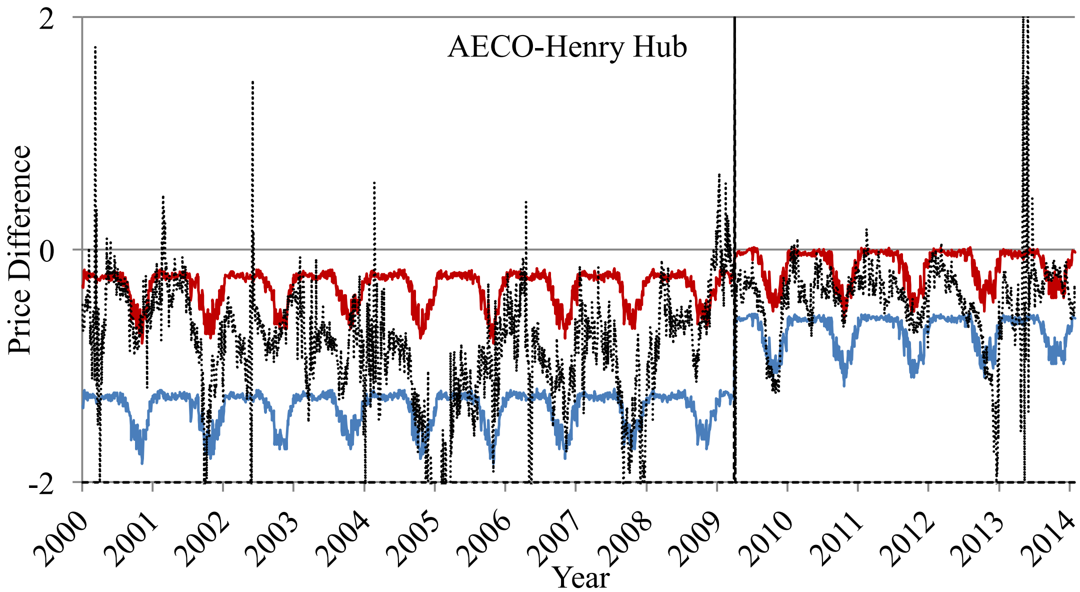

Figure 4.3. Time-varying upper and lower threshold values (red and blue solid line) and original daily price differences (black dotted lines) between AECO Hub and Henry Hub (October 2, 2000 to October 31, 2014)

Note: The vertical (dashed) line indicates January 1, 2010.

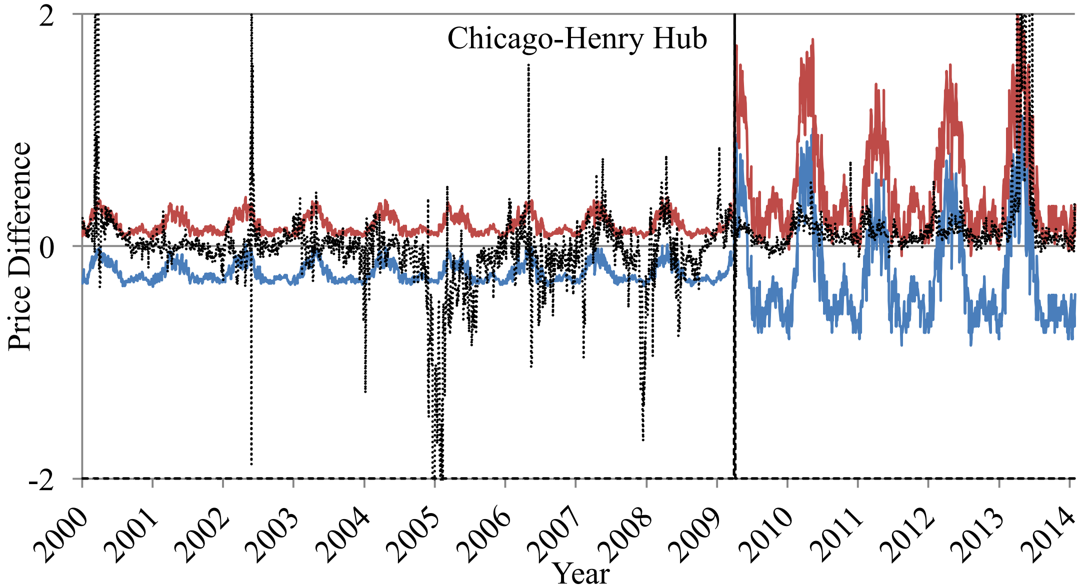

Figure 4.4. Time-varying upper and lower threshold values (red and blue solid line) and original daily price differences (black dotted lines) between Chicago and Henry Hub (October 2, 2000 to October 31, 2014)

Note: The vertical (dashed) line indicates January 1, 2010.

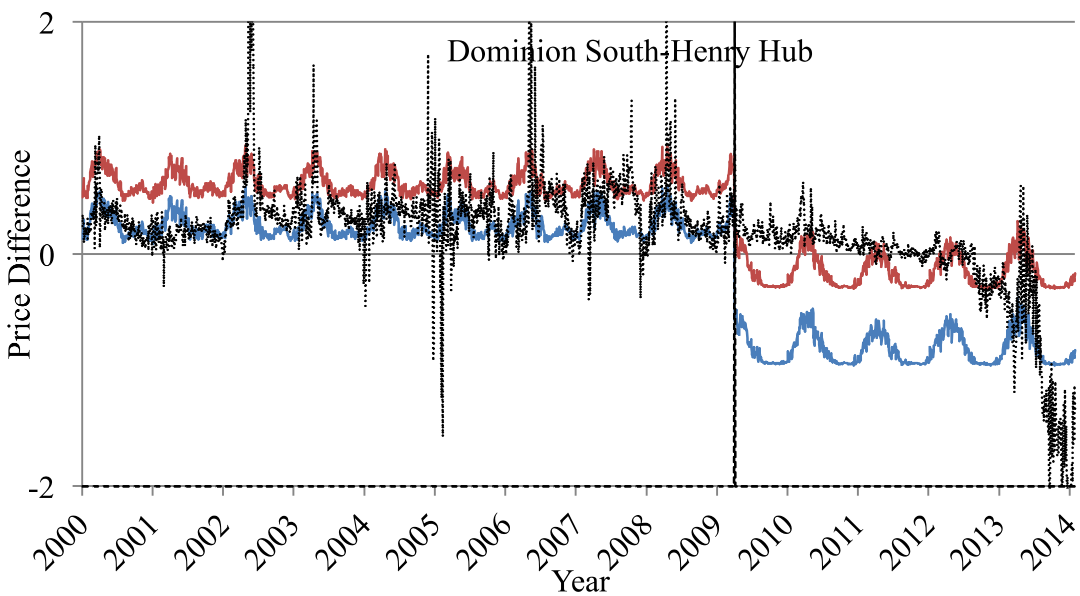

Figure 4.5. Time-varying upper and lower threshold values (red and blue solid line) and original daily price differences (black dotted lines) between Dominion South and Henry Hub (October 2, 2000 to October 31, 2014)

Note that the vertical (dashed) line separates the first and second subsamples.

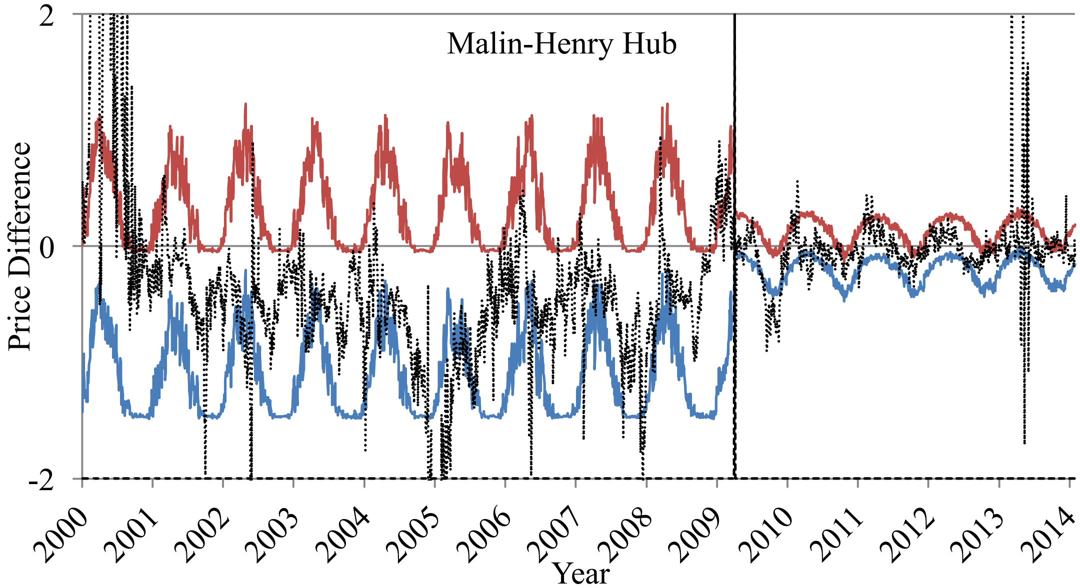

Figure 4.6. Time-varying upper and lower threshold values (red and blue solid line) and original daily price differences (black dotted lines) between Malin and Henry Hub (October 2, 2000 to October 31, 2014)
Note: The vertical (dashed) line indicates January 1, 2010.

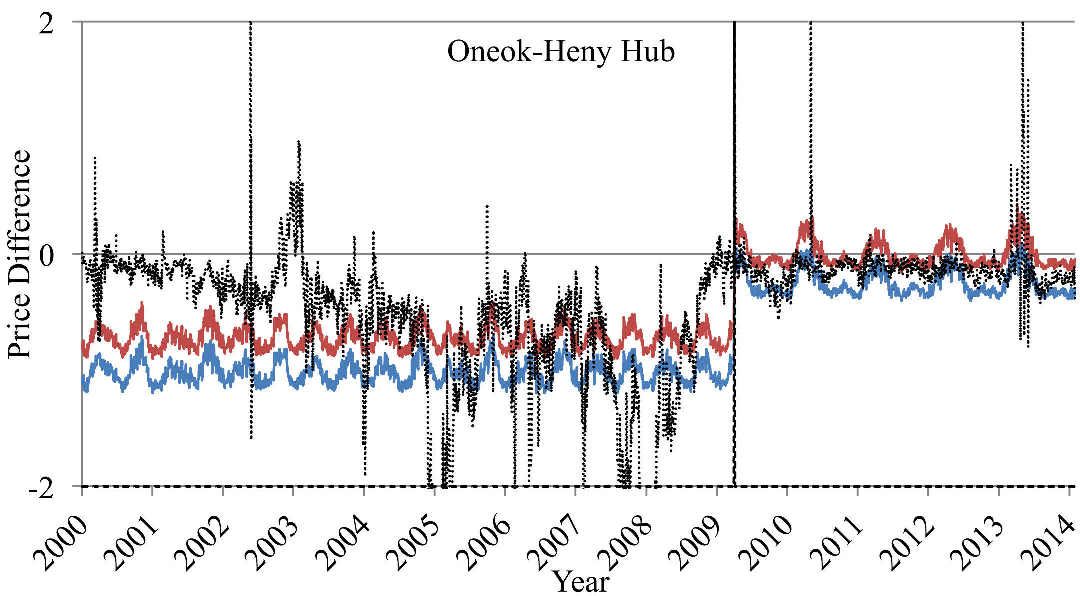

Figure 4.7. Time-varying upper and lower threshold values (red and blue solid line) and original daily price differences (black dotted lines) between Oneok and Henry Hub (October 2, 2000 to October 31, 2014)
Note: The vertical (dashed) line indicates January 1, 2010.

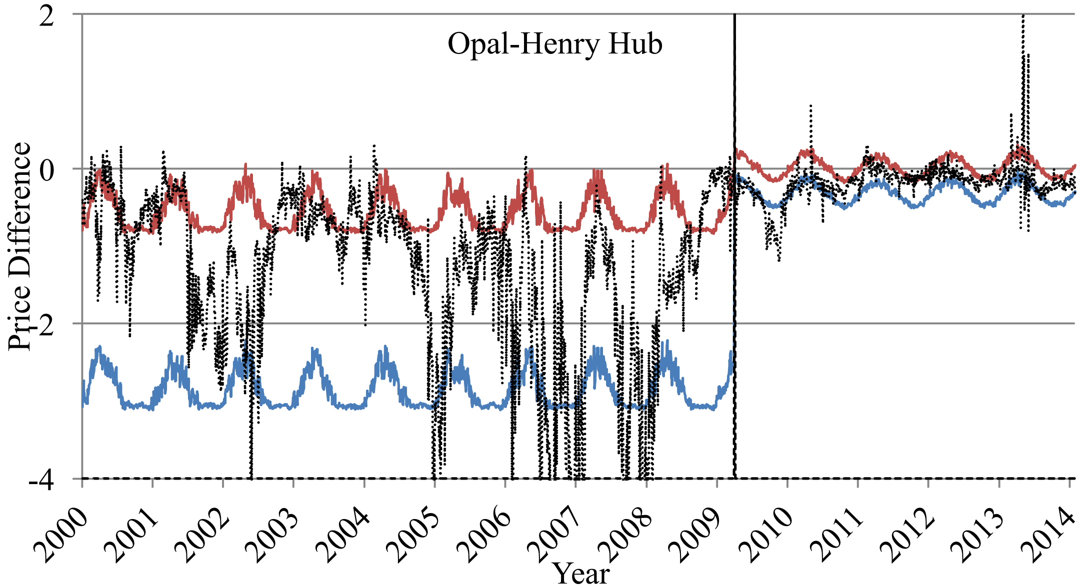

Figure 4.8. Time-varying upper and lower threshold values (red and blue solid line) and original daily price differences (black dotted lines) between Opal and Henry Hub (October 2, 2000 to October 31, 2014)

Note: The vertical (dashed) line indicates January 1, 2010.

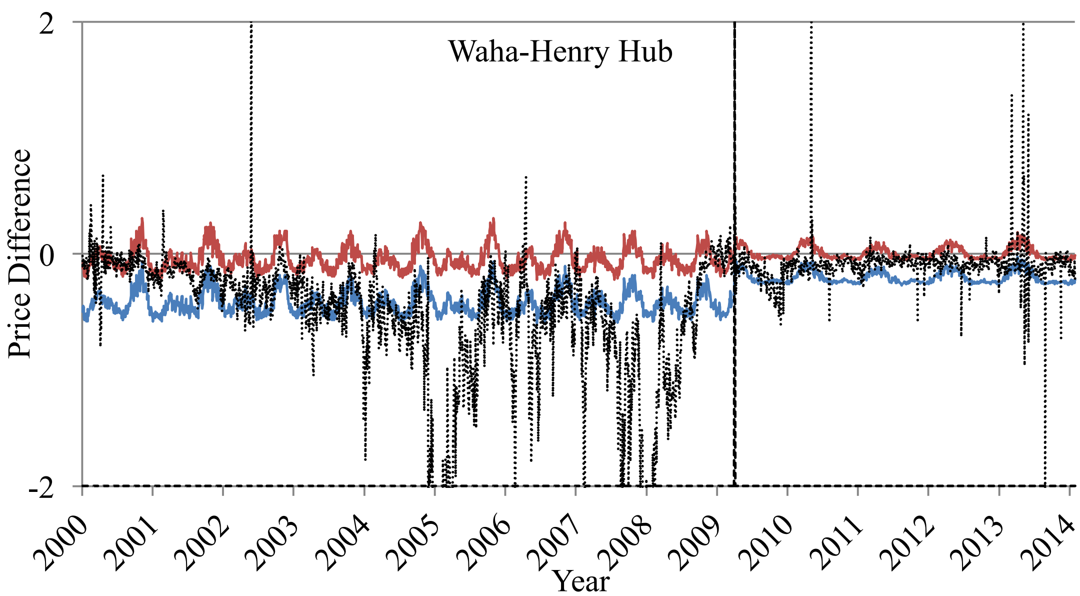

Figure 14. Figure 4.9. Time-varying upper and lower threshold values (red and blue solid line) and original daily price differences (black dotted lines) between Waha Hub and Henry Hub (October 2, 2000 to October 31, 2014)

Note: The vertical (dashed) line indicates January 1, 2010.

Averages of recovered time-varying threshold values are presented in table 4.10. Gaps between the time-varying upper and lower bounds are equal to gaps between the constant threshold bounds. Threshold gaps of each market-pair appear to be different between the two subsamples. Threshold intervals of AECO Hub-Henry Hub, Malin-Henry Hub, Oneok-Henry Hub, Opal-Henry Hub, and Waha Hub-Henry Hub become narrower in the second subsample while threshold intervals of Chicago-Henry Hub and Dominion South-Henry Hub become wider in the second subsample.

Table 4.10. Averages of Recovered Time-Varying Threshold Values

Market Pairs	Average of $c_t^{(1)}$	Average of $c_t^{(2)}$	$c_t^{(2)} - c_t^{(1)}$
	First Subsample (October 2, 2000 - December 31, 2009)		
AECO-Henry Hub	-1.3599	-0.3235	1.0364
Chicago-Henry Hub	-0.2299	0.1803	0.4102
Dominion South-Henry Hub	0.2516	0.6116	0.3600
Malin-Henry Hub	-1.1140	0.3160	1.4300
Oneok-Henry Hub	-1.0188	-0.7235	0.2953
Opal- Henry Hub	-2.8544	-0.5734	2.2810
Waha-Henry Hub	-0.4219	-0.0456	0.3763
	Second Subsample (January 1, 2010 - October 31, 2014)		
AECO-Henry Hub	-0.7032	-0.1355	0.5677
Chicago-Henry Hub	-0.2433	0.5266	0.7699
Dominion South-Henry Hub	-0.8142	-0.1525	0.6617
Malin-Henry Hub	-0.2051	0.1357	0.3408
Oneok-Henry Hub	-0.2487	-0.0005	0.2482
Opal- Henry Hub	-0.3165	0.0214	0.3379
Waha-Henry Hub	-0.2031	0.0065	0.2096

Discussion

Threshold values of each market-pair during the period of October 2, 2000 to December 31, 2009 appear to be different from what is observed for the period of January 1, 2010 to October 31, 2014. Threshold bands of AECO Hub-Henry Hub, Malin-Henry Hub, Oneok-Henry Hub, Opal-Henry Hub, and Waha Hub-Henry Hub become narrower in the latter period; whereas, threshold bands of Chicago-Henry Hub and Dominion South-Henry Hub become wider. As threshold intervals may be induced by transaction costs including transportation costs, the narrower (wider) intervals likely suggest the lower (higher) transaction costs. Lower (higher) transaction costs potentially lead to more (less) natural gas trading between hubs.

As noted in literature, transaction costs may not solely cause price differences between markets but also monetary policies, policy interventions, the behavior of inventories, and other economic risk factors (Balke and Fomby 1997; Tsay 1998). In natural gas spot markets, in additional to transaction costs, the location of each market, whether it is located in/near production/reserve area, each individual market's role in price discovery, and/or pipeline/transportation constraints may be behind price differences between markets.

Between January 2010 and October 2014, most observed price differences of all market-pairs except Dominion South-Henry Hub are in the arbitrage-free range (regime two). It should be note that natural gas trading may occur for reasons other than arbitrage. The major changes in the percentage of time spent in each regime for the three market-pairs, Oneok-Henry Hub, Waha Hub-Henry Hub, and Dominion South-

Henry Hub, are noteworthy. The percentages of Oneok-Henry Hub and Waha-Henry Hub (Dominion South-Henry Hub) observations in regime two noticeably increase (decrease) even though their threshold bands get narrower (wider). Because of the shale gas revolution resulting in the potential presence of structural change during 2009, the amount of natural gas produced from these areas has increased. In 2013, Texas, Louisiana, Pennsylvania, Arkansas, and Oklahoma were the top five largest shale gas producing states (U.S. EIA 2014h, 2014i, 2014j). One potential difference between Oneok and Waha Hub and Dominion South, however, is the existing pipeline system; Oneok and Waha Hub are in the regions that pipelines have been built to transport natural gas while Dominion South is in a region that there are no sufficient pipelines to takeaway substantial volumes of natural gas.

There is no incentive to trade natural gas between markets unless transaction costs are less than price differences. Since the shale gas bloom, trading between markets that are located in/near the production areas and Henry Hub may be less active. Transaction costs of these market-pairs might be cut down to induce trading. This may be behind lower transaction costs of AECO Hub-, Oneok-, Opal-, and Waha Hub-Henry Hub as these markets are located in/nearby major natural gas production areas. Given the model only includes prices one can only speculate on trading, but, as production areas experienced an increase in production, these areas are able to supply their associated demand areas without as much natural gas from other regions. Transaction costs must decrease in this case to induce trade. This explanation is mostly likely what is occurring between Waha Hub-Henry Hub and Oneok-Henry Hub. Both Waha Hub and Oneok are experiencing shale gas increases; the Eagle Ford in Texas and the Woodford in Oklahoma (U.S. EIA 2014i, 2014j; American Petroleum Institute 2014) and have pipeline capacity to supply their demand areas.

The higher transaction costs between Dominion South and Henry Hub are possibly the results of the lack of takeaway capacity. Pennsylvania previously relied on natural gas from the Gulf Coast. With the development in the Marcellus shale Pennsylvania now can fulfil its own demand becoming less dependent to no dependent on natural gas inflows from other states (U.S. EIA 2013c, 2015g). Dry natural gas production in Pennsylvania has increased since 2010 (figure 4.10) and Pennsylvania has become one of the top five natural gas producing states (U.S. EIA 2015e, 2015g). Because of the massive production in the Marcellus area, pipelines are being transformed to supply natural gas to the Midwest and the Gulf Coast (U.S. EIA 2015g). The Dominion South area does not have the necessary pipeline capacity to efficiently transport increased natural gas causing bottlenecks (Grimes 2014; American Petroleum Institute 2014). This increase in pipeline demand coupled with a relatively fixed pipeline supply is most likely the cause of the increase in transaction costs. These changes likely result in less natural gas trading between Dominion South and Henry Hub.

Physical trading between some natural gas hubs, for example, AECO Hub and Henry Hub, may not be feasible because of transportation constraints. Examining transaction costs between natural gas market centers where only physical trading occurs is left for future studies. It is recommended that future research should be undertaken such that markets are paired based on the information of the major natural gas transportation corridors. Moreover, the development of threshold models in which both threshold values and threshold intervals are time-variant would provide a contribution to the study of transaction costs.

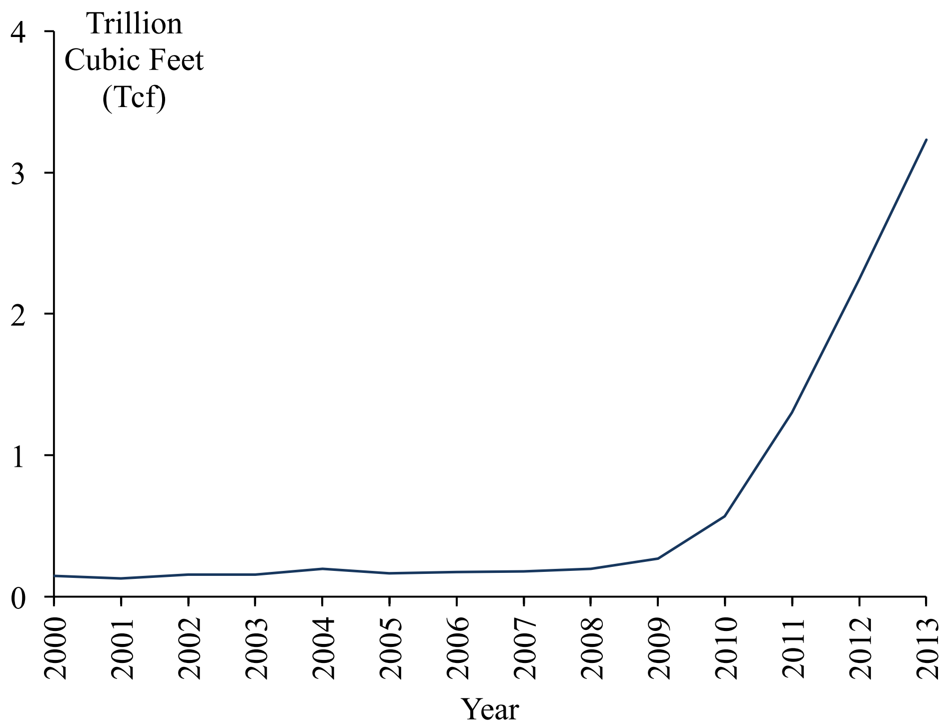

Figure 4.10. Annual Pennsylvania dry natural gas production (Tcf) (2000 – 2013) (U.S. EIA 2015e)

CHAPTER V

NUMBER OF FACTORS EFFECTS ON FACTOR-AUGMENTED VECTOR AUTOREGRESSIVE (FAVAR) PERFORMANCES

Using a vector autoregressive (VAR) approach to study economic issues constrains the ability to understand structural information, because standard VARs usually limit the analysis to approximately eight or fewer variables (Stock and Watson 2002; Bernanke, Boivin, and Eliasz 2005). As such, relevant information may not be reflected in a VAR analysis because of the small number of variables (Bernanke, Boivin, and Eliasz 2005). Incorporating richer information data sets has caught the attention of academics (Sargent and Sims 1997; Stock and Watson 2002; Bernanke and Boivin 2003; Bernanke, Boivin and Eliasz 2005; Moench 2008; Zagaglia 2010). These studies usually assume that variation in economic time series can be captured by a small set of influencing variables; these variables are considered the set of common factors (Sargent and Sims 1997). Bai and Ng (2002) propose several criteria to determine the appropriate number of common factors to include in factor models. In empirical applications, Bai and Ng's (2002) various criteria, however, may lead to differing number of factors; issues of parsimony and appropriateness may arise (Moench 2008; Zagaglia 2010).

The objective is to investigate whether and how the number of unobservable components from a data-rich model influences inferences from and probabilistic forecasting performance of various models. The factor-augmented vector autoregressive (FAVAR) approach, proposed by Bernanke, Boivin, and Eliasz (2005), is employed to characterize unobservable components. Innovation accounting analysis (impulse response functions and forecast error variance decompositions) are applied to discover dynamic effects. Then, the prequential forecasting approach introduced by Dawid (1984) is applied to evaluate predictive distributions for out-of-sample data. Two FAVAR models differing in their number of factors (five and ten factors) based on the rage of optimal number of factors derived from Bai and Ng's (2002) criteria, along with a five variable VAR and a univariate autoregressive (AR) model, are compared.

In keeping with the energy theme of this dissertation, shale gas gross withdrawals are the main variable of interest. Shale gas has noticeably become a "game changer" for the U.S. natural gas market through the use of horizontal drilling and hydraulic fracturing, colloquially known as "fracking". These techniques have prominently increased the capability of producers to commercially recover natural gas and oil from low-permeability geologic formations, mainly shale formations (U.S. EIA 2011). The advent of commercial viable shale gas production began in earnest the 1980s and 1990s when Mitchell Energy and Development Corporation started to produce deep shale gas economically in the Barnett Shale in North Central Texas (U.S. EIA 2011). By 2005, the Barnett shale yielded nearly 0.5 trillion cubic feet of natural gas per year (U.S. EIA 2011). Mitchell Energy and Development achievement induced other companies to explore and produce from other shale sources (U.S. EIA 2011). Increasing shale gas production leads to increased domestic natural gas supply. The rapid growth of domestic natural gas supply resulted in lower natural gas prices. Natural gas, consequently, has become an attractive energy source for electric power generating, industrial, and exporting sectors. Shale gas production may be the "game changer" in not only the U.S. natural gas market but also the entire energy sector of the U.S. Technologies fostering the proliferation of shale gas may affect other industries in the U.S. energy sector. As shale gas production is claimed to be a "game changer," examining its effects on energy dynamics is noteworthy. Unfortunately, because of the limited availability of data on shale gas production, natural gas gross withdrawals are considered as a proxy for shale gas production.

Literature on A Data-Rich Environment

Studies using a data-rich environment suggest that the use of large data sets provides reasonable results and improves forecast precision (Stock and Watson 2002; Bernanke and Boivin 2003; Bernanke, Boivin, and Eliasz 2005; Moench 2008; Zagaglia 2010). Stock and Watson (2002) extract common factors from a large data set using principal components methods. They show that forecasting models which include these common factors outperform univariate autoregressive, traditional vector autoregressive, and leading indicator models. Bernanke and Boivin (2003) employ the factor-model approach developed by Stock and Watson (2002) to estimate and forecast the Fed's policy reaction function. Their findings are in line with Stock and Watson's (2002) results that allowing the systematic information in large data sets to be summarized by a relatively few estimated indicators improve forecasting performance.

Bernanke, Boivin, and Eliasz (2005) propose a FAVAR model in which both unobservable factors and observable economic variables (such as a policy indicator, measures of economic activities, and/or prices) characterize the common forces that determine the dynamics of the macroeconomic economy. They apply the model to measure the effects of monetary policy; exploiting information derived from the FAVAR model significantly increases the ability of identifying the monetary transmission mechanism. Bernanke, Boivin, and Eliasz (2005, p. 406) claim, "The FAVAR approach is successful at extracting pertinent information from a large data set of macroeconomic indicators."

Employing the FAVAR approach in a data-rich environment helps improving forecasting performance (Moench 2008; Zagaglia 2010). Moench (2008) uses the short-term interest rates as policy instrument and factors from of a large number of macroeconomic variables to forecast the yield curve under a no-arbitrage restriction. He finds that most of the variation in interest rates is explained by macroeconomic variables. The no-arbitraging FAVAR model results in an improvement in predicting the yield curve for the out-of-sample over the Duffee (2002) model and the Nelson-Seigel model modified by Diebold and Li (2006). Zagaglia (2010) extracts common factors from a large data set including global macroeconomic indicators, financial market indices, and quantities and prices on energy products to study the dynamics of oil futures prices traded at NYMEX using a FAVAR model. He finds that the estimated factors can be categorized into energy prices, energy quantities, and macroeconomic and financial data. Combining these factors with oil returns improves the forecasting performance of oil futures prices over a VAR model of returns only, a factor-included VAR model, and a random walk model.

Studies using a FAVAR model to evaluate the response of macroeconomic variables to shocks in policy indicators and observable measures of economic activity and prices include Bernanke, Boivin, and Eliasz (2005), Lescaroux and Mignon (2009), and Lombardi, Osbat, and Schnatz (2012). Bernanke, Boivin, and Eliasz (2005) construct the impulse responses for key macroeconomic variables to a monetary policy (federal fund rate) shock using the FAVAR approach. Responding to a negative monetary policy shock, real activity measures decline, prices goes down, money aggregates decline, and the dollar appreciates. Lescaroux and Mignon (2009) apply the FAVAR model to examine the impacts of oil prices on the Chinese economy. They find that an oil price shock induces a contemporaneous increase in consumer and producer price indexes, leading to a rise in interest rates, a delayed negative effect on GDP, investment, and consumption, and a deferred increase in coal and power prices. Lombardi, Osbat, and Schnatz (2012) find that exchanges rates and industrial production influence individual non-energy commodity prices; whereas, robust spillovers from oil to non-oil commodity prices and an oil price impacts on the interest rate are not found.

Methodology

Let Y_t be a $(M\times 1)$ vector of observable variables at time *t* driving the dynamics of the system. Bernanke, Boivin, and Eliasz (2005, p. 391) suggest, "Y_t could contain a policy indicator and observable measures of real activity and prices." Let F_t be a $(K\times 1)$ vector of unobserved factors, which summarize additional information, not fully captured by Y_t, involved in explaining the dynamics of the series of interest (Bernanke, Boivin, and Eliasz 2005). The joint dynamics of F_t and Y_t are

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + \upsilon_t, \tag{5.1}$$

where $\Phi(L)$ is a lag polynomial of finite order d and υ_t is $((K + M) \times 1)$ vector of error terms with zero mean and covariance matrix Q (Bernanke, Boivin, and Eliasz 2005). Equation (5.1) is the FAVAR model (Bernanke, Boivin, and Eliasz 2005).

Because the factors F_t are unobservable, equation (5.1) cannot be directly estimated. Assume that the factors can be inferred from X_t, a $(N \times 1)$ vector of informational time series. The number of informational time series N can be large; in particular, N can be greater than T (number of observations) and N is much greater than the number of factors plus observed variables in the FAVAR system, $N \gg (K + M)$ (Bernanke, Boivin, and Eliasz 2005). The informational time series, X_t, are assumed to be related to the unobservable factors, F_t, and the observed variables, Y_t, by the following equation

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + \varepsilon_t, \tag{5.2}$$

where Λ^f is an $(N \times K)$ matrix of factor loadings, Λ^y is an $(N \times M)$ matrix of parameters, and ε_t is an $(N \times 1)$ vector of error terms with mean zero (Bernanke, Boivin, and Eliasz 2005). Assumptions on the covariance matrix of ε_t dictate on the estimation approach. If ε_t is normal and uncorrelated then maximum likelihood methods are implemented; however, if ε_t is allowed a small level of cross-correlation then a two-step principal components approach is implemented (Bernanke, Boivin, and Eliasz 2005).

Estimation

In line with Stock and Watson 2002; Bernanke and Boivin 2003; Bernanke, Boivin, and Eliasz 2005; Moench 2008; Lescaroux and Mignon 2009; Zagaglia 2010; Lombardi, Osbat, and Schnatz 2012, the two-step principal components approach is applied in estimating the FAVAR model because of its computational simplicity and implemental convenience. In the first step, equation (5.2) is used to estimate the unobservable factors, F_t, by applying principal component approach. In the second step, the FAVAR is estimated by using equation (5.1) with the estimated factors, $\hat{F}_t$, in place of the unobservable factors, F_t.

Common Factors and Number of Common Factors

Common factors in the large data set, X_t, are estimated by the nonparametric method of asymptotic principal components. The number of factors estimated by this method is min{*N, T*}, which is, however, much larger than permitted by estimation of state space models (Bai and Ng 2002). To determine which of these factors are statistically significant, all the common factors must first be consistently estimated when both *N* and *T* are large (Bai and Ng 2002). Let C_t be a $(Q\times 1)$ matrix of common factors extracted from X_t. Because X_t contains dynamic information on both F_t and Y_t, the common factors estimated from X_t are denoted as $C(F_t, Y_t)$. Estimates of common factors, $\hat{C}(F_t, Y_t)$, and factor loading, Λ, are obtained by solving the following optimization problem

(5.3) $V(Q) = \min_{\Lambda, C_t} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it} - \lambda_i^{Q'} C_t\right)^2,$

subject to the normalization of either $\frac{\Lambda^{Q'}\Lambda^{Q}}{N} = I_Q$ or $\frac{C_t^{Q'}C_t^{Q}}{T} = I_Q$ where Λ^{Q} is a (Q x N) matrix containing λ_i^Q, a (Q x 1) vector of factor loadings denoted individually by i and I_Q is a (Q x Q) identity matrix (Bai and Ng 2002).

When all common factors are observed but the factor loadings are not, the problem becomes to choose Q common factors that capture the variations in X_t and estimate the corresponding factor loadings (Bai and Ng 2002). λ_i can be estimated by applying ordinary least squares to each equation as the model is linear and the factors are observed (Bai and Ng 2002). Divided by *NT,* the sum of squared residuals from regressing X_i on the Q common factors for all i becomes

(5.4) $V(Q, C_t) = \min_{\Lambda}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \lambda_i^{Q'}C_t\right)^2.$

The appropriate number of common factors, Q, can be determined using a loss function $V(Q, C_t) + Q\mathrm{g}(N, T)$, where $\mathrm{g}(N, T)$ is the penalty for model over-fitting (Bai and Ng 2002). Let

(5.5) $V(Q, \hat{C}_t) = \min_{\Lambda}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \lambda_i^{Q'}\hat{C}_t\right)^2$

be the sum of square residuals (divided by *NT*) when Q common factors are estimated.

The following 12 criteria to determine the number of factors to include in factor models are proposed in Bai and Ng (2002)

(5.6) $PC_{p1}(Q) = V(Q, \hat{C}_t) + Q\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$

(5.7) $PC_{p2}(Q) = V(Q, \hat{C}_t) + Q\hat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\, C_{NT}^2,$

(5.8) $PC_{p3}(Q) = V(Q, \hat{C}_t) + Q\hat{\sigma}^2\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right),$

(5.9) $IC_{p1}(Q) = \ln V(Q, \widehat{C}_t) + Q(\frac{N+T}{NT})\ln(\frac{NT}{N+T})$,

(5.10) $IC_{p2}(Q) = \ln V(Q, \widehat{C}_t) + Q(\frac{N+T}{NT})\ln\ C_{NT}^2$,

(5.11) $IC_{p3}(Q) = \ln V(Q, \widehat{C}_t) + Q(\frac{\ln C_{NT}^2}{C_{NT}^2})$,

(5.12) $AIC_1(Q) = V(Q, \widehat{C}_t) + Q\hat{\sigma}^2(\frac{2}{T})$,

(5.13) $BIC_1(Q) = V(Q, \widehat{C}_t) + Q\hat{\sigma}^2(\frac{\ln T}{T})$,

(5.14) $AIC_2(Q) = V(Q, \widehat{C}_t) + Q\hat{\sigma}^2 2(\frac{2}{N})$,

(5.15) $BIC_2(Q) = V(Q, \widehat{C}_t) + Q\hat{\sigma}^2(\frac{\ln N}{N})$,

(5.16) $AIC_3(Q) = V(Q, \widehat{C}_t) + Q\hat{\sigma}^2 2(\frac{N+T-Q}{NT})$, and

(5.17) $BIC_3(Q) = V(Q, \widehat{C}_t) + Q\hat{\sigma}^2(\frac{N+T-Q}{NT})\ln(NT)$,

where $V(Q, \widehat{C}_t) = N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2$, $\hat{\sigma}_i^2 = \frac{\hat{e}_i'\hat{e}_i}{T}$, $\hat{e}_i = X_i - \Lambda_i^Q\widehat{C}_t, C_{NT}^2 = \min\{N, T\}$.
PC_p criteria refer to panel criteria, IC_p criteria refer to information criteria, *AIC* refer to Akaike information criteria, and *BIC* refer to Bayesian information criteria (Bai and Ng 2002). The penalty term of PC_p criteria includes $\hat{\sigma}^2$ which provides a proper scaling to the penalty term (Bai and Ng 2002). $\hat{\sigma}^2$ is not required in the penalty term of IC_p criteria because scaling by $\hat{\sigma}^2$ is implicitly applied by the natural logarithm transformed $V(Q, \widehat{C}_t)$ (Bai and Ng 2002). As such, IC_p criteria may be desirable because they do not rely on the choice of maximum number of factors and $\hat{\sigma}^2$ (Bai and Ng 2002).

The PC_p and IC_p criteria differ from the information criteria used in cross-section and time-series analysis in that g(*N, T*) is a function of both *N* and *T*. The penalty terms in AIC_1 and BIC_1 are standard terms used in time-series applications. Two conditions are required to get a consistently estimated number of common factors; first, $\text{g}(N,T) \rightarrow 0$ and second, $C_{NT}^2.\text{g}(N,T) \rightarrow \infty$ as N, $T \rightarrow \infty$ (Bai and Ng 2002). AIC_1 fails to achieve the second condition for all N and T while BIC_1 fails to meet the second condition when $N \ll T$ (Bai and Ng 2002). Similarly, AIC_2 fails in the second condition while BIC_2 works only if $N \ll T$ (Bai and Ng 2002). The penalty term in AIC_3 and BIC_3 involves both N and T. However, AIC_3 violates the second condition while BIC_3 violates the first condition for some N and T (Bai and Ng 2002). Therefore, BIC_3 may perform well under some data structures, for example data with the presence of cross-section correlations (Bai and Ng 2002). PC_{p1}, PC_{p2}, IC_{p1}, IC_{p2}, AIC_3, and BIC_3 are considered in this study.

Removing the Influence of Y_t

Because $\hat{C}(F_t, Y_t)$ corresponds to an arbitrary linear combination of Y_t, obtaining $\hat{F}_t$ requires removing the dependency of $\hat{C}(F_t, Y_t)$ on Y_t. Following Bernanke, Boivin, and Eliasz's (2005), $\hat{C}(F_t, Y_t)$ is regressed on $\hat{F}_t^{slow}$ and Y_t to get the dependency of $\hat{C}(F_t, Y_t)$ on Y_t, where $\hat{F}_t^{slow}$ refers to K (= $Q - M$) factors estimated from slow-moving variables. Slow-moving variables do not to respond contemporaneously to shocks in Y_t while fast-moving variables are allowed to respond contemporaneously to Y_t. The regression to remove the influence of Y_t is

(5.18) $\hat{C}(F_t, Y_t) = \beta^{slow}\hat{F}_t^{slow} + \beta^{Y} Y_t + \eta_t,$

where β^{slow} is a (Q x K) matrix of coefficients of $\hat{F}_t^{slow}$, $\beta^Y$is a (Q x M) matrix of coefficients indicating the dependency of $\hat{C}(F_t, Y_t)$ on Y_t, and η_t is a (Q x 1) vector of error terms. Using $\hat{\beta}^Y$ estimated from equation (5.18), $\hat{F}_t$ is derived by

(5.19) $\hat{F}_t = \hat{C}(F_t, Y_t) - \hat{\beta}^Y Y_t$.

The FAVAR model (equation 5.1) is estimated using $\hat{F}_t$ derived from equation (5.19) and the observable variable, Y_t.

Data

One hundred and seventy-nine monthly series of energy and macroeconomic data are used. Natural gas gross withdrawals are treated as the observable variable, Y_t. The informational vector, X_t, includes the remaining 178 variables. The natural logarithm transformation is applied to all series except those in percentages before any estimation. Based on Augmented Dickey Fuller test results, series that are non-stationary (have a unit root) are transformed such that they become stationary before estimating the FAVAR.

Monthly series, the classification of variables into the slow- and fast- moving variables, their transformation (if necessary to make stationary), and data sources are presented in Appendix C. The FAVAR model is fitted over the period of February 2001 to December 2012. Out-of-sample data period for one step-ahead forecasts is January 2013 to December 2014.

Empirical Results

Because of the high variation in the variables, the data are demeaned and standardized to have zero mean and unit variance. As such, a small number of factors can represent the dynamics of the 178 informational time series (Bai and Ng 2002). Common factors are estimated using principal component approach and the number of common factors are determined based on PC_{p1}, PC_{p2}, IC_{p1}, IC_{p2}, AIC_3, and BIC_3 criteria.

Allowing the maximum number of common factors equals to 20, PC_{p1}, PC_{p2}, and AIC_3 all suggest 20 common factors. IC_{p1} suggests 13 common factors, IC_{p2} suggests 10, and BIC_3 suggests five common factors (table 5.1). For parsimony and comparison purposes, FAVAR models with five and 10 factors are investigated in this study. In addition, as noted earlier a five variable VAR model and a univariate AR model are also compared.

Table 5.1. Test Results of Numbers of Factors Suggested by Several Criteria

Criteria	Number of Factors
PC_{p1}	20
PC_{p2}	20
IC_{p1}	13
IC_{p2}	10
AIC_3	20
BIC_3	5

Structural Break in Factors

There appears to be potential structural changes in pricing relationship among North America natural gas spot market (see Chapter II). As such, the possible presence of structural change in the U.S. energy dynamics, represented by the common factors obtained from the principal components, is tested for structural breaks in factor loadings using tests proposed by Han and Inoue (2015). Han and Inoue's (2015) tests are constructed under the joint null hypothesis that all factor loadings are constant over time against the alternative that at least one of the factor loadings is not constant over time. The tests utilize the second moment of estimated factors, rather than a simple regression-based approach, to avoid dimensionality problems. Han and Inoue's (2015) tests use more information than other tests as they directly test for the differences between before and after the break in all elements of the covariance matrix of the estimated factors. The idea of these tests is that a structural break in factor loading also appears in second moments of factors estimated from the full sample principal components. Results of Han and Inoue's (2015) tests indicate that there is no structural break in the common factors over the period of February 2001 to December 2012; the joint null hypothesis that all factor loadings are constant over time cannot be rejected.

Several reasons are postulated why the potential existence of structural breaks is found in Chapter II but not in the common factors. First, the frequency of the data may matter in detecting structural breaks. The Potential breaks are found in Chapter II using the high frequency daily data, whereas no structural breaks are found in the common factors which use relatively low frequency monthly data. Second, the relationships among variables between the two studies are different. The study in Chapter II focuses only on the pricing relationship in the natural gas sector, while the common factors examine broader relationships. Not only are energy prices considered but also energy activities such as production and consumption. Further, besides natural gas, other energy sources such as coal, crude oil and other petroleum products, electricity, renewable energy along with several macroeconomic variables are included. Finally, the methodologies differ. Tests applied in Chapter II are the tests for parameter instability in long-run relationship parameters, while tests applied in this chapter is the tests for parameter instability in estimated factors.

Interpreting the Estimated Factors

Five and 10 common factors, $\hat{C}(F_t, Y_t)$, are used in equations (5.18) and (5.19) to obtain the estimated factors, $\widehat{F_t}$. To estimate equation (5.18), common factors extracted from slow-moving variables, $\widehat{F_t}^{slow}$, are required. Because the dimension of Y_t equals one (*M* =1) (only natural gas gross withdrawals is treated as an observable variable), to obtain five and 10 (*Q* = 5, 10) estimated factors, four and nine slow-moving factors (*K* = *Q* – *M* = 4, 9) are necessary. After removing the influence of Y_t from the common factors, $\hat{C}(F_t, Y_t)$, the five estimated factors, $\widehat{F_t}$, from the system with five common factors are not the same as the first five estimated factors from the system with 10 common factors. They are not the same because of the difference in the number of slow-moving factors.

To interpret the estimated factors, each factor is regressed individually on all the variables in X_t. R-squared values from regressing each of the five factors on the 178 variables are illustrated in figures 5.1 – 5.5. R-squared values for the 10 factors are presented in figures 5.6 – 5.15. To help in the interpretation, the variables are classified into 20 categories. The variables in each category are presented in Appendix C.

It appears that each estimated factor is better represented by a group/groups of variables than an individual variable. For most of the factors, many of the 178 variables contribute to explaining that factor. Factor 1 of the five factor model is a prime example of this issue (figure 5.1). Over one-half of the categories have at least one variable with an R-squared greater than 0.30. The following discussion is based strictly on observations of figures 5.1-5.15. To be considered as representing a factor, the category as a group should have relatively high R-squared and not just one or two of the variables within that group having high R-squared. Factor 1 of the five factors appears to represent a mixture of carbon dioxide emissions, crude oil and petroleum product consumption and production, electricity net generation, electricity prices, natural gas prices and storages, and renewable energy consumption. The R-squared values from regressing factor 1 on most of the variables in these categories are relatively high. Factor 2 of the five factor system is apparently explained by crude oil and petroleum product prices. Factor 3 from the system with the five factors seems to be the combination of crude oil and petroleum product prices, electricity consumption, electricity net generation, electricity prices, and natural gas consumption and prices. Electricity prices, natural gas consumption and prices, and renewable energy consumption and production appear to explain factor 4. Factor 5 appears to be mostly explained by natural gas consumption and prices, renewable energy consumption and production, and weather.

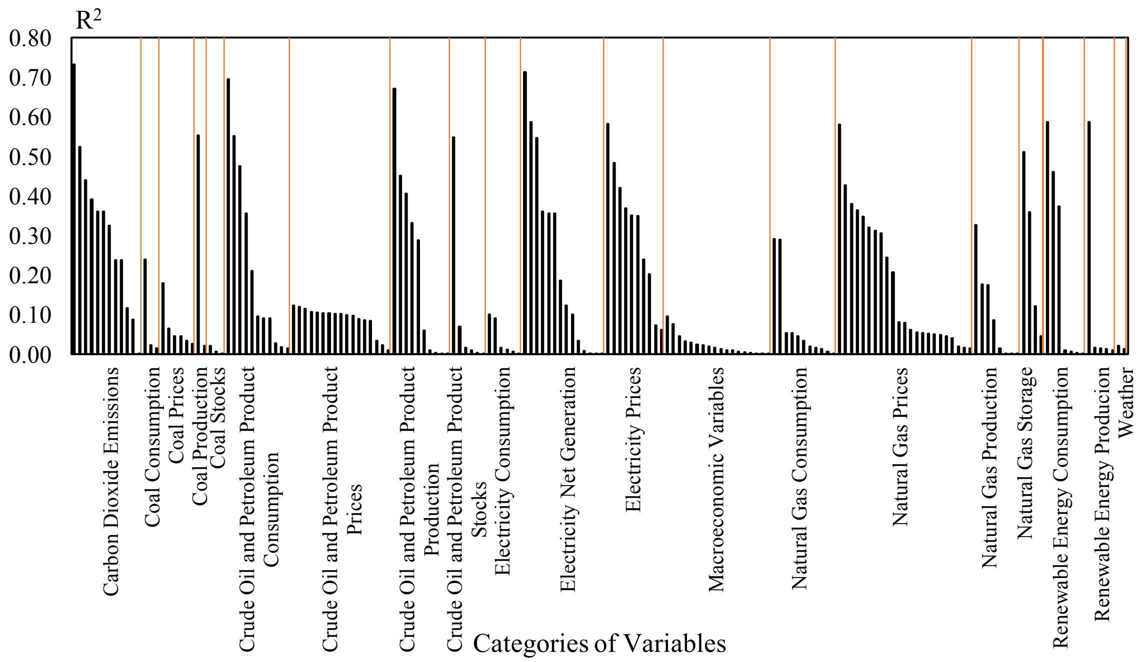

Figure 5.1. R-squared values from regressing factor 1 of the five factors on each of 178 variables

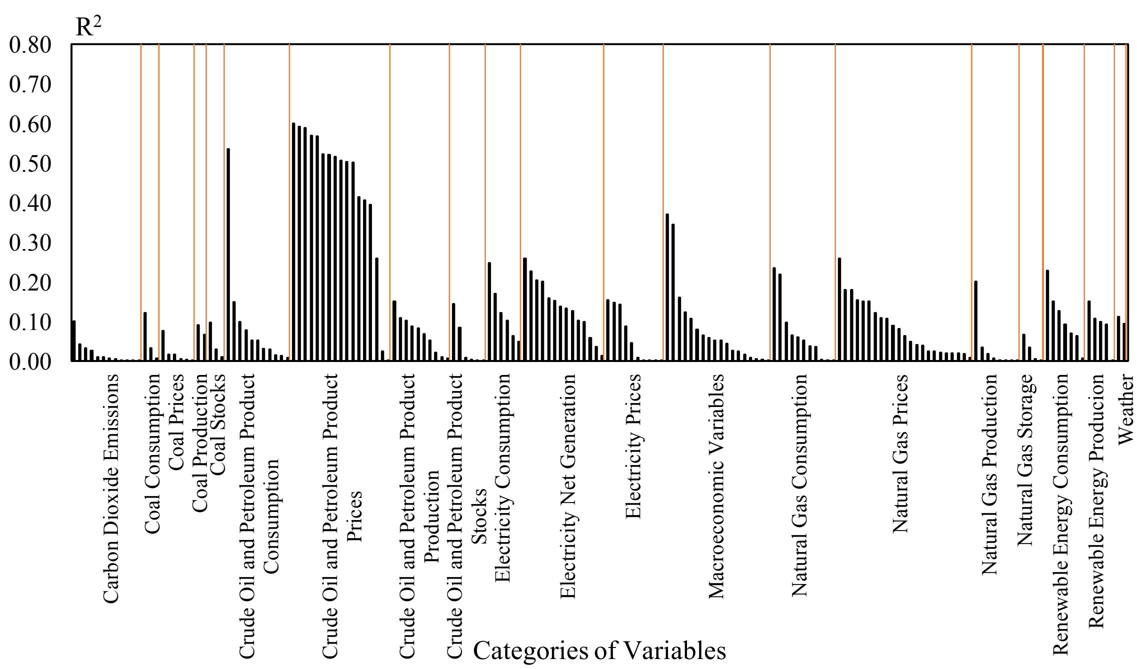

Figure 5.2. R-squared values from regressing factor 2 of the five factors on each of 178 variables

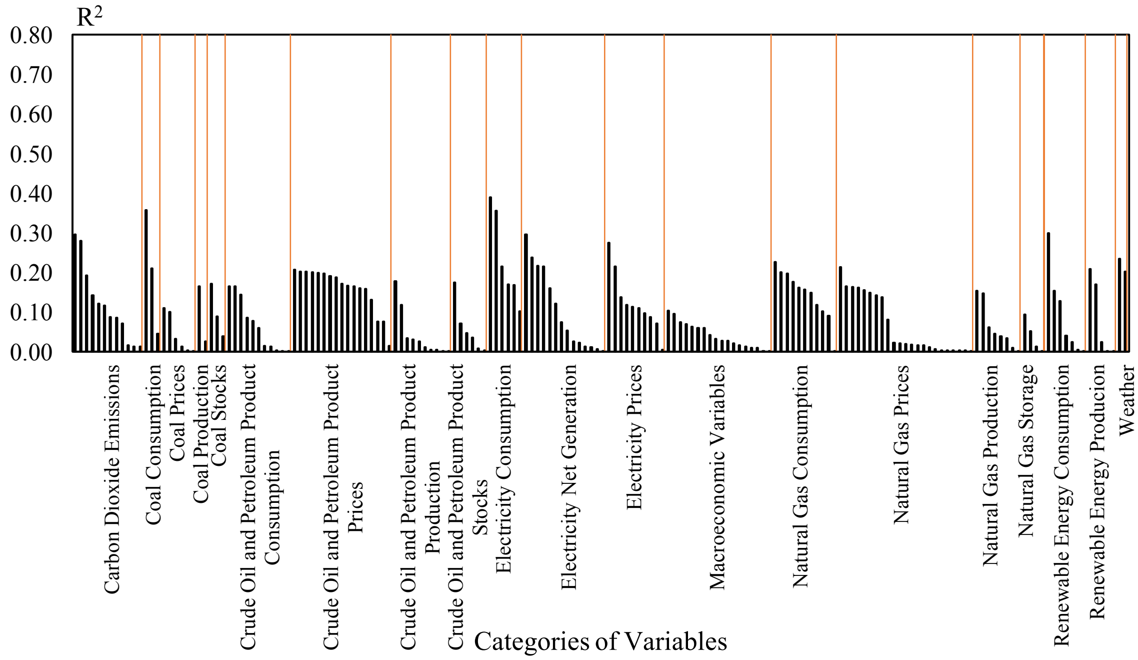

Figure 5.3. R-squared values from regressing factor 3 of the five factors on each of 178 variables

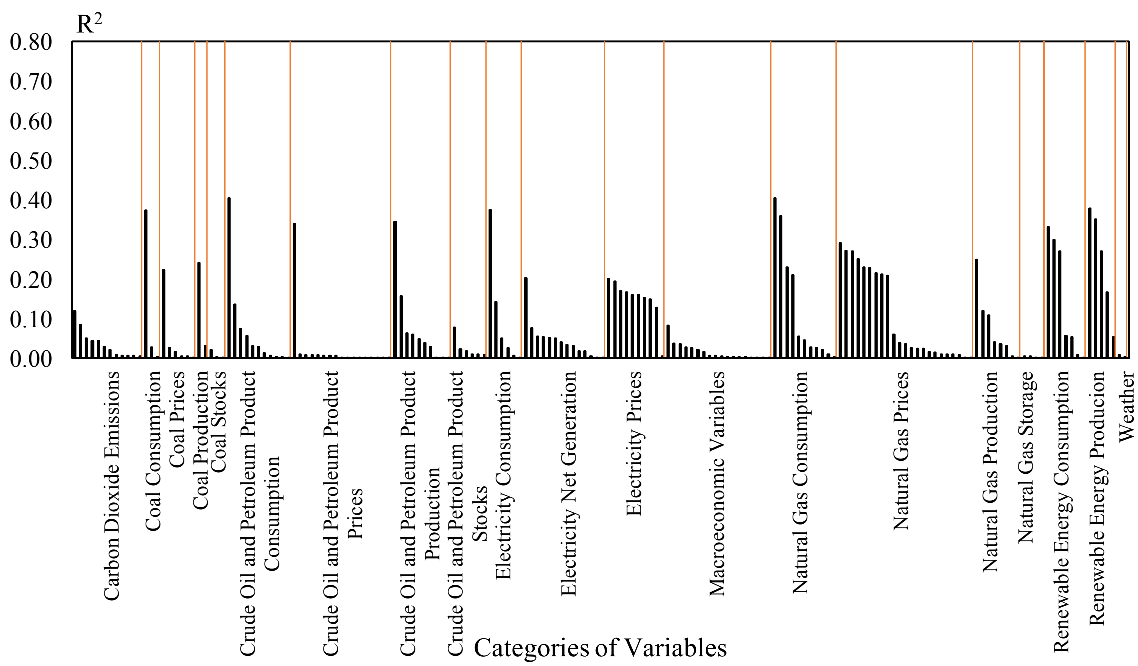

Figure 5.4. R-squared values from regressing factor 4 of the five factors on each of 178 variables

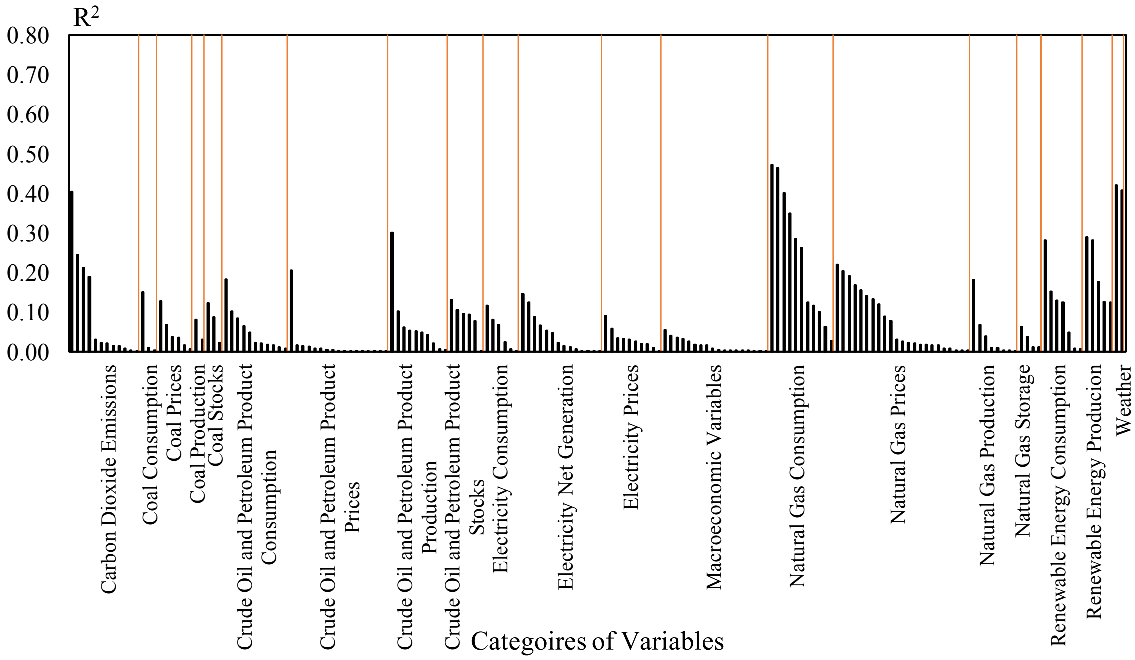

Figure 5.5. R-squared values from regressing factor 5 of the five factors on each of 178 variables

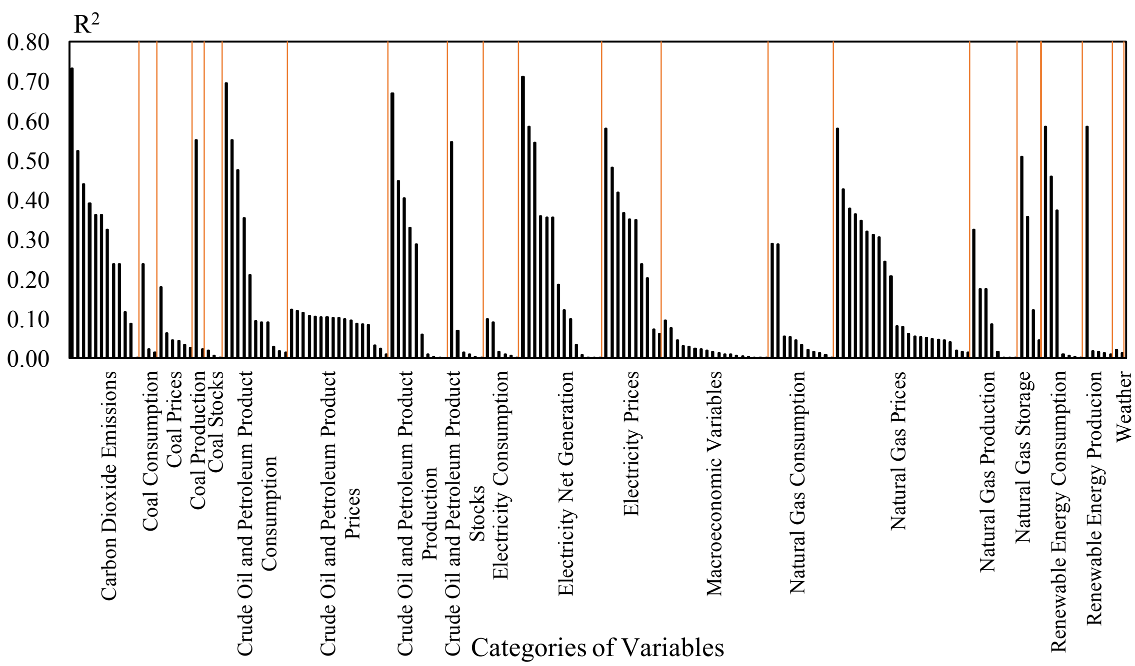

Figure 5.6. R-squared values from regressing factor 1 of the 10 factors on each of 178 variables

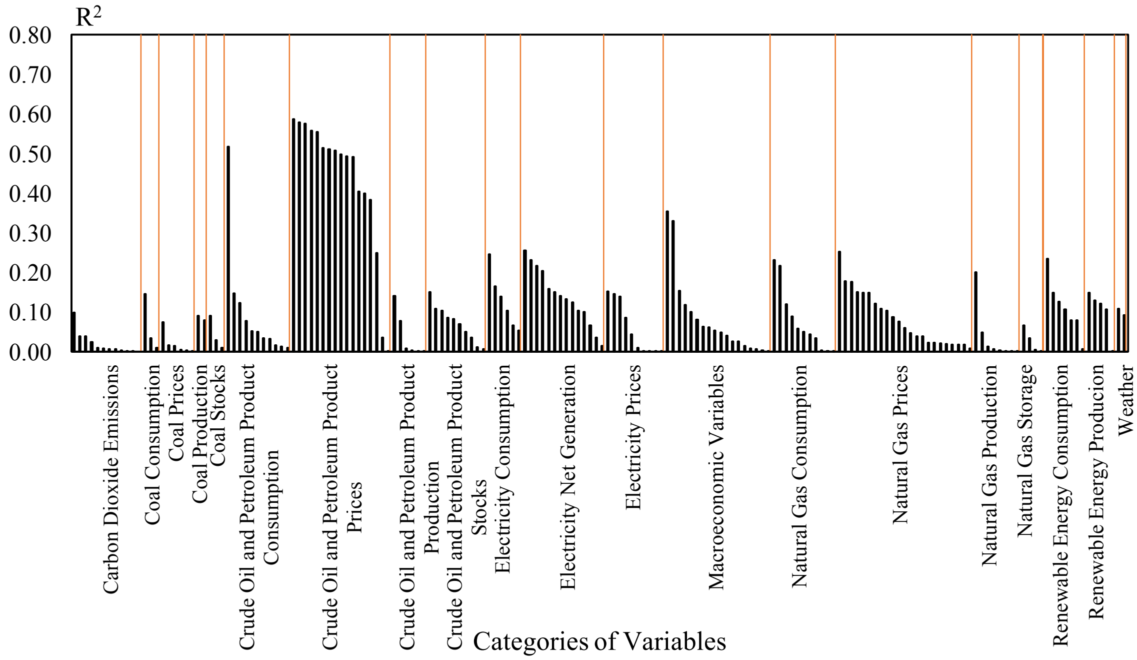

Figure 5.7. R-squared values from regressing factor 2 of the 10 factors on each of 178 variables

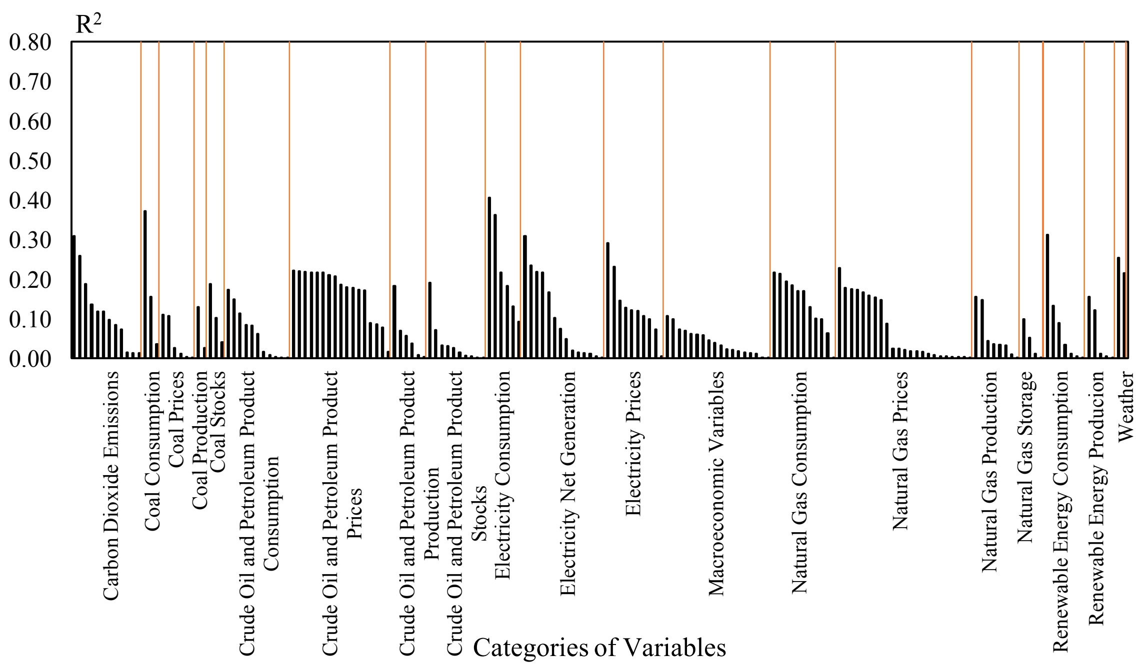

Figure 5.8. R-squared values from regressing factor 3 of the 10 factors on each of 178 variables

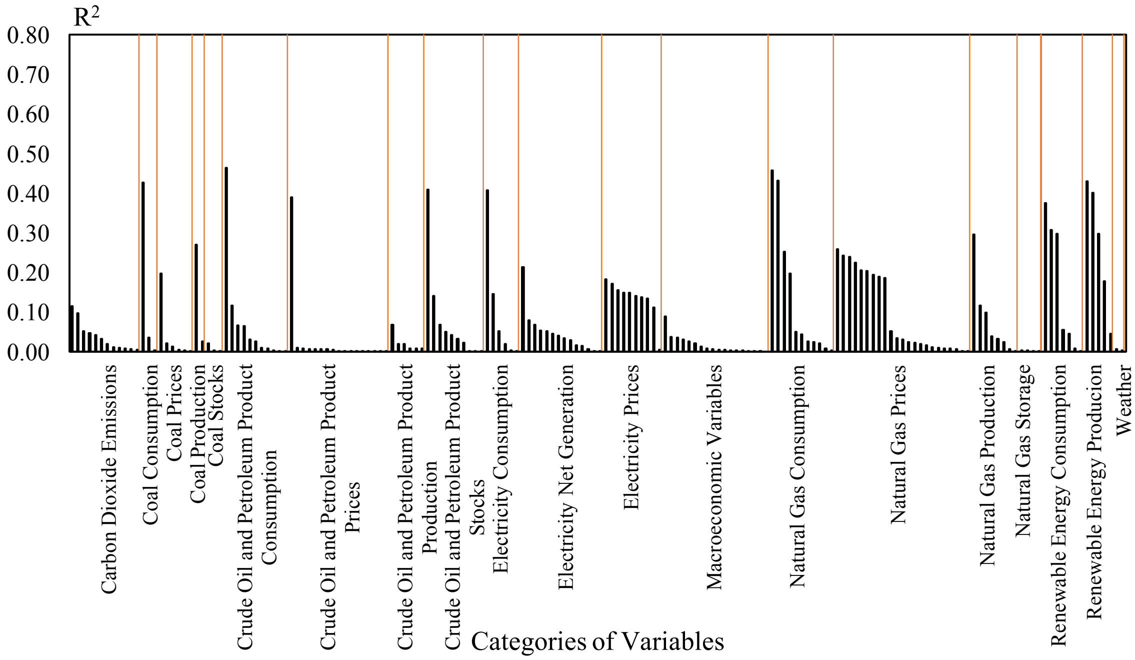

Figure 5.9. R-squared values from regressing factor 4 of the 10 factors on each of 178 variables

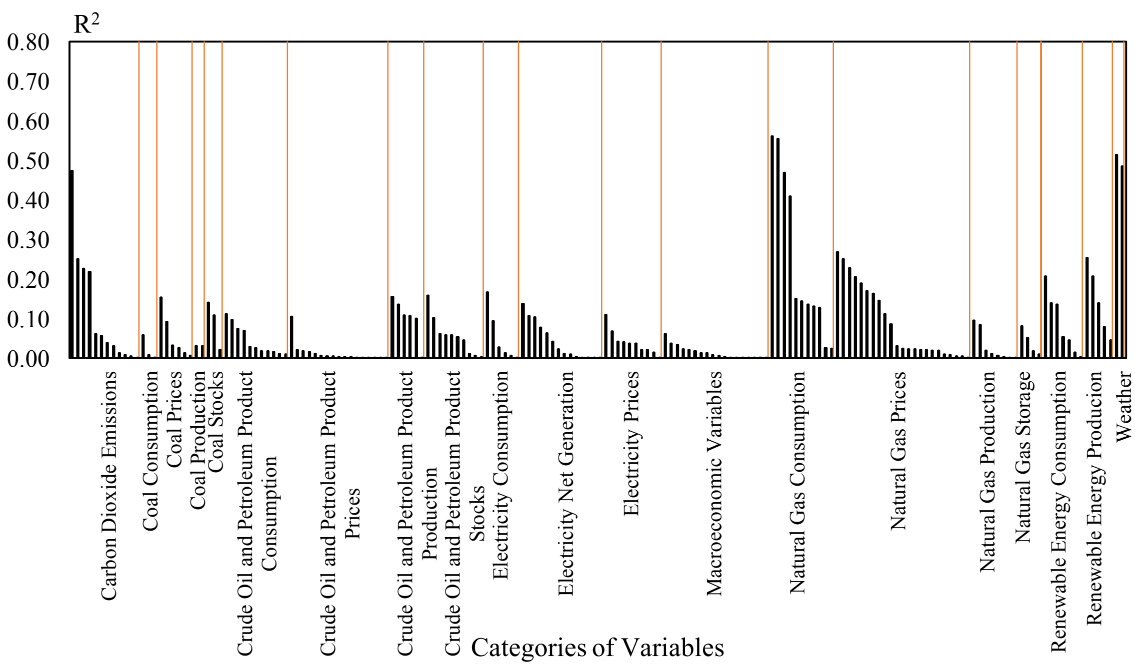

Figure 5.10. R-squared values from regressing factor 5 of the 10 factors on each of 178 variables

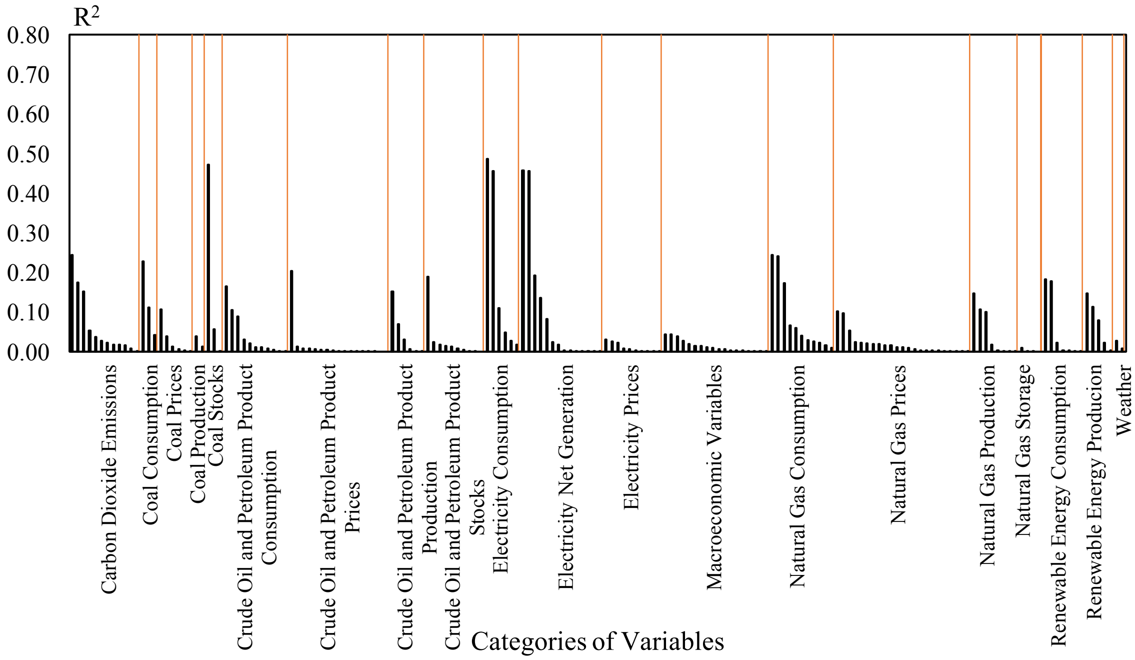

Figure 5.11. R-squared values from regressing factor 6 of the 10 factors on each of 178 variables

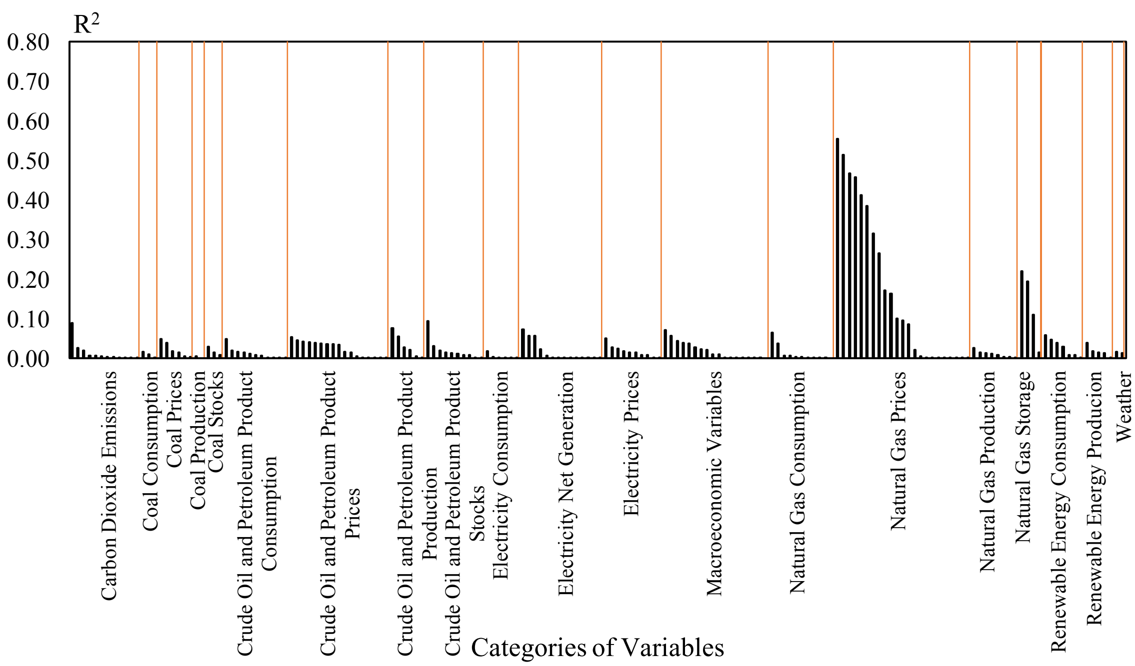

Figure 5.12. R-squared values from regressing factor 7 of the 10 factors on each of 178 variables

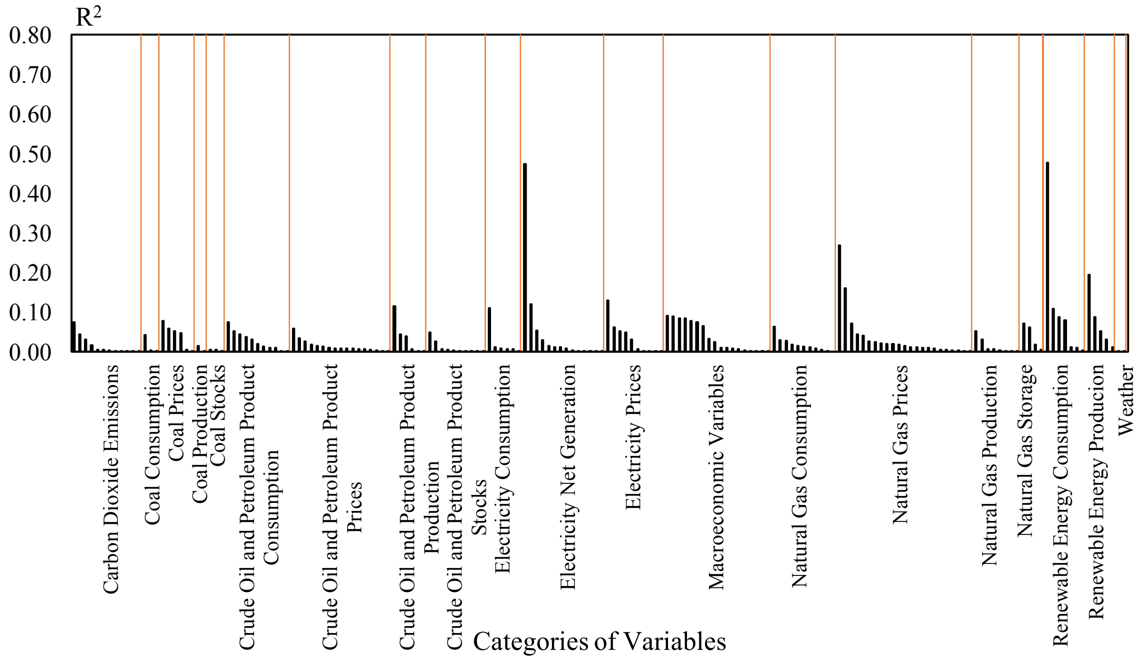

Figure 5.13. R-squared values from regressing factor 8 of the 10 factors on each of 178 variables

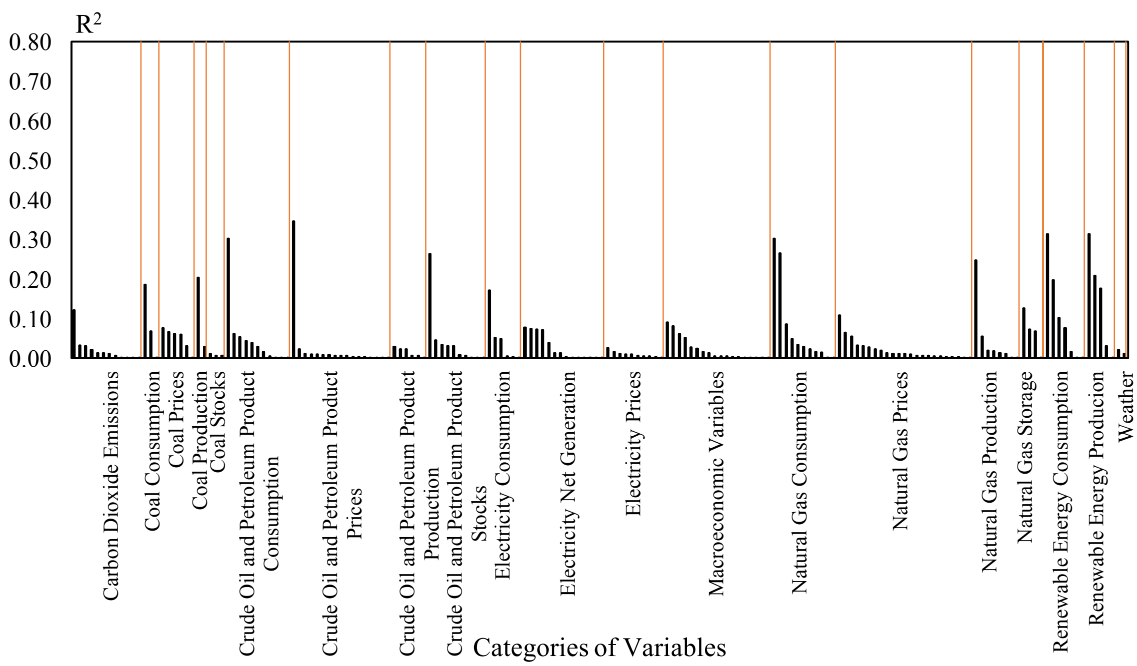

Figure 5.14. R-squared values from regressing factor 9 of the 10 factors on each of 178 variables

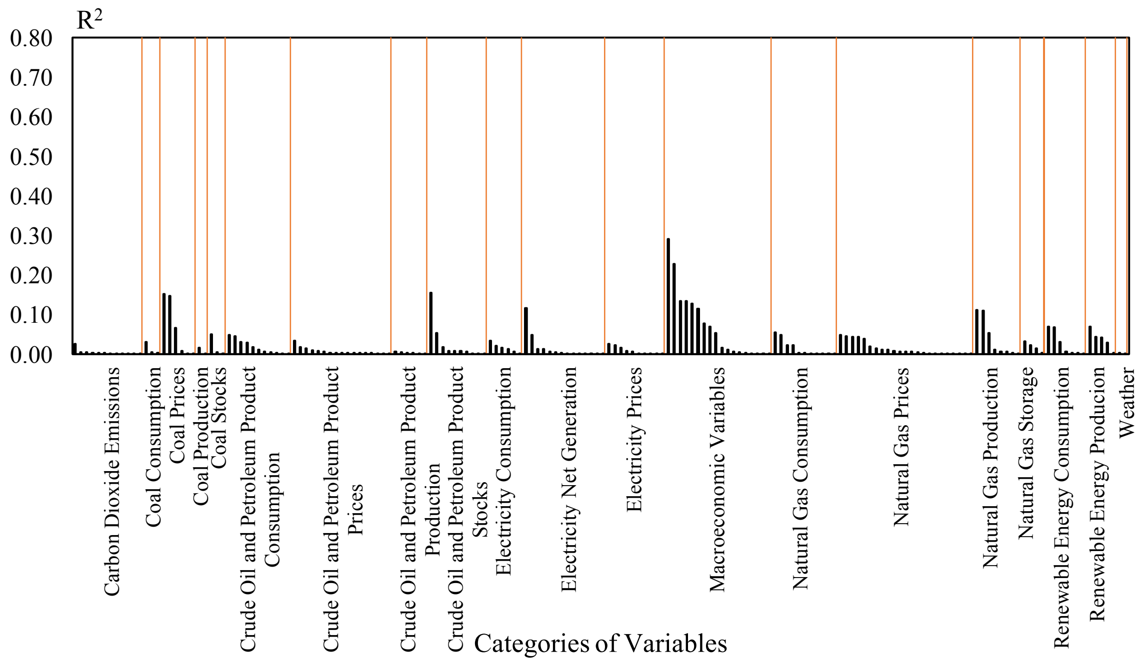

Figure 5.15. R-squared values from regressing factor 10 of the 10 factors on each of 178 variables

Even though they are derived from slightly different system of equations (5.18) and (5.19), the first five factors of the 10 factor system appear to represent the same energy categories as of the five factors of the five factor system. Factor 6 of the 10 factors likely represents coal stocks, electricity consumption, and electricity net generation. Factor 7 appears to represent natural gas prices. Most R-squared values from the regressions on the eighth factor are relatively low; relatively high R-squared values are from regressing factor 8 on some variables of electricity net generation, natural gas prices, and renewable energy consumption. Similar to factor 8, most R-squared values from each regression of factor 9 are relatively low. Compared to other variables in other categories, R-squared values from regressing factor 9 on most variables of natural consumption and renewable energy consumption and production are relatively high. Factor 10 appears to be explained by macroeconomic variables.

FAVAR Models vs VAR model

Two FAVAR models are constructed; one includes five estimated factors and the another includes 10 estimated factors, with natural gas gross withdrawals as the observable variable (hereafter FAVAR(5F) and FAVAR(10F)). To examine whether the data-rich FAVAR models are more advantageous than a VAR model, a VAR model with five variables VAR(5)) of natural gas gross withdrawals, natural gas consumption, Henry Hub natural gas spot price, West Texas Intermediate (WTI) crude oil spot price, and the S&P500 index are constructed.

Given its formula in Chapter II, Schwarz loss measure is used to determine lag length in each model. The Schwarz loss criteria suggest two lags are appropriate for each model (table 5.2).

Table 5.2. Schwarz Loss Measures on One to Five Lags on Each Model

Lag Length	FAVAR(5F)[a]	FAVAR(10F)[b]	VAR(5)[c]
1	-33.5611	-63.4724	-0.8546
2	-34.7067*	-63.7792*	-1.1147*
3	-34.5484	-61.6724	-0.5862
4	-34.3403	-59.3259	-0.2986
5	-34.0043	-56.7813	0.0950

Note: The asterisk '*' indicates minimum values of Schwarz loss measure.
[a] The FAVAR(5F) model includes five estimated factors and natural gas gross withdrawals as the observable variable.
[b] The FAVAR(10F) model includes 10 estimated factors and natural gas gross withdrawals as the observable variable.
[c] The VAR(5) model includes natural gas gross withdrawals, Henry Hub natural gas spot price, natural gas consumption, WTI crude oil spot price, and S&P500.

Contemporaneous Causal Flows

In equation (5.1), the innovation terms, υ_t, are assumed to be independent but contemporaneous correlations among the elements are allowed. If the elements of innovation term are contemporaneously uncorrelated, then innovation accounting procedures can be performed using the moving average representation of the estimated FAVAR/VAR (Hamilton 1994). Nevertheless, contemporaneous correlations usually exist in economic data. Following Bernanke (1986) to obtain contemporaneously uncorrelated innovations, the observed innovations, υ_t, are modeled as a function of more fundamental driving sources of variation, e_t, which are independent (orthogonal) to other sources of variation

(5.8) $\upsilon_t = A^{-1}e_t$,

where A is a matrix representing how each non-orthogonal innovation is caused by the orthogonal variation in each equation (Bernanke 1986). Usual innovation accounting procedure can be preformed by pre-multiplying the FAVAR/VAR models by A. To obtain an identified model, zero restrictions on A are identified using directed acyclic graphs (DAGs) on innovations from the estimated models.

Jarque-Berra normality tests on residual series from the three models' estimation suggest that all residual series are normal. The greedy equivalency search (GES) algorithm (Chickering 2002, 2003) executed in Tetrad version five is applied to the residual series from the estimated FAVAR/VAR models to create the orthogonal innovations for innovation accounting analysis (impulse response functions and forecast error variance decompositions). DAGs estimated from the FAVAR(5F), FAVAR(10F), and VAR(5) models are presented in figures 5.16 – 5.18.

The FAVAR(5F) and FAVAR(10F) models have one common information flow; information flow from natural gas gross withdrawals to factor 4. Factor 4 contemporaneously responds to natural gas gross withdrawals in the FAVAR(5F) and FAVAR(10F) models. In the FAVAR(5F), natural gas gross withdrawals are not only an information provider but also an information receiver; it receives information from factor 5. In the FAVAR(10F) model, natural gas gross withdrawals behave only as an information sender; sending information to factors 4 and 9.

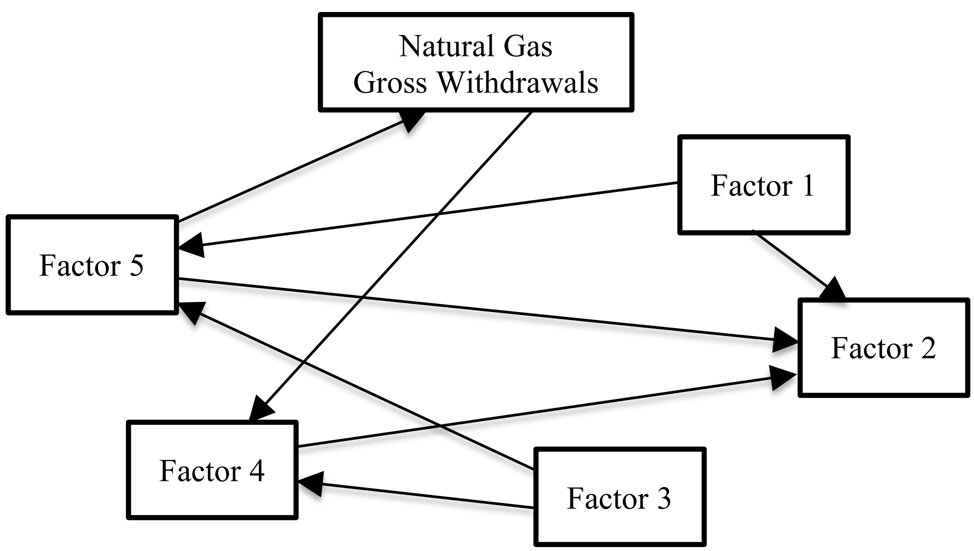

Figure 5.16. Contemporaneous casual flows of the residual series estimated from the FAVAR(5F) model

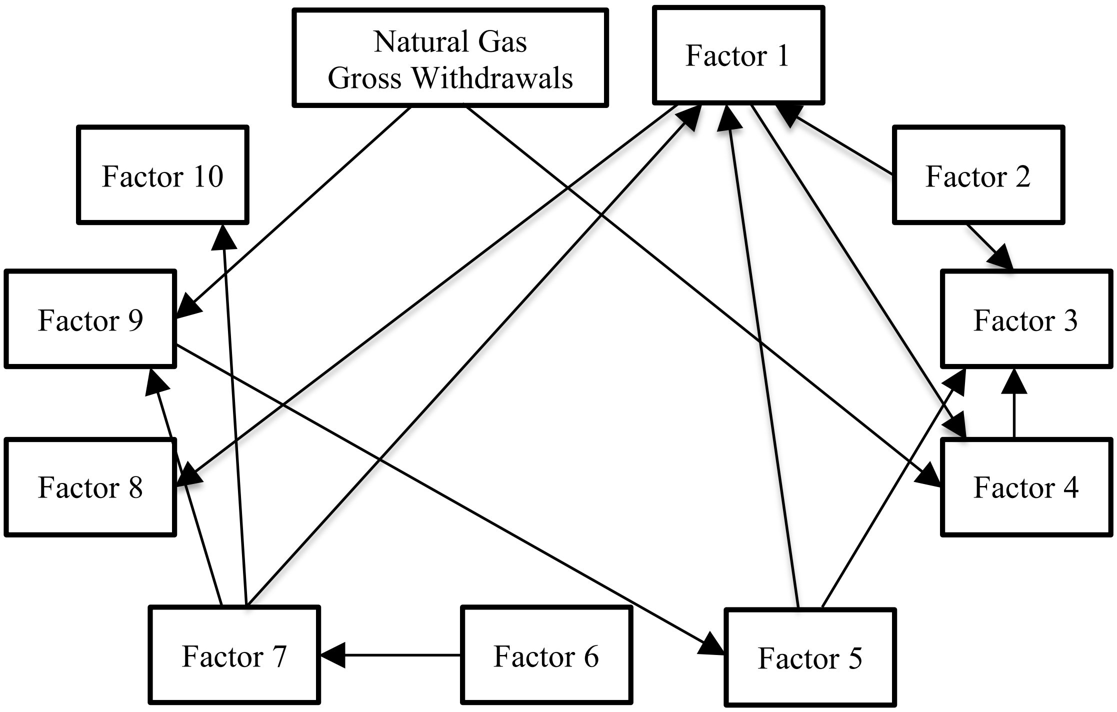

Figure 5.17. Contemporaneous casual flows of the residual series estimated from the FAVAR(10F) model

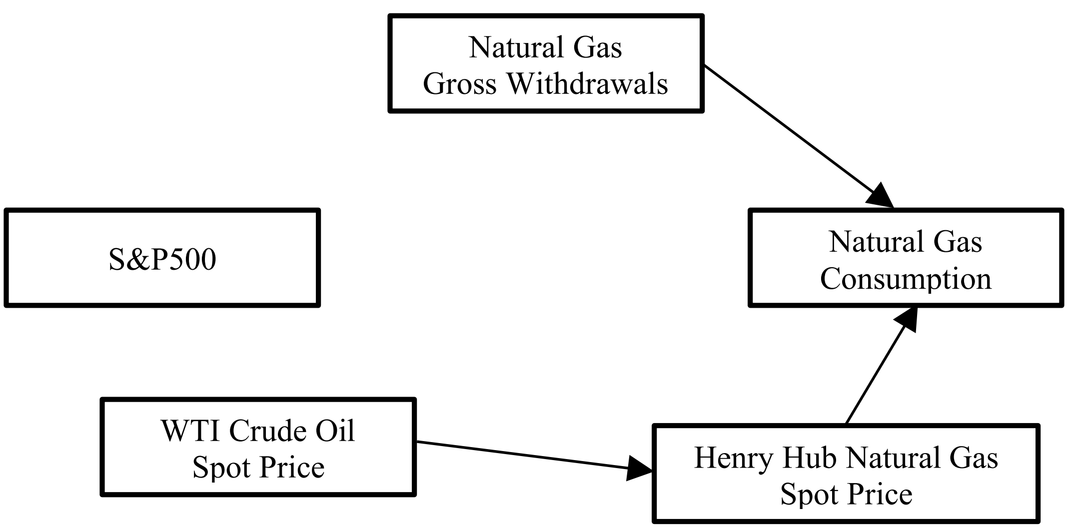

Figure 15. Figure 5.18. Contemporaneous casual flows of the residual series estimated from the VAR(5) model

In the FAVAR(5F) model, factor 1 transmits information to factors 2 and 5. In addition to receiving information from factor 1, factor 2 contemporaneously responds to information from factors 4 and 5. As a transmitter, factor 3 passes information to factors 4 and 5.

In the FAVAR(10F) model, factor 1 gathers information from factors 2, 5, and 7 and passes information to factors 4 and 8. As an information supplier, factor 2 sends information to factors 1 and 3. Factor 3 is an information sink; it takes information from factors 2, 4, and 5 but do not transmit any information. Factor 5 receives information from factor 9 and passes information to factors 1 and 3. Obtaining information from factor 6, factor 7 sends information to factors 1, 9, and 10.

In the VAR(5) model, natural gas gross withdrawals do not receive any information from other variables, but provide information to natural gas consumption. Natural gas consumption contemporaneously responds to information from not only natural gas gross withdrawals but also Henry Hub natural gas spot price. Information flows from WTI crude oil spot price to Henry Hub natural gas spot price. S&P500 is exogenous in contemporaneous time, no information transferring between S&P500 and others.

Impulse Response Functions

Impulse response functions provide the dynamic responses of each series to a one-time shock in each series. For comparision purposes, the responses are normalized such that each response is divided by the standard error of its innovations. Impulse response functions are presented for 12 months. Each sub-graph provides the response of the market given by the row heading to a one-time shock in the series listed in the column heading.

Impulse response functions of the FAVAR(5F) model are presented in figure 5.19. Of primary importance is not only how natural gas withdrawals respond to shocks in the other variables and how the other variables respond to a shock in withdrawals but also differences between the models. Factors 1, 3, and 5 do not respond, while factor 2 has a negative response and factor 4 has a positive response to a shock in natural gas gross withdrawals in the first month. Factors 1, 2, and 5 positively respond to a shock in natural gas gross withdrawals in the second month, whereas factors 2 and 4 negatively respond. All impulse response functions are stable tending toward zero as the number of months out increases.

Natural gas gross withdrawals initially respond negatively to shocks in factors 1 and 5, but the responses become positive in the second month. The response to a shock in factor 1 is relatively larger than that to a shock in factor 5. The response of natural gas gross withdrawals to a shock in factor 3 are similar to the response to a shock in factor 4, but the response to the shock in factor 3 are relatively larger. Natural gas gross withdrawals barely respond to a shock in factor 2.

Dynamic responses of the first five factors to a shock in natural gas gross withdrawals in the FAVAR(10F) model (figure 5.20) generally have similar patterns to those of the FAVAR(5F) model. Responses of factors 7, 8, and 10 to a shock in natural gas gross withdrawals are relatively small. A month after the shock, a response of factor 9 to a shock in natural gas gross withdrawals is negative, whereas two months after the

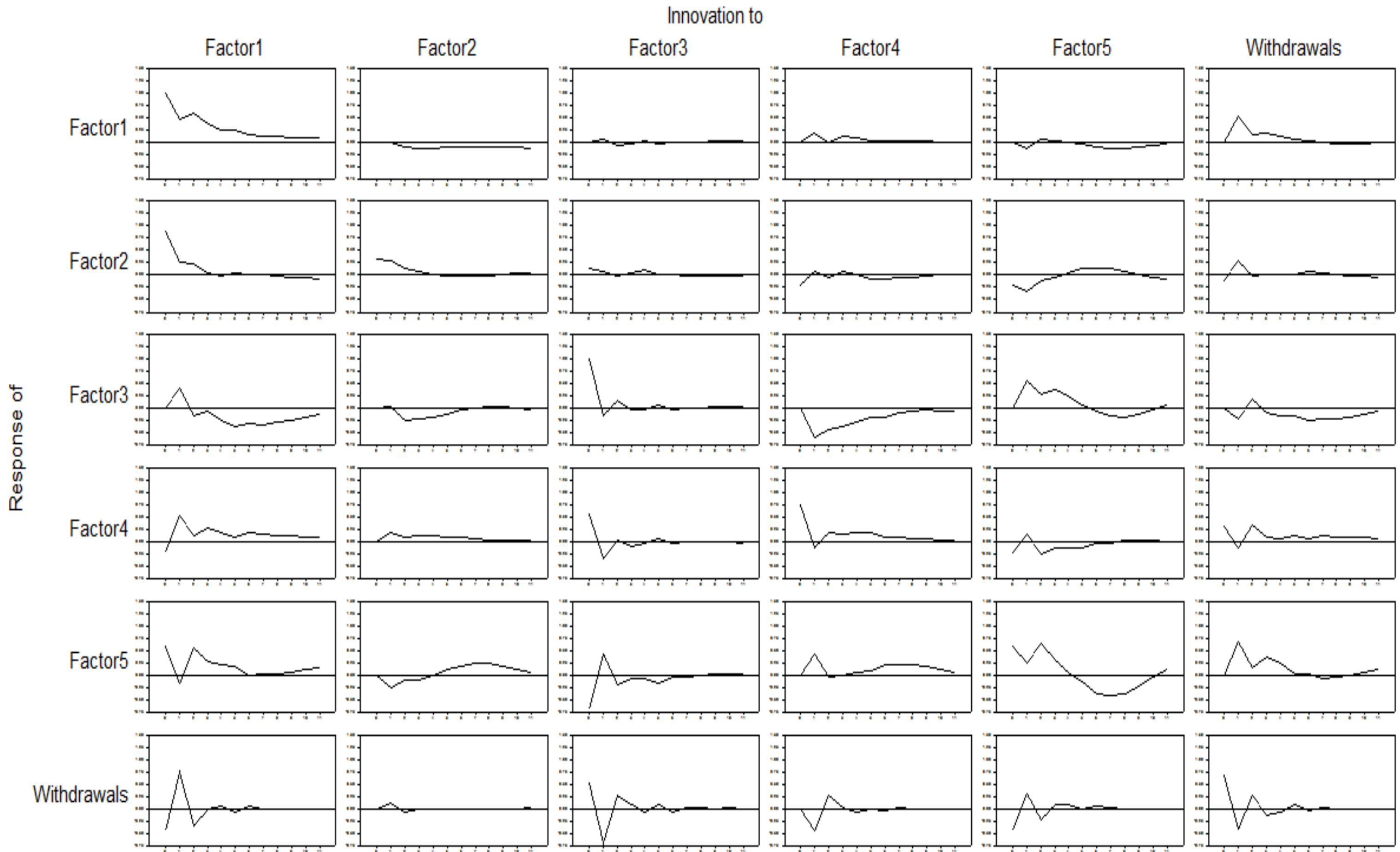

Figure 5.19. Impulse response functions of five factors and natural gas gross withdrawals from the FAVAR(5F) model

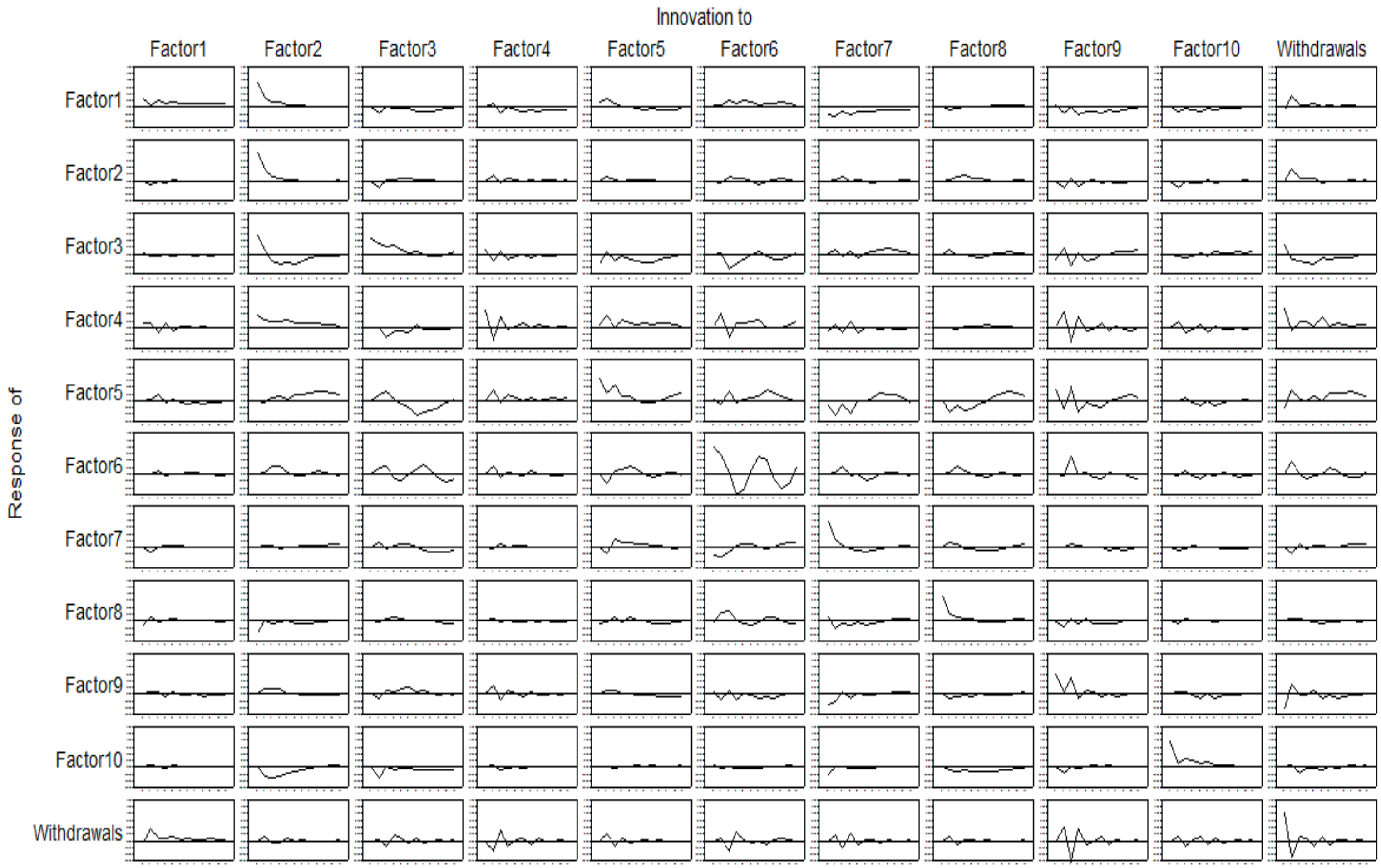

Figure 5.20. Impulse response functions of ten factors and natural gas gross withdrawals from the FAVAR(10F) model

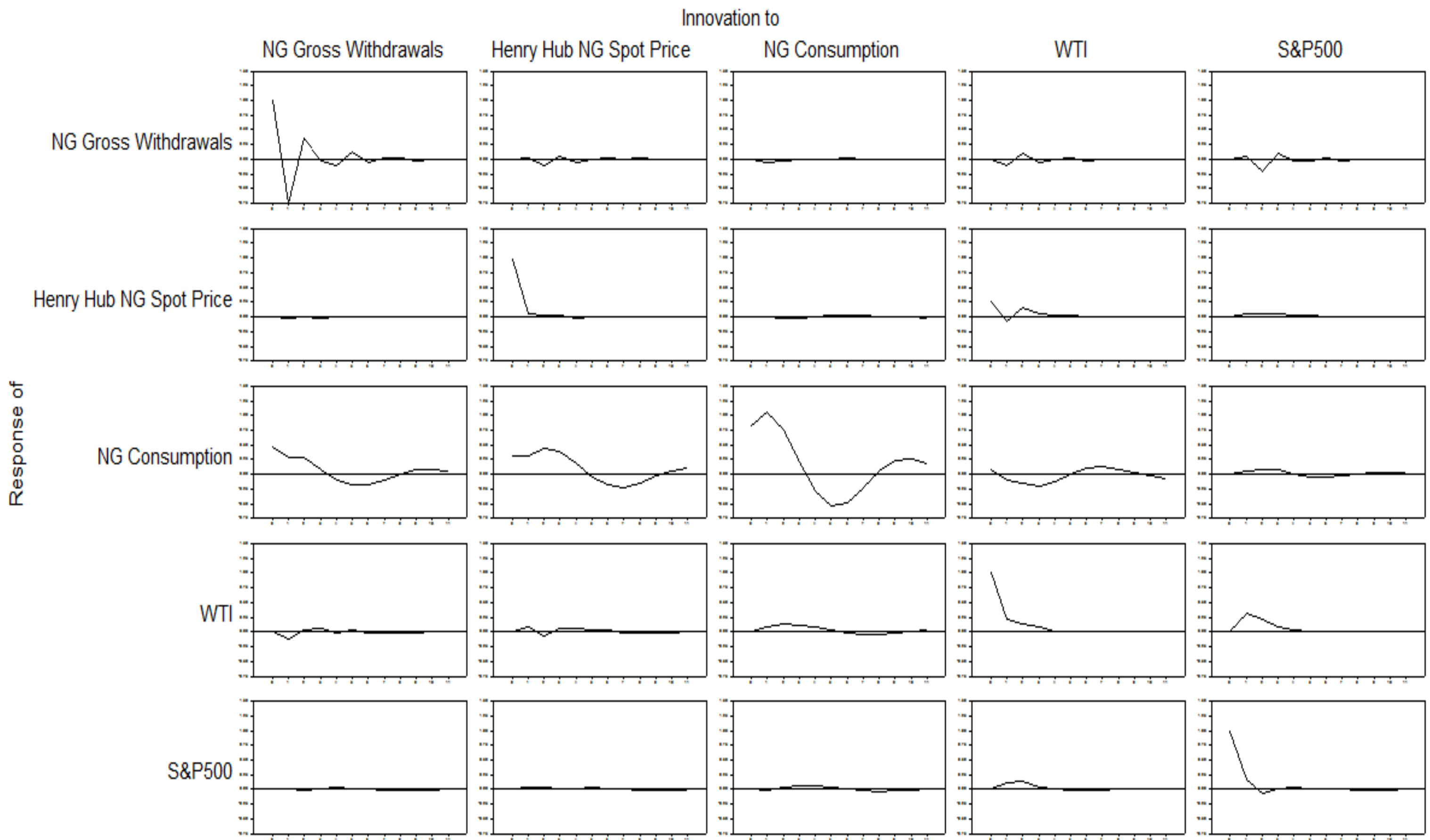

Figure 5.21. Impulse response functions of natural gas gross withdrawals, Henry Hub natural gas spot price, natural gas consumption, WTI crude oil spot price, and S&P500 from the VAR(5) model

shock, the response is positive. Responses of natural gas gross withdrawals to shocks in the first five factors of the FAVAR(10F) model are inconsiderably different to those of the FAVAR(5F). Natural gas gross withdrawals responses to shocks in factors 6 and 7 have similar patterns to that of factor 4. Natural gas gross withdrawals appear to have large responses to shocks in factors 4, 6, and 9, and itself.

Impulse response functions of the VAR(5) model are presented in figure 5.21. Natural gas gross withdrawals have small to no responses to shocks in the other four variables. Other than itself, Henry Hub natural gas price responds to only a shock in WTI. Natural gas consumption has relatively positive large responses to shocks in Henry Hub spot price and natural gas gross withdrawals. Responses of natural gas consumption to a shock in WTI oscillate between positive and negative. Natural gas consumption barely responds to a shock in S&P500. Responses of WTI crude oil spot price to a shock in S&P500 are larger than to shocks in other variables. S&P500 appears not to respond to shocks in the other variables.

Forecast Error Variance Decompositions

Forecast error variance decompositions at horizons of one, six, and 12 months ahead are presented to observe how each series depends on its own innovations and other series' innovations. Values in each row indicate, at each time horizon, how much variation in each series is due to itself and the other price series; the sum of the values in each row must be 100.

Forecast error variance decompositions from the FAVAR(5F) model are presented in table 5.3. Innovations in natural gas gross withdrawals do not influence the forecast error variances of the factors at one month ahead except for factor 4. Natural gas gross withdrawals partially influence forecast error variances of all the factors at six and 12 months ahead, explaining between four and 20 percent of the variances. At one month ahead, uncertainty in natural gas gross withdrawals is explained by itself and factors 1, 3, and 5. At six and 12 months ahead, over 20% of the uncertainties in natural gas gross withdrawals are explained by factors 1 and 3 and itself with factors 4 and 5 explaining about 10%.

Forecast error variance decompositions from the FAVAR(10F) model are presented in table 5.4. Innovations in natural gas gross withdrawals explain 10% or less the forecast error variances in factors 1, 2, 5, 6, 7, 8, and 10 at any time horizon. Variance in factor 3 is explained by natural gas gross withdrawals, 9% at one month ahead, 12% at six and 12 months ahead. Natural gas gross withdrawals are the major source of variation in factor 4 (43%) at one month ahead, but decrease to 17% at 12 months ahead. Natural gas gross withdrawals play an important role in explaining forecast error variance in factor 9 at one month ahead and become the most important explainer of variance in factor 9 at six and 12 months ahead other than itself. Natural gas gross withdrawals are dependent on only itself at the one-month horizon; its uncertainty at six and 12 months ahead is primarily initiated by innovations of itself and factor 9. Factors 4 and 6 explain variance in natural gas gross withdrawals over seven percent each at the 12 month ahead.

Forecast error variance decompositions from the VAR(5) model are presented in table 5.5. Variation in natural gas gross withdrawals is 100% explained by itself at one

Table 5.3. Forecast Error Variance Decompositions of Five Factors and Natural Gas Gross Withdrawals from the FAVAR(5F) Model

Horizon	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Natural Gas Gross Withdrawals
				Factor 1		
1	100.00	0.00	0.00	0.00	0.00	0.00
6	78.33	2.73	0.59	2.38	1.12	14.86
12	74.44	5.30	0.57	2.36	3.58	13.75
				Factor 2		
1	78.52	10.84	1.79	4.50	3.48	0.88
6	60.23	14.61	2.32	4.11	12.15	6.58
12	57.66	14.27	2.24	4.50	14.53	6.80
				Factor 3		
1	0.00	0.00	100.00	0.00	0.00	0.00
6	12.58	4.54	33.71	24.09	20.96	4.12
12	19.87	3.83	27.62	20.95	19.05	8.67
				Factor 4		
1	4.47	0.00	29.29	51.74	4.42	10.09
6	21.61	3.73	21.32	32.05	8.51	12.78
12	24.33	4.09	19.56	30.42	7.93	13.67
				Factor 5		
1	29.72	0.00	40.90	0.00	29.39	0.00
6	23.60	2.99	21.07	5.61	26.97	19.76
12	19.38	6.99	16.50	8.72	32.38	16.04
				Natural Gas Gross Withdrawals		
1	17.78	0.00	24.47	0.00	17.58	40.17
6	27.97	0.54	27.23	9.31	11.38	23.57
12	27.97	0.56	27.19	9.36	11.43	23.50

Table 5.4. Forecast Error Variance Decompositions of Ten Factors and Natural Gas Gross Withdrawals from the FAVAR(10F) Model

Horizon	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Natural Gas Gross Withdrawals
					Factor 1						
1	10.27	78.31	0.00	0.00	3.10	0.53	6.55	0.00	0.77	0.00	0.47
6	10.00	39.95	2.54	2.77	5.44	8.22	12.16	0.51	6.85	2.13	9.44
12	11.74	31.97	4.79	4.23	5.24	10.16	12.06	1.52	8.20	2.03	8.06
					Factor 2						
1	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	1.71	62.22	4.86	3.15	0.98	2.42	1.11	4.28	4.75	4.83	9.70
12	1.90	59.67	4.84	3.15	1.35	3.43	1.65	4.22	5.59	4.83	9.39
					Factor 3						
1	0.12	40.36	30.98	2.03	13.40	0.03	0.42	0.00	3.32	0.00	9.33
6	0.62	25.05	19.78	3.55	9.62	14.07	1.93	1.65	10.70	1.13	11.91
12	0.88	22.93	16.94	3.42	13.45	13.51	4.58	1.67	9.69	1.46	11.47
					Factor 4						
1	2.17	16.55	0.00	36.15	0.66	0.11	1.38	0.00	0.16	0.00	42.82
6	2.61	10.64	5.10	18.60	8.19	11.40	3.28	0.24	18.77	3.00	18.17
12	2.40	11.75	5.13	16.85	9.11	12.90	3.36	0.49	17.63	3.05	17.35
					Factor 5						
1	0.00	0.00	0.00	0.00	68.35	0.32	3.98	0.00	16.96	0.00	10.39
6	1.59	2.16	6.54	5.91	29.37	3.59	14.49	10.91	17.00	1.44	6.99
12	1.49	8.66	13.02	4.13	19.43	6.61	11.27	11.75	12.10	1.39	10.15

Table 5.4. Continued

Horizon	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Natural Gas Gross Withdrawals
						Factor 6					
1	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00
6	0.46	4.44	5.38	2.33	6.52	59.10	3.58	2.32	9.18	1.02	5.68
12	0.55	3.50	8.35	1.94	4.99	60.87	3.29	1.84	7.69	1.04	5.93
						Factor 7					
1	0.00	0.00	0.00	0.00	0.00	7.46	92.54	0.00	0.00	0.00	0.00
6	2.53	1.01	3.92	1.33	11.89	13.03	57.03	4.43	0.88	0.74	3.22
12	2.17	3.06	7.66	1.21	10.75	14.55	47.84	5.45	1.72	0.91	4.67
						Factor 8					
1	2.59	19.71	0.00	0.00	0.78	0.13	1.65	74.83	0.19	0.00	0.12
6	2.96	12.13	1.71	0.42	2.23	16.19	9.57	47.72	4.50	0.90	1.69
12	2.70	12.03	2.71	0.51	4.03	17.02	9.98	42.81	5.65	0.90	1.66
						Factor 9					
1	0.00	0.00	0.00	0.00	0.00	1.01	12.57	0.00	53.60	0.00	32.82
6	1.21	5.90	5.96	7.16	2.13	4.81	10.41	1.14	38.64	1.41	21.23
12	1.47	5.81	6.41	6.79	3.89	6.19	10.05	1.26	35.72	2.22	20.19
						Factor 10					
1	0.00	0.00	0.00	0.00	0.00	0.62	7.67	0.00	0.00	91.72	0.00
6	0.58	19.41	7.82	1.36	0.36	0.55	3.95	4.54	2.29	56.59	2.55
12	0.63	18.70	8.91	1.45	1.08	0.92	3.73	6.51	2.63	52.53	2.93

Table 5.4. Continued

Horizon	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Natural Gas Gross Withdrawals
					Natural Gas Gross Withdrawals						
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
6	2.75	1.08	2.51	8.40	3.24	6.76	4.88	1.05	26.88	2.97	39.48
12	2.72	1.06	3.04	8.59	3.14	7.33	4.86	1.05	26.78	3.25	38.17

Table 5.5. Forecast Error Variance Decompositions of Five Variables from the VAR(5) Model

Horizon	Henry Hub Natural Gas Spot Price	Natural Gas Consumption	WTI Crude Oil Spot Price	S&P500	Natural Gas Gross Withdrawal
			Henry Hub Natural Gas Spot Price		
1	93.86	0.00	6.14	0.00	0.00
6	89.44	0.11	9.34	0.93	0.18
12	89.29	0.25	9.32	0.95	0.19
			Natural Gas Consumption		
1	9.09	69.42	0.59	0.00	20.91
6	14.66	71.71	2.36	0.49	10.78
12	15.13	71.04	2.87	0.53	10.44
			WTI Crude Oil Spot Price		
1	0.00	0.00	100.00	0.00	0.00
6	1.42	3.49	82.10	11.20	1.79
12	1.54	3.96	81.51	11.13	1.86
			S&P500		
1	0.00	0.00	0.00	100.00	0.00
6	0.46	0.97	3.31	95.10	0.16
12	0.53	1.14	3.31	94.83	0.19
			Natural Gas Gross Withdrawal		
1	0.00	0.00	0.00	0.00	100.00
6	0.81	0.15	1.21	2.88	94.96
12	0.82	0.17	1.24	2.91	94.86

month ahead and approximately 95% at six and 12 months ahead. Similarly, at any time horizon, the key source of uncertainty in Henry Hub natural gas spot price is its own innovation (90 – 94%). Natural gas consumption itself, natural gas gross withdrawals, and Henry Hub spot price play the most important roles in explaining variation in natural gas consumption. WTI crude oil spot price is dependent on only itself (100%) at one month ahead; in addition to itself (82%), the S&P500 plays a role in explaining WTI's variance (11%) at six and 12 months ahead. Similar to WTI, S&P500 variance is completely self-dependent at one month ahead; at six and 12 months ahead, S&P variance is triggered by itself 95% and 3% by an innovation in WTI.

In summarizing the above discussion, natural gas gross withdrawals appear to primarily influence and are influenced by the dynamics of electricity consumption, electricity net generation, electricity prices, natural gas consumption, and natural gas prices. These are the categories which primarily characterize factors 1, 3, 4, 5, and 9. Results from the VAR(5) analysis agree with the above summary as natural gas gross withdrawals is one of the major sources of variation in natural gas consumption. In the FAVAR(5F) model, natural gas gross withdrawals play a significant role in explaining dynamics in factors 1, 4, and 5. Natural gas consumption and natural gas prices are the main categories describing these three factors. In the FAVAR(10F) model, natural gas gross withdrawals play a significant role in explaining dynamics in factors 4 and 9, which are primarily represented by natural gas consumption and renewable energy consumption and production. Dynamics in natural gas gross withdrawals are mainly caused by innovations in factors 1, 3, and itself in the FAVAR(5F) model, while the dynamics are mainly caused by innovations in factors 9 and itself. The key categories representing both factors 1 and 3 are electricity net generation, electricity prices, natural gas consumption, and natural gas prices. Factor 9 is generally descripted by natural consumption and renewable energy consumption and production.

Prequential Analysis[19]

Calibration measures (calibration plots and chi-squared goodness-of-fit test statistics), root mean-squared error, the Brier score and its decompositions, and the ranked probability score are calculated to assess out-of-sample forecasting ability of the various models. In addition to the FAVAR(5F), FAVAR(10F), and VAR(5) models, an univariate autoregressive (AR) model of natural gas gross withdrawals is fitted over in-sample data (February 2001 to December 2012) to forecast out-of-sample values for one step-ahead horizon over the period of January 2013 to December 2014. Schwarz criteria suggests that six lags are appropriate for the univariate AR model, hereafter AR(6).

Calibration Measures

A graphical representation and a goodness-of-fit test statistic are used to evaluate calibration performance. Calibration plots between the relative frequency and realized fractiles of the natural gas gross withdrawals forecasts from the four models are illustrated in figure 5.22. A forecast is well calibrated, when its calibration plot is close to the 45-degree line. Comparing plots of natural gas gross withdrawals forecasts from the four models, it appears that the plot from the FAVAR(5F) model is closest to the 45-degree line.

[19] See Chapter III for details on the prequential methodology.

Chi-squared test statistics are calculated based on 20 non-overlapping subintervals of the observed fractiles. Chi-squared goodness-of-fit test statistics on the forecasts of natural gas gross withdrawals from the four models are less than the 5% critical value of $\chi^2(19)$ (table 5.6), implying that all models yield well calibrated probabilistic forecasts of natural gas gross withdrawals.

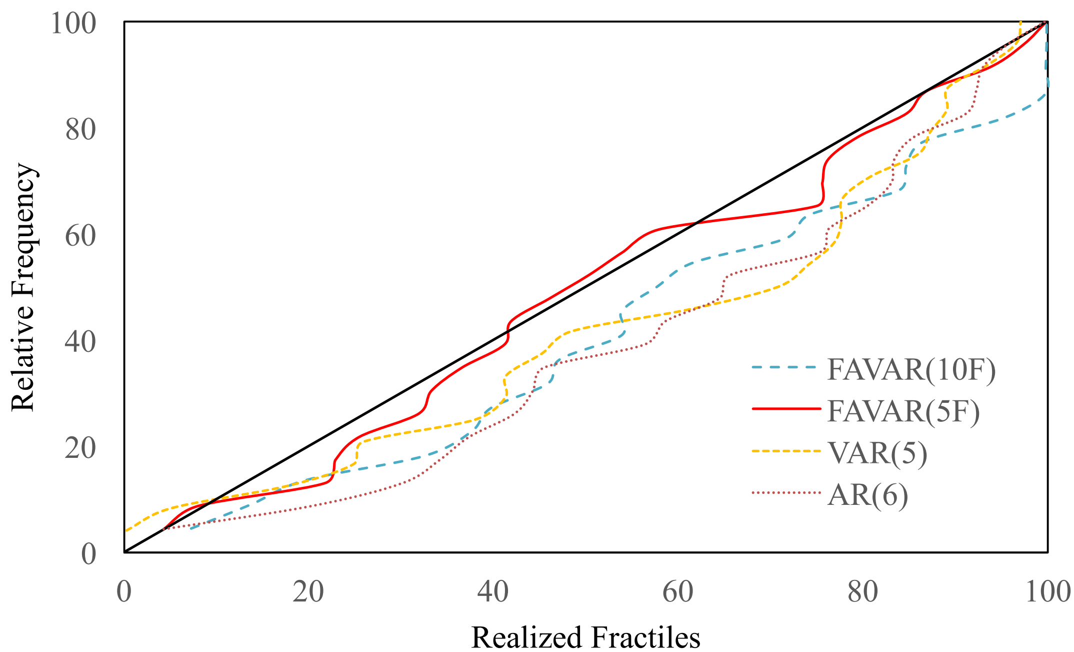

Figure 5.22. Calibration Plots of the Natural Gas Gross Withdrawals Forecast from the Four Models

Table 5.6. Test Statistics on the Probabilistic Forecast of Natural Gas Gross Withdrawals from the Four Models

Models	Chi-Squared Test Statistics	Root Mean-Squared Error	Brier Score	Ranked Probability Score
FAVAR(5F)[a]	10.9130	0.7660	0.8237	0.1418
FAVAR(10F)[b]	25.2727	1.0699	0.9105	0.1879
VAR(5)[c]	17.6667	0.6956	0.8358	0.1237
AR(6)[d]	16.1304	0.7426	0.8294	0.1364

Note: The null hypothesis of well calibration cannot be rejected if the chi-squared test statistic is less than the 5% critical value of $\chi^2(19)$ =30.144.

[a] The FAVAR(5F) model includes five estimated factors and natural gas gross withdrawals as the observable variable.

[b] The FAVAR(10F) model includes 10 estimated factors and natural gas gross withdrawals as the observable variable.

[c] The VAR(5) model includes natural gas gross withdrawals, Henry Hub natural gas spot price, natural gas consumption, WTI crude oil spot price, and S&P500.

[d] The AR(6) model is the univariate autoregressive model of natural gas gross withdrawals with six lags.

Root Mean-Squared Error (RMSE)

To calculate the RMSE for the probability forecasts, the means of the probability distributions are used as point forecast. Smaller RMSE indicates better forecasts. The forecast of natural gas gross withdrawals from the VAR(5) model has the smallest RMSE, while the forecast from the FAVAR(10F) model has the largest RMSE (table 5.6). Based on RMSE, it appears that the probability forecasts of natural gas gross withdrawals from the VAR(5) model is more desirable than those from factor models, although the AR(6) and FAVAR(5F) models have RMSE that are closer to the VAR(5) model than the FAVAR(10F) model.

The forecast error (the difference between observed withdrawals and mean of forecasted withdrawals) from each model is regressed on the difference between itself and the other models' forecast errors. This is called encompassing regression (Harvey and Newbold 2000). The null hypothesis of encompassing cannot be rejected for all directions, implying that the withdrawals forecasts from the four models are not different in encompassing. When a forecast encompasses the others, it means that the other forecasts contain no useful information not present in the encompassing forecast.

The Brier Score and Yates' Covariance Decomposition

The Brier score is a probability score that involves both calibration and sorting ability. The smaller the Brier score, the better probability forecasting. As with RMSE, the three models, AR(6), VAR(5), and FAVAR(5F) have Brier scores that are close (table 5.6). The FAVAR(5F) model yields the smallest Brier score on the natural gas gross withdrawals forecast, while the FAVAR(10F) model yields the largest Brier score. The AR(6) model has a smaller Brier score than the VAR(5).

Yates (1988) suggests partitioning the Brier score, $\overline{PS}(f,d)$, into five elements[20] such that $\overline{PS}(f,d) = Var(d) + MinVar(f) + Scat(f) + Bias^2 - 2Cov(f,d)$. Ideally, in a good forecast, the three components, *MinVar*, scatter, and $Bias^2$, should be small, while the last component, covariance term, should be large. *Var(d)* reflects the underlying variance of observed data and does not depend on the forecasts. As such, *Var(d)* of natural gas gross withdrawals in the four models are the same; they are calculated over the same out-of-sample data (table 5.7).

[20] See Chapter III for interpretation on each element.

Even though the FAVAR(5F) model has the smallest Brier score on the natural gas gross withdrawals forecast, it has the largest *Minvar* and second largest scatter. With smallest Bias^2 and the largest covariance term, the Brier score suggests the forecast of natural gas gross withdrawals from the FAVAR(5F) model are the most desirable among the four models. Although the FAVAR(10F) model has second largest covariance term, its largest scatter and relatively large *MinVar* and Bias^2 terms produce the largest Brier score on the natural gas gross withdrawals forecast.

Table 5.7. The Brier Score and Yates' Decomposition on the Probabilistic Forecast of Natural Gas Gross Withdrawals from the Four Models

Models	Brier Score	*Var(d)*	*MinVar*	*Scat(f)*	*Bias2*	*Cov(f, d)*
FAVAR(5F)[a]	0.8237	0.8819	0.0463	0.1211	0.0184	0.1220
FAVAR(10F)[b]	0.9105	0.8819	0.0372	0.1795	0.0211	0.1046
VAR(5)[c]	0.8358	0.8819	0.0265	0.0710	0.0279	0.0858
AR(6)[d]	0.8294	0.8819	0.0206	0.0687	0.0255	0.0837

Note: Brier Score = $Var(d) + MinVar(f) + Scat(f) + Bias^2 - 2Cov(f, d)$.
[a] The FAVAR(5F) model includes five estimated factors and natural gas gross withdrawals as the observable variable.
[b] The FAVAR(10F) model includes 10 estimated factors and natural gas gross withdrawals as the observable variable.
[c] The VAR(5) model includes natural gas gross withdrawals, Henry Hub natural gas spot price, natural gas consumption, WTI crude oil spot price, and S&P500.
[d] The AR(6) model is the univariate autoregressive model of natural gas gross withdrawals with six lags.

Even though the *MinVar* and scatter are relatively small, the Brier score on the natural gas gross withdrawals forecast from the VAR(5) model is ranked the third best. This is because it has the largest Bias^2 and a relatively small covariance term.

The AR(6) model gives the smallest *MinVar* and scatter on the natural gas gross withdrawals forecast. With a relatively large Bias^2 and the smallest covariance, the AR(6) gives the second best of the Brier score on the natural gas gross withdrawals forecast.

The Ranked Probability Score (RPS)

Similar to the Brier score, the smaller the RPS score, the better the probabilistic forecasts. The RPS suggests the forecast of natural gas gross withdrawals from the VAR(5) model is the most desirable (table 5.6). The AR(6) model provides the second smallest RPS on the forecast of natural gas gross withdrawals. Consistent with the other test statistics, the RPS on the natural gas gross withdrawals forecast from the FAVAR(10F) model is the largest.

Comparison of the Four Systems

Prequential analysis of the natural gas gross withdrawals forecasts from the four models are discussed in the previous subsections. Results on the forecasts of the other variables in the FAVAR(5F), FAVAR(10F), and VAR(5) models are presented in Appendix D. In this subsection, the forecasting performance of the entire system from the four models is explored. Being a univariate model, results on the AR(6) system in table 5.8 are the same as those in tables 5.6 and 5.7.

RMSE of each system is the square root of the average MSE of all forecasts or variables in the system. The RMSE on the FAVAR(5) system is the smallest while the RMSE on the AR(6) system is the largest (table 5.8). It appears that factor models provide useful information in forecasting because the RMSE on the factor models are smaller than those of the models not involving estimated factors.

The Brier score and its partitions of each system are calculated by treating all variables in the system as a single variable. When considering the entire system, variances of the observed outcome (*Var(d)*) are different across the four systems as each system incorporates different series. Unlike when considering only forecasts on natural gas gross withdrawals, the Brier score on the AR(6) system is the smallest with the Brier score on the FAVAR(5F) being the second smallest (table 5.8). One likely reason for the previous results is that the two models consider fewer variables than the FAVAR(10F) model. The FAVAR(10F) system has the largest Brier score with the largest *Var(d)* and the largest scatter. This is probably because the FAVAR(10F) system comprises 11 variables. Even though the FAVAR(5F) has relatively large *Var(d)* and scatter and has the largest *MinVar*, with the smallest Bias^2 and the largest covariance term, the Brier score on the FAVAR(5F) system is the second best. The VAR(5) system has the smallest *Var(d)* and smallest *MinVar*. It, however, has the smallest covariance term; the Brier score on the VAR(5) system is ranked the third best. Among the four systems, the AR(6) has the largest Bias^2, although its Brier score is the smallest.

The RPS on each system is an average RPS for all forecasts in the system. Consistent with the Brier score, the RPS suggests that the AR(6) system is the best among the four systems while the FAVAR(10) system is the worst (table 5.8). The VAR(5) system is ranked the second best while the FAVAR(5) system is ranked the third best.

Table 5.8. Test Statistics on Probabilistic Forecasts of the Four Systems

Systems	Root Mean-Squared Error	Brier Score	*Var(d)*	*MinVar*	*Scat(f)*	*Bias2*	*Cov(f, d)*	Ranked Probability Score
FAVAR(5F)[a]	0.3218	0.8398	0.8798	0.0321	0.1349	0.0067	0.1068	0.1576
FAVAR(10F)[b]	0.3353	0.9160	0.8916	0.0140	0.1389	0.0076	0.0680	0.1865
VAR(5)[c]	0.6758	0.8692	0.8790	0.0138	0.0734	0.0093	0.0532	0.1470
AR(6)[d]	0.7426	0.8294	0.8819	0.0206	0.0687	0.0255	0.0837	0.1364

Note: Brier Score = $Var(d) + MinVar(f) + Scat(f) + Bias^2 - 2Cov(f, d)$.

[a] The FAVAR(5F) model includes five estimated factors and natural gas gross withdrawals as the observable variable.

[b] The FAVAR(10F) model includes 10 estimated factors and natural gas gross withdrawals as the observable variable.

[c] The VAR(5) model includes natural gas gross withdrawals, Henry Hub natural gas spot price, natural gas consumption, WTI crude oil spot price, and S&P500.

[d] The AR(6) model is the univariate autoregressive model of natural gas gross withdrawals with six lags.

Discussion

The objective involves addressing two major questions: (1) whether and how the number of unobservable components (estimated factors) from a data-rich model influence inferences; and (2) whether and how the number of unobservable components (estimated factors) from a data-rich model influence probabilistic forecasting performance. To answer the first question, innovation accounting analysis is applied. Results suggest the inferences are only minimally affected by the number of estimated factors. Dynamic responses of the first five factors and natural gas gross withdrawals of the FAVAR(10F) model are slightly different from dynamic response of the FAVAR(5F) model. These finding are in line with Bernanke, Boivin, and Eliasz's (2005) finding that increasing the number of factors does not alter dynamic response results.

Natural gas gross withdrawals influence the U.S. energy sector, particularly electricity net generation, electricity prices, natural gas consumption, and natural gas prices. Natural gas gross withdrawals started to increase with increases in shale gas withdrawals. At the same time, U.S. natural gas consumption also increased (figure 5.23). Increases in domestic natural gas production have led to decreasing natural gas prices. Low natural gas prices induce increases in natural gas consumption, especially in the industrial and electric power generating sectors. Electricity net generation from natural gas in 2014 is approximately 75% greater than that in 2001 and 25% greater than that in 2007, the year that the U.S. EIA started reporting data on natural gas withdrawals from shale gas (U.S. EIA 2015a).

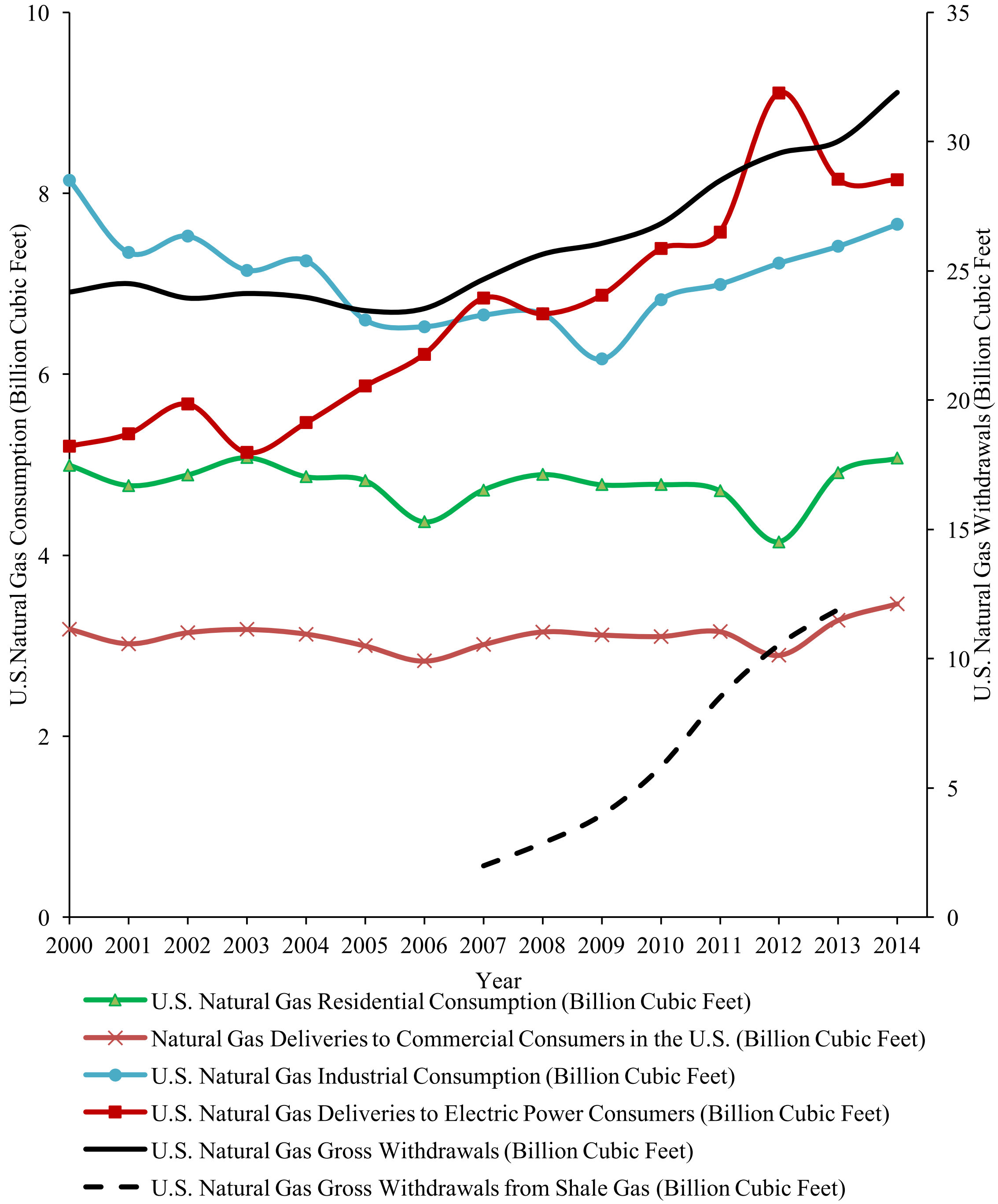

Figure 5.23. U.S. annual natural gas consumption by sectors, natural gas gross withdrawals, and natural gas gross withdrawals from shale gas (U.S. EIA 2015c, 2015d)

The second objective is addressed through the use of prequential forecasting (Dawid 1984) for four models, the FAVAR(5F), FAVAR(10F), VAR(5), and AR(6). The FAVAR(5F) model appears to be the most desirable forecasting model as it yields the smallest or the second-smallest Brier scores and RMSE when considering only natural gas withdrawals forecasts and when considering the entire forecasting system. Visually, the calibration plot for the natural gas gross withdrawals from the FAVAR(5F) model is the closest to a 45-degree line of the four systems. Further, in terms of the Brier score's partition, the FAVAR(5F) model has the minimum Bias^2, which is a miscalibration measure, and the maximum covariance, which is the essential indicator of forecasting ability.

Although the RMSE and the RPS suggest that forecasts of natural gas gross withdrawals from the VAR(5) is the best forecasting model among the four models, results for the entire system differ. With only one variables considered, the AR(6) model is sometimes indicated having "best" forecasting performance. Univariate models, however, generally do not provide sufficient information for policy implementation, as they cannot explicitly capture the relationships among variables (Bessler and Kling 1986).

Consistent with literature (Stock and Watson 2002, 2009; Moench 2008; Zagaglia 2010; Breitung and Eickmeier 2011), it appears that factor models provide useful information for forecasting. Two smallest Bias^2 and two largest covariances belong to the natural gas gross withdrawals forecasts from the FAVAR(5F) and the FAVAR(10F) models. Moreover, the RMSE on the FAVAR(5F) and FAVAR(10F) systems are the smallest and the second smallest. Nevertheless, including too many factors may lower forecasting performance. The FAVAR(10F) always yields the largest RMSE, Brier score, and RPS on the forecast of natural gas gross withdrawals. Including five factors appears to sufficiently explain the dynamics of the U.S. energy system. This is most likely because the relationships contained in factors 6 to 10 are already captured in factors 1 to 5. R-squared values from regressing factors 6 to 10 on each variable are generally smaller relative to R-squared values from regressing factors 1 to 5 on each variable. Thus, probabilistic forecasting performance appears to be affected by the number of estimated factors; including estimated factor results in better probabilistic forecasts but including too many estimated factors tends to worsen probabilistic forecasts.

CHAPTER VI

CONCLUSIONS AND DISCUSSION

Natural gas, an essential energy source in the U.S., is increasing in importance. Technological advances have been transforming the natural gas sector by increasing economically feasible production. Natural gas prices have been steadily decreasing since reaching their peak in 2003, leading to increased consumption. Have the dynamics of this sector changed as a result?

Issues generally ignored in literature are addressed in this dissertation. Studies about energy price and/or energy activity relationships are commonly conducted assuming that the relationships are constant over time (Serletis and Rangel-Ruiz 2004; Park, Mjelde, and Bessler 2008; Olsen, Mjelde, and Bessler 2014). There, however, are economic or energy events that may affect the economic dynamics and relationships. Economic viability of technological advances, for example, improvements in offshore drilling ability have increased well depth. Furthermore, the development from traditional vertical drilling to horizontal drilling are continuously evolving. A relatively recent technology in terms of economic viability is the fracking phenomenon that could possibly be inducing structural changes in the energy sector. A few recent studies (Mohammadi 2011; Lin and Wesseh 2013; Wakamatsu and Aruga 2013; Apergis, Bowden, and Payne 2015) have found breaks induced by the implementation of regulatory reform, oil shortage, financial crisis, and shale gas revolution. Moreover, because of limitations of empirical models in considering a large data set, the scopes of time series energy studies are usually restricted such that either variables from only a specific sub-sector or a couple selected variables from assorted sub-sectors are considered. In the time series literature, there exist few studies focusing on the entire energy system interacting with the economy.

Failure to consider structural changes brought about by technology and regulatory changes may lead to misleading inferences from any modeling exercise (Breitung and Eickmeier 2011; Chen, Dolado, and Gonzalo 2014). Tests for parameter constancy are implemented to investigate the potential existence of structural changes in long-run daily pricing relationships in the natural gas sector using the cointegrated (VAR) models introduced by Hansen and Johansen (1999). The presence of structural changes sheds some light on why previous studies may have conflicting results in natural gas pricing relationships. Altering pricing relationships (prices and transaction costs) as the results of structural changes may influence natural gas trading and transportation. Stakeholders in natural gas policy and infrastructure, such as pipeline systems, need to be aware of such changes. As such, results of the studies presented here should be of interest not only to those interested in energy markets from traders to policy makers, but also researchers interested in modeling energy issues, market structural changes, and time series analysis.

The potential presence of structural breaks in natural gas pricing relationships lead to the question how incorporating or ignoring the breaks affect the forecasting ability of a forecasting system. Under many circumstances, ignoring structural breaks may worsen forecasting ability of a model (Stock and Watson 2002; Banerjee, Marcellino, and Masten 2008). Allowing for uncertainty and time-varying parameters, the prequential approach introduced by Dawid (1987) is applied to probabilistically forecast natural gas returns. Unlike point forecasts, probabilistic forecasts can capture information related to uncertainty of unknown events.

Besides influencing the natural gas sector, technological and regulatory changes in the natural gas sector may influence the entire U.S. energy system. The factor-augmented vector autoregressive (FAVAR) method is an attractive methodology to examine the U.S. energy sector because of the large amount of information in this sector and its interactions with the macro-economy (Zagaglia 2010). Utilizing a data-rich environment, the FAVAR model is implemented to explore U.S. energy dynamics. One issue that arises in FAVAR methodology is the number of factors to include; different criteria suggest different numbers of factors. Besides the energy dynamics contributions, this study makes a methodology contribution by examining how the number of factors included affects the estimated U.S. energy dynamics and forecasting ability.

Dynamics in Daily Natural Gas Pricing

To investigate the possible existence of structural changes with unknown break points among North America natural gas spot markets, tests for parameter constancy in a cointegrated VAR model introduced by Hansen and Johansen (1999) are employed. The idea of the tests is to compare if the long-run relationship parameters recursively estimated by day from a vector error correction model differ from the parameters estimated for the entire data period. This methodology is applied to eight daily natural gas markets located throughout the U.S. (Chicago, Dominion South Point, Henry Hub, Malin, Oneok, Opal, and Waha Hub) and Canada (AECO Hub).

Test results suggest that long-run pricing relationships among the eight natural gas spot markets in the North America are not constant over the period of 1994 to 2014. Two potential points of structural changes are found, one during 2000 and the other in 2009. The structural change occurring around 2000 is probably induced by high natural gas prices and volatility in prices, the U.S. Federal Energy Regulatory Commission (FERC) Order No. 637 (which involves removing some pipeline price ceilings), and changes in imports. The likely major contributing factor to the break occurring around 2009 is the shale gas revolution.

As expected, in general, regulatory agencies are able to alter markets. Specifically, it appears FERC with its policies can and does alter the natural gas sector. However, not only was there a major FERC order in 2000, but also there were a multitude of major events that impacted the natural gas sector around this time. It is shown that more transitory events such as weather shocks can alter relationships but these alterations are short lived. The break occurring around 2000 was longer lived than the breaks associated with a weather shock, implying that time is necessary for the markets to learn and respond to regulatory changes.

Based on the possible existence of structural shifts, the data is divided into three sub-periods to investigate price dynamics. The first sub-period (May 1994 – September 2000) is the phase that the natural gas industry was maturing and becoming competitive as a result of the development of natural gas trading hubs and natural gas spot, term, and derivatives markets (Joskow 2013). The second sub-period (October 2000 – December 2009) is the phase that the U.S. natural gas sector was more import-intensive; ratios of U.S. natural gas imports to U.S. dry natural gas production are high relative to other periods (Joskow 2013; U.S. EIA 2015d, 2015h). Natural gas prices are relatively high and volatile during this period. In the third sub-period (January 2010 – October 2014), the natural gas industry becomes less import-reliance as the result of shale gas bloom (U.S. EIA 2014b).

Each individual market's role in price discovery differs in the three sub-periods; markets that were important for price discovery may be less important as a result of changes in the industry. Such information is helpful to energy traders. Because of the shale gas revolution, excess demand regions become excess supply regions; in response to such conversions, adjusting and expanding existing pipeline systems and constructing new systems to be bidirectional may improve transportation in the natural gas industry.

From an academic standpoint, inconsistent results in the literature, such as importance of markets in pricing relationships and whether there exists an east-west split in North America natural gas markets or not, are possibly the result of not only different methodologies employed and markets included, but also the time period of the data considered. Rather than considering data with long periods of time, researchers should realize when a structural break occurs and use an appropriate time period of data set to obtain appropriate inferences.

Because of the potential presence of structural breaks, the prequential forecasting approach introduced by Dawid (1984) is applied to determine whether and how the presence of structural breaks affects the performance of out-of-sample probability forecasting of natural gas returns. Vector autoregressive (VAR) models are estimated for three in-sample periods: (1) full model - May 4, 1994 to October 31, 2014; (2) two-period model - October 2, 2000 to October 31, 2014; and (3) recent model - January 1, 2010 to October 31, 2014. Calibration measures, root mean-squared error, the Brier score and its decompositions, and the ranked probability score are applied to evaluate out-of-sample forecasts (November 3, 2014 to April 30, 2015). Regardless of the model (or the data used), Henry Hub and AECO Hub are either the first or second easiest markets to forecast; whereas, Dominion South and Chicago are either the hardest or second hardest markets to forecast in terms of RMSE and the scoring rules. In terms of the Brier score's partition, the covariance between the forecast and the observed outcome is the best indicator of forecasting ability. Individually, Dominion South, Chicago, and Waha Hub returns are easier to predict when using the full, two-period, and recent models. Returns in Malin appear to be more difficult to predict, irrespective the model.

Considering different in-sample data results in different probability forecasts. Prequential analysis indicates that models that produce better forecasts are the two-period and full models, which incorporate a larger set of in-sample data, covering one and two break points. It appears that the existence of the structural changes does not affect the prequential forecasts; this is probably because the application of the Kalman filter is applied to address the issue of time-varying parameters. As a caution for future prequential studies, probabilistic forecasting performance may be poor if the presence of structural breaks and the application of the Kalman filter are ignored.

Transaction Costs

The presence of threshold cointegration between market pairs before and after structural changes in the long-term pricing relationship among the North America natural gas spot markets are examined. Based on the structural changes occurring around 2000 and 2009 and the forecasting results, two subsamples of data are considered; the first subsample covers the period of October 2, 2000 to December 31, 2009 and the second subsample covers the period of January 1, 2010 to October 31, 2014. Threshold values of each market-pair during the period of October 2, 2000 to December 31, 2009 differ from what is observed for the period of January 1, 2010 to October 31, 2014. Threshold bands of AECO Hub-Henry Hub, Malin-Henry Hub, Oneok-Henry Hub, Opal-Henry Hub, and Waha Hub-Henry Hub become narrower in the latter period; whereas, threshold bands of Chicago-Henry Hub and Dominion South-Henry Hub become wider.

As threshold intervals may be induced by transaction costs (including transportation costs), the narrower (wider) intervals likely suggest the smaller (larger) transaction costs. Changes in transaction costs between market pairs are most likely the result of changes in natural gas flows, which are the consequence of the structural change associated with the shale gas bloom. Changes in natural gas flows may be a signal that the industry needs to consider modifying and/or improving the pipeline system. In addition, lower/ higher transaction costs probably influence natural gas traders' decisions. Smaller (Larger) transaction costs potentially lead to more (less) natural gas trading between hubs.

Dynamics in the U.S. Energy Sector

To investigate whether and how the number of unobservable components from a data-rich model influences inferences from and probabilistic forecasting performance of various models, a factor-augmented vector autoregressive (FAVAR) approach, proposed by Bernanke, Boivin, and Eliasz (2005), is employed to characterize unobservable components. Innovation accounting analyses are applied to discover dynamic inferences. Then, the prequential forecasting approach is applied to evaluate predictive distributions for out-of-sample data. The structural breaks found in the daily natural gas pricing are not found in the monthly data rich factors using the methodology developed by Han and Inoue (2015). Two FAVAR models differing in their number of factors (five and ten factors), based on the range of optimal number of factors derived from Bai and Ng's (2002) criteria, are compared along with a five variable vector autoregressive (VAR) and a univariate model.

Based on innovation accounting analysis, inferences appear to be minimally affected by the number of estimated factors. Dynamic responses of the first five factors and natural gas gross withdrawals of the FAVAR model with 10 factors are slightly different from dynamic responses of the FAVAR model with five factors. These finding are in line with Bernanke, Boivin, and Eliasz's (2005) finding that increasing the number of factors does not substantially alter dynamic response results.

Consistent with literature (Stock and Watson 2002, 2009; Moench 2008; Zagaglia 2010; Breitung and Eickmeier 2011), it appears that factor models provide useful information in forecasting. Two smallest Bias^2 and two largest covariance terms on the natural gas gross withdrawals forecasts are from the FAVAR(5F) and FAVAR(10F) models. Moreover, the RMSE on the two FAVAR systems are the smallest and the second smallest. These imply that the probabilistic forecasting performance is affected by the number of estimated factors. Using estimated factor results in better probabilistic forecasts but including too many estimated factors tends to worsen probabilistic forecasts. When issues of parsimony and appropriateness arise, building a data-rich FAVAR model with parsimony factors tends to be more desirable.

Limitations and Further Research

It should be noted that inferences in this dissertation must be viewed in light of the studies' limitations. In Chapter II, the data used appears to be non-Gaussian but the tests for parameter constancy in the cointegrated VAR model are based on the assumption of Gaussian error terms. The robustness of tests for parameter constancy in the co-integrated VAR model under the assumption of Gaussian data when applying non-Gaussian data is left for future research. Moreover, any break dates is left at the direction of the analyst to decide when applying Hansen and Johansen's (1999) tests to detect the structural changes. The advent of formal tests for structural break dates in the cointegrated VAR model would be a contribution to the literature in long-run relationships and structural changes. Besides, AECO Hub is considered in Chapter II to represent Canada; however, it provides limited information to the system of pricing dynamics. Including other markets in Canada may provide a fuller picture of the natural gas market. Inclusion of additional markets, however, comes at a large cost in time series methods.

The issue of the number of in-sample observations arises in Chapter III. It was expected that the model in which no structural shifts occurring during the period of in-sample data used would yield the best probability forecasts among the three models. The model, which incorporates smallest data period, however, has a poorer forecasting performance than the other two models, which incorporate data sets that include periods of structural breaks. Does the number of in-sample observations matter in forecasting performance? Addressing this question is left for further studies.

The AECO Canadian natural gas spot market is included but the price at this market is converted to U.S. dollars. A study considering potential effect of exchange rates and using Canadian currency is left for future research.

Physical trading between some natural gas hubs, for example, AECO Hub, Henry Hub, and other markets located between these two hubs may not occur, because of transportation constraints. Examining transaction costs between natural gas market centers where only physical trading occurs is left for future studies. Further, it is recommended to pair markets based on the information of the major natural gas transportation corridors. Consistent long periods of daily pipeline data are difficult to obtain. Besides, the development of threshold cointegration models in which both threshold values and threshold intervals vary across time would provide a contribution in transaction costs' studies.

Structural changes in the dynamics of the natural gas pricing system are found but these structural changes did not appear in dynamics of the U.S. energy system. Whether the frequency of data considered or methodologies used to inspect structural changes causes such inconsistency is left for further research. Moreover, to broaden understandings of the U.S. energy dynamics, observable measures of economic activity and prices such as crude oil production, electricity net generation, Henry Hub natural gas spot price, NYMEX natural gas futures price, may be incorporated, along with or in place with natural gas gross withdrawals.

All impulse response functions presented in this dissertation are point estimates. Confidence intervals for impulse responses could be reported to capture the uncertainty in the results. In addition, test statistics on probability forecasts from different models are close; as such, it is difficult to determine which models provide better forecasts. Developing formal tests on whether these statistics are significantly different is left a suggestion for future studies.

REFERENCES

Alterman, S. 2012. Natural Gas Price Volatility in the UK and North America. Oxford Institute for Energy Studies. Accessed October 22, 2014. Available at: http://www.oxfordenergy.org/wpcms/wp-content/uploads/2012/02/NG_60. pdf.

American Petroleum Institute. 2014. Understanding Natural Gas Markets. Accessed August 12, 2015. Available at: http://www.api.org/~/media/ files/oil-and-natural-gas/natural-gas-primer/understanding-natural-gas-markets-primer-low.pdf.

Andrews, D.W. 1993. Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61(4): 821-856.

Apergis, N., N. Bowden, and J.E. Payne. 2015. Downstream Integration of Natural Gas Prices Across U.S. States: Evidence from Deregulation Regime Shifts. *Energy Economics* 49: 82-89.

Ardeni, P.G. 1989. Does the Law of Price Really Hold for Commodity Prices? *American Journal of Agricultural Economics* 71(3): 661-669.

Bai, J. 1997. Estimating Multiple Breaks One at a Time. *Econometric Theory* 13(3): 315-352.

Bai, J., and S. Ng. 2002. Determining the Number of Factors in Approximate Factor Models. *Econometrica* 70(1): 191-221.

Bai J., and P. Perron. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66(1): 47-78.

———. 2003. Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Economics* 18 (1): 1-22.

Balke, N.S., and T.B. Fomby. 1997. Threshold Cointegration. *International Economic Review* 38(3): 627-645.

Baltagi, B. H. 2011. *Econometrics.* New York, NY: Springer-Verlag.

Banerjee, A., M. G. Marcellino, and I. Masten. 2008. Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change. *CEPR Working Paper 6706.* Accessed August 12, 2015. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1141612.

Bekkerman, A., B.K. Goodwin, and N.E. Piggott. 2013. A Variable Threshold Band Approach to Measuring Market Linkages. *Applied Economics* 45(19): 2705-2714.

Bernanke, B.S. 1986. Alternative Explanations of the Money-Income Correlation. *Carnegie-Rochester Conference Series on Public Policy* 25: 49-99.

Bernanke, B.S., and J. Boivin. 2003. Monetary Policy in a Data-Rich Environment. *Journal of Monetary Economics* 50(3): 525-546.

Bernanke, B.S., J. Boivin, and P. Eliasz. 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics* 120(1): 387-422.

Bessler, D.A., and J.L. Kling. 1990. Prequential Analysis of Cattle Prices. *Applied Statistics* 39(1): 95-106.

———. 1986. Forecasting Vector Autoregressions with Bayesian Priors. *American Journal of Agricultural Economics* 68(1): 144-151.

Bessler, D.A., and R. Ruffley. 2004. Prequential Analysis of Stock Market Returns. *Applied Economics* 36(5): 399-412.

Bessler, D.A., and Z. Wang. 2012. D-Separation, Forecasting, and Economic Science: A Conjecture. *Theory and Decision* 73(2): 295-314.

Bloomberg L.P. 2015. North America Natural Gas Spot Prices. Retrieved May 15, 2015 from Bloomberg Professional Service.

Breitung, J., and S. Eickmeier. 2011. Testing for Structural Breaks in Dynamics Factor Models. *Journal of Econometrics* 163(1): 71-84.

Brier, G.W. 1950. Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review* 78(1): 1-3.

Brown, R., J. Durbin, and J. Evans. 1975. Techniques for Testing the Constancy of Regression Relationships Over Time. *Journal of the Royal Statistical Society. Series B (Methodological).* 37(2): 149-192.

Bruggeman, A., P. Donati, and A. Warne. 2003. Is the Demand for Euro Area M3 Stable? *ECB Working Paper 255*. Accessed May 31, 2015. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=457521.

Bunn, D.W. 1984. *Applied Decision Analysis*. New York, NY: McGraw-Hill.

Casillas-Olvera, G., and D.A. Bessler. 2006. Probability Forecasting and Central Bank Accountability. *Journal of Policy Modeling* 28(2): 223-234.

Chen, L., J.J. Dolado, and J. Gonzalo. 2014. Detecting Big Structural Breaks in Large Factor Models. *Journal of Econometrics* 180(1): 30-48.

Chickering, D.M. 2002. Leaning Equivalence Classes of Bayesian-Network Structures. *Journal of Machine Learning Research* 2: 445-498.

———. 2003. Optimal Structure Identification with Greedy Search. *Journal of Machine Learning Research* 3: 507-554.

Chow, G.C. 1960. Tests for Equality between Sets of Coefficients in Two Linear Regressions. *Econometrica* 28(3): 591-605.

Cuddington, J.T., and Z. Wang. 2006. Assessing the Degree of Spot Market Integration for U.S. Natural Gas: Evidence from Daily Price Data. *Journal Regulatory Economics* 29 (2): 195-210.

Dawid, A.P. 1984. Statistical Theory: The Prequential Approach. *Journal of the Royal Statistical Society* 147(2): 278-292.

DeJong, D.N., J.C. Nankervis, N.E. Savin, and C.H. Whiteman. 1992. The Power Problems of Unit Root Test in Time Series with Autoregressive errors. *Journal of Econometrics* 53(1-3): 323-343.

Dennis, J.G. 2006. *CATs in RATs Cointegration Analysis of Time Series Version 2*. Evanston, IL: Estima.

DeVany, A., and W.D. Walls. 1993. Pipeline Access and Market Integration in the Natural Gas Industry: Evidence from Cointegration Tests. *The Energy Journal* 14(4): 1-19.

———. 1994a. Natural Gas Industry Transformation, Competitive Institutions and the Role of Regulation. *Energy Policy* 22(9): 755-763.

———. 1994b. Open Access and the Emergence of a Competitive Natural Gas Market. *Contemporary Economic Policy* 12(2): 77-96.

———. 1996. The Law of One Price in a Network: Arbitrage and Price Dynamics in Natural Gas City Gate Markets. *Journal of Regional Science* 36(4): 555-570.

Diebold, F.X., and C. Li. 2006. Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics* 127(1-2): 337-364.

Doan, T. 2000. *RATS: User's Manual 4.3*. Evanston, IL: Estima.

Doane, M.J., and D.F. Spulber. 1994. Open Access and the Evolution of the United States Spot Market for Natural Gas. *Journal of Law and Economics* 37(2): 477-517.

Duffee, G.R. 2002. Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance* 57(1): 405-443.

Enders, W. 2004. *Applied Econometric Time Series*. New York: John Wiley & Sons.

Enders, W., and K. Chumrusphonlert. 2004. Threshold Cointegration and Purchasing Power Parity in the Pacific Nations. *Applied Economics* 36(9): 889-896.

Engel, R.F., and C.W.J. Granger. 1987. Cointegration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55(2): 251-276.

Epstein, E.S. 1969: A Scoring System for Probability Forecasts of Ranked Categories. *Journal of Applied Meteorology and Climatology* 8(6): 985-987.

Fair, R.C. 1986. Evaluating the Predictive Accuracy of Models. In Z. Griliches and M.D. Intriligator, eds. *Handbook of Econometrics Vol. III*. New York: Elsevier Science Publisher BV, pp.1980-1995.

Federal Reserve Bank of St. Louis. 2015. FRED Economic Data. Accessed May 4, 2015. Available at: https://research.stlouisfed.org/fred2/.

Goodwin, B.K., and N.E. Piggott. 2001. Spatial Market Integration in the Presence of Threshold Effects. *American Journal of Agricultural Economics* 83(2): 302-317.

Granger, C. W. J. 1981. Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics* 16(1): 121-130.

Grimes, J. 2014. Eastern Gas Basis: Pipeline Pain or Gain? Last modified August 26. Accessed August 12, 2015. Available at: http://www.morningstar commodity.com/Research/Morningstar_Eastern_Gas_Basis_Research.pdf.

Hamilton, J.D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Han, X., and A. Inoue. 2015. Tests for Parameters Instability in Dynamic Factor Models. *Econometric Theory* 31(5): 1117-1152.

Hansen, B.E., 1992. Tests for Parameter Instability in Regressions with I(1) Processes. *Journal of Business and Economic Statistics* 10(3): 321-335.

———. 1996. Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis. *Econometrica* 64(2): 413-430.

———. 1999. Testing for Linearity. *Journal of Economic Surveys* 13(5): 551-576.

Hansen, B.E., and B. Seo. 2002. Testing for Two-Regime Threshold Cointegration in Vector Error-Correction Models. *Journal of Econometrics* 110(2): 293-318.

Hansen, H., and S. Johansen. 1999. Some Tests for Parameter Constancy in Cointegrated VAR-Models. *The Econometrics Journal*, 2(2): 306-333.

Hansen, H., and K. Juselius. 1995. *CATs in RATs*. *Cointegration Analysis of Time Series*. Evanston, IL: Estima.

Hansen, P.R. 2003. Structural Changes in the Cointegrated Vector Autoregressive Model. *Journal of Econometrics* 114(2): 261-295.

Harvey, D., and P. Newbold. 2000. Tests for Multiple Forecast Encompassing. *Journal of Applied Econometrics* 15(5): 471-482.

Hoover, R. 2005. Automatic Inference of the Contemporaneous Causal Order of a System of Equations. *Econometric Theory* 21(1): 69-77.

IndexMundi. 2015. Uranium Monthly Price. Accessed May 4, 2015. Available at: http://www.indexmundi.com/commodities/?commodity=uranium &months=360.

Johansen, S. 1988. Statistical Analysis Testing of Cointegrating Vectors. *Journal of Economic Dynamics and Control* 12(2): 231-254.

———. 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59: 1551-1580.

———. 1995. *Likelihood-Based Inference on Cointegrated Vector Autoregressive Models*. New York, NY: Oxford University Press.

Johansen S., R. Mosconi, and B. Nielsen. 2000. Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend. *The Econometrics Journal* 3(2): 216-249.

Joskow, P. 2013. Natural Gas: From Shortages to Abundance in the United States. *American Economic Review* 103(3): 338-343.

King, M., and M. Cuc. 1996. Price Convergence in North American Natural Gas Spot Markets. *The Energy Journal* 17(2): 17-42.

Kleit, A.N. 1998. Did Open Access Integrate Natural Gas Markets? An Arbitrage Cost Approach. *Journal of Regulatory Economics* 14(1): 19-33.

Kling, K.L., and D.A. Bessler. 1989. Calibration-Based Predictive Distributions: An Application of Prequential Analysis to Interest Rates, Money, Prices, and Output. *The Journal of Business* 62(4): 477-499.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin. 1992. Testing the Null Hypothesis of Stationary against the Alternative of a Unit Root. *Journal of Econometrics* 54(1): 159-178.

Lai, P.C., and D.A. Bessler. 2015. Price Discovery between Carbonated Soft Drink Manufacturers and Retailers: A Disaggregate Analysis with PC and LiNGAM Algorithms. *Journal of Applied Economics* 18(1):173-198.

Lescaroux, F., and V. Mignon. 2009. Measuring the Effects of Oil Prices on China's Economy: A Factor-Augmented Vector Autoregressive Approach. *Pacific Economic Review* 14(3): 410-425.

Lin, B., and P.K. Wesseh Jr. 2013. What Causes Price Volatility and Regime Shifts in the Natural Gas Market. *Energy* 55: 553-563.

Lo, M.C., and E. Zivot. 2001. Threshold Cointegration and Nonlinear Adjustment to the Law of One Price. *Macroeconomic Dynamics* 5(4): 533-576.

Lombardi, M.J., C. Osbat, and B. Schnatz. 2012. Global Commodity Cycles and Linkages: a FAVAR Approach. *Empirical Economics* 43(2): 651-670.

Lüthepohl H., P. Saikkonen, and C. Trenkler. 2004. Testing for the Cointegrating Rank of a VAR Process with Level Shift at Unknown Time. *Econometrica* 72(2): 647-622.

Moench, E. 2008. Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach. *Journal of Econometrics* 146(1): 26-43.

Mohammadi, H. 2011. Market Integration and Price Transmission in the U.S. Natural Gas Market: From the Wellhead to End Use Markets. *Energy Economics* 33(2): 227-235.

Murphy, A.H. 1970. The Ranked Probability Score and the Probability Score: A Comparison. *Monthly Weather Review* 98(12): 917-924.

National Energy Technology Laboratory. 2013. Modern Shale Gas Development in the United States: An Update. Accessed October 22, 2014. Available at: http://www.netl.doe.gov/File%20Library/Research/Oil-Gas/shale-gas-primer-update-2013.pdf.

National Oceanic and Atmospheric Administration. 2014. Degree Days Statistics. Last modified January 3. Accessed March 31, 2015. Available at: http://www.cpc.ncep.noaa.gov/products/analysis_ monitoring/cdus/degreedays/.

Newey, W., and K. West. 1994. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* 61(4): 631-653.

Nyblom, J. 1989. Testing for the Constancy of Parameters Over Times. *Journal of the American Statistical Association* 84(405): 223-230.

Olsen, K., J.W. Mjelde, and D.A. Bessler. 2014. Price Formulation and the Law of One Price in Internationally Linked Markets: An Examination of the Natural Gas Markets in the USA and Canada. *The Annals of Regional Science* 54(1): 117-142.

O'Connell, P.G.J., and S.J. Wei. 2002. The Bigger They Are, the Harder They Fall: Retail Price Differences Across U.S. Cities. *Journal of International Economics* 56(1): 21-53.

Park, H., J.W. Mjelde, and D.A. Bessler. 2007. Time-varying Threshold Cointegration and the Law of One Price. *Applied Economics* 39(9): 1091-1105.

———. 2008. Price Interactions and Discovery among Natural Gas Spot Markets in North America. *Energy Policy* 36(1): 290-302.

Pearl, J. 2000. Causality: Models, Reasoning, and Inference. Cambridge, UK: Cambridge Press.

Phillips, P.C.B., and B.E. Hansen. 1990. Statistical Inference in Instrumental Variables Regession With I(1) Processes. *The Review of Economic Studies* 57(1): 99-125.

Quandt R. 1960. Tests of the Hypothesis that a Linear Regression Obeys Two Separate Regimes. *Journal of the American Statistical Association* 55 (290): 324-330.

Said, S.E., and D.A. Dickey. 1984. Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order. *Biometrika* 71(3): 599-607.

Saikkonen, P., and H. Lüthepohl. 2000. Testing for the Cointegrating Rank of a VAR Process with Structural Shifts. *Journal of Business and Economics Statistics* 18(4): 451-464.

Sargent, T., and C. Sims. 1977. Business Cycle Modeling without Pretending to Have Too Much a-priori Economics Theory. In C. Sims et al., eds. *New Methods in Business Research.* Minneapolis, MN: Federal Reserve Bank of Minneapolis, pp.145-168.

Seo, B. 1998. Tests for Structural Change in Cointegrated Systems. *Econometric Theory* 14(2): 222-59.

Serletis, A. 1997. Is There an East-West Split in North American Natural Gas Markets? *The Energy Journal* 18(1): 47-62.

Serletis, A., and R. Rangel-Ruiz. 2004. Testing for Common Features in North American Energy Markets. *Energy Economics* 26(3): 401-414.

Shimizu, S., P.O. Hoyer, A. Hyvarinen, and A. Kerminen. 2006. A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research* 7: 2003-2030.

Spirtes, P., and C. Glymour. 1991. An Algorithm for Fast Recovery of Sparse Causal Graphs. *Social Science Computer Review* 9(1): 62-72.

Spirtes, P., Glymour, C. and Scheines, R. 2000. *Causation, Prediction, and Searc*h. Cambridge, MA: MIT Press

Stock, J.H., and M.W. Watson. 2002. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics* 20(2): 147-162.

———. 2009. Forecasting in Dynamic Factor Models Subject to Structural Instability. In J. Castle and N. Shephard, eds. *The Methodology and Practice of Econometrics, A Festschrift in Honor of Professor David F. Hendry.* Oxford, UK: Oxford University Press, pp.173-205.

Swanson, N.R., and C.W.J, Granger. 1997. Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions. *Journal of the American Statistical Association* 92(437): 357-367.

Tsay, R.S. 1998. Testing and Modeling Multivariate Threshold Models. *Journal of the American Statistical Association* 93(443): 1188-1202.

U.S. Energy Information Administration. 1996. Short-Term Energy Outlook April 1996. Accessed March 2, 2015. Available at: http://www.eia.gov/forecasts/steo/archives/2Q96.pdf.

———. 2010. Energy Timeline: Natural Gas. Last modified June 2010. Accessed February 10, 2015. Available at: http://www.eia.gov/kids/energy.cfm?page=tl_naturalgas.

———. 2011. Review of Emerging Resources: U.S. Shale Gas and Shale Oil Plays. Last modified July 18. Accessed October 22, 2014. Available at: http://www.U.S.EIA.gov/analysis/studies/usshalegas/.

———. 2013a. 2012 Brief: Average Wholesale Natural Gas Price Fell 31% in 2012. Last modified January 8. Accessed October 22, 2014. Available at: http://www.eia.gov/todayinenergy/detail.cfm?id= 9490.

———. 2013b. Annual Energy Outlook 2013 with Projections to 2040. Accessed October 22, 2014. Available at: http://www.eia.gov/forecasts/aeo/pdf/0383%282013%29.pdf.

———. 2013c. Increased Northeast Natural Gas Production Reduced Net Inflow of Supply from Other Areas. Last modified November 19. Accessed August 6, 2015. Available at: http://www.eia.gov/todayinenergy/detail.cfm?id=13851

———. 2013d. Natural Gas Explained: Imports and Exports. Accessed October 22, 2014. Available at: http://www.eia.gov/energyexplained/index.cfm?page=natural_ gas_ imports.

———. 2014a. 32% 0f Natural Gas Pipeline Capacity into the Northeast Could be Bidirectional by 2017. Last modified December 2, 2014. Accessed March 24, 2015. Available at: http://www.eia.gov/todayinenergy/detail. cfm?id=19011.

———. 2014b. Annual Energy Outlook 2014 with Projections to 2040. Accessed October 22, 2014. Available at: http://www.eia.gov/forecasts/aeo/pdf/0383(2014).pdf.

———. 2014c. Natural Gas Explained: Factor Affecting Natural Gas Prices. Accessed October 22, 2014. Available at: http://www.eia.gov/energyexplained/index.cfm?page=natural_gas_factors_affecting._prices.

———. 2014d. Natural Gas Explained: Where Our Natural Gas Comes from. Last modified December 18. Accessed March 2, 2015. Available at: http://www.eia.gov/energyexplained/index.cfm?page= natural_ gas_where.

———. 2014e. Natural Gas Weekly Update. Last modified January 9. Accessed March 2, 2015. Available at: http://www.eia.gov/naturalgas/weekly/.

———. 2014f. Natural Gas Spot and Futures Prices (NYMEX). Accessed October 22, 2014. Available at: http://www.eia.gov/dnav/ng/ng_pri_fut_s1_d.htm.

———. 2014g. State Profile: California. Last modified June 19. Accessed March 2, 2015. Available at: http://www.eia.gov/state/analysis.cfm?sid= CA.

———. 2014h. State Profile: Louisiana. Last modified November 20. Accessed July 31, 2015. Available at: http://www.eia.gov/state/analysis.cfm?sid=LA.

———. 2014i. State Profile: Oklahoma. Last modified November 20. Accessed Aug 9, 2015. Available at: http://www.eia.gov/state/analysis.cfm?sid= OK.

———. 2014j. State Profile: Texas. Last modified November 20. Accessed Aug 12, 2015. Available at: http://www.eia.gov/state/analysis.cfm?sid=TX.

———. 2015a. Electricity Data Browser. Accessed September 8, 2015. Available at: http://www.eia.gov/electricity/data/browser/.

———. 2015b. Glossary. Accessed March 12, 2015. Available at: http://www.eia.gov/tools/glossary/index.cfm?id=H.

———. 2015c. Natural Gas Consumption by End Use. Last modified July 31, 2015. Accessed July 31, 2015. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_dcu_nus_a.htm.

———. 2015d. Natural Gas Gross Withdrawals and Production. Accessed March 12, 2015. Available at: http://www.eia.gov/dnav/ng/ng_prod_sum_dcu_ NUS_a.htm.

———. 2015e. Pennsylvania Dry Natural Gas Production. Last modified July 31, 2015. Accessed July 31, 2015. Available at: http://www.eia.gov/dnav/ng/hist/na1160_spa_2a.htm.

———. 2015f. State Profile: Illinois. Last modified March 19. Accessed July 17, 2015. Available at: http://www.eia.gov/state/analysis.cfm?sid=IL.

———. 2015g. State Profile: Pennsylvania. Last modified May 21. Accessed July 17, 2015. Available at: http://www.eia.gov/state/analysis.cfm?sid= PA.

———. 2015h. U.S. Natural Gas Imports by Country. Last modified February 27. Accessed March 12, 2015. Available at: http://www.eia.gov/dnav/ng/ng_move_impc_s1_a.htm

U.S. Federal Energy Regulatory Commission. 2000. Regulation of Short-Term Natural Gas Transportation Services, and Regulation of Interstate Natural Gas Transportation Services. Accessed May 15, 2015. Available at: http://www.ferc.gov/legal/maj-ord-reg/land-docs/rm98-10.pdf.

Wakamatsu, H., and K. Aruga. 2013. The Impact of the Shale Gas Revolution on the US and Japanese Natural Gas Markets. *Energy Policy* 62: 1002-1009.

Walls. W.D. 1994a. A Cointegration Rank Test of Market Linkages with an Application to the U.S. Natural Gas Industry. *Review of Industrial Organization* 9(2): 181-191.

———. 1994b. Price Convergence across Natural Gas Fields and City Markets. *The Energy Journal* 15(4): 37-48.

Wang, Z. and D.A. Bessler. 2005. A Monte Carlo Study on the Selection of Cointegrating Rank Using Information Criteria. *Econometric Theory* 21(3): 593-620.

Working Group of Commercial Energy Firms. 2009. Significant Price Discovery Contract Proceeding, AECO Financial Basis Contract. Accessed July 11, 2015. Available at: http://www.cftc.gov/ucm/groups/public/@lrfederalregister/documents/frcomment/09-016c002.pdf.

Yates, F. 1988. Analyzing the Accuracy of Probability Judgements for Multiple Events: An Extension of the Covariance Decomposition. *Organizational Behavior and Human Decision Process* 41(3): 281-299.

Yang, J., D.A. Bessler, and D.J. Leatham. 2000. The Law of One Price: Developed and Developing Country Market Integration, *Journal of Agricultural and Applied Economics* 32(3): 429–40.

Zagaglia, P. 2010. Macroeconomic Factors and Oil Futures Prices: A Data-Rich Model. *Energy Economics* 32(2): 409-417.

Zellner. A., C. Hong, and C.K. Min. 1991. Forecasting Turning Points in International Output Growth Rates Using Bayesian Exponentially Weighted Autoregression, Time-Varying Parameter, and Polling Techniques. *Journal of Econometrics* 49(1): 475-304.

APPENDIX A

DESCRIPTON OF PRICE SERIES CONSIDERED IN CHAPTERS II THROUGH IV

The following descriptions are taken verbatim from the Bloomberg Professional Service (2014)

Spot Natural Gas Price/AECO C Hub (NGCAAECO)
Tickers such as this in {ALLX NGCA <GO>} represent Canadian spot natural gas prices in U.S. dollars per million Btu. Corresponding prices in Canadian dollars per gigajoule for many of these prices can be found at {ALLX NGCD <GO>}. Natural gas at EnCana Corp.'s AECO C Hub in Alberta where TransCanada Pipeline's Alberta system, also known as Nova Gas Transmission, connects to Foothills Pipeline and Alberta Natural Gas. Prices converted from Canadian to U.S. dollars.

Mid-Continent Natural Gas Spot Price/Chicago City Gate (NAGANGPL)
Natural gas delivered to Chicago utilities including Nicor, Peoples Gas Light & Coke, Northern Indiana Public Service (NIPSCO). Major pipelines providing deliveries include NGPL, Alliance and ANR.

Dominion South Point Natural Gas Spot Price (NGNECNGO)
Natural gas at Dominion Transmission's South Point pool, which runs through parts of Ohio, Pennsylvania, and West Virginia on two separate lines. Major pricing locations in the pool are at the Lebanon, Ohio, interconnect with ANR Pipeline and the Oakford, Pennsylvania, storage facility. Bloomberg priced those two locations separately until Sept. 22, 2004 when the ticker for Lebanon was discontinued. The Oakford ticker was used for the consolidated South Point price.

Henry Hub (NGUSHHUB)
Please note that prices are no longer updated intraday. Prices are end-of-day. Starts Jan. 25, 1991 with weekly changes, moves to daily updates on March 10, 1994. Natural gas for next-day delivery at the Henry Hub, the benchmark U.S. pricing point and delivery point for New York Mercantile Exchange futures. The Henry Hub is operated by Sabine Pipe Line LLC and located in Erath, Louisiana. The hub has interconnects with Gulf South, Sonat, NGPL, Texas Eastern, Sabine, Columbia Gulf, Transco, Trunkline, Jefferson Island and Acadian Gas. To see older data please go to HP and change the Source to BGAP.

Gas Transmission Northwest Malin Oregon Spot Natural Gas Price (NGWCPGSP)
Natural gas delivered into PG&E's California Gas Transmission from TransCanada's Gas Transmission Northwest at their Malin, Oregon interconnect near the Oregon-California border.

Oneok Gas Transportation OGT Natural Gas Spot Price (NTGSOKON)
Natural gas delivered into Oneok Gas Transportation's intrastate pipeline in Oklahoma. Oneok, or OGT, was previously known as ONG Transmission and traders still often refer to the pipeline by that name.

Rocky Mountain Natural Gas Spot Price/Kern River Opal Wyoming (NGRMKERN)
Natural gas at the Opal, Wyoming, processing plant and Muddy Creek compressor station in southwestern Wyoming where Kern River pipeline connects with Northwest Pipeline, Questar Pipeline, and Colorado Interstate Gas.

Natural Gas Waha Hub Spot Price (NGTXOASI)
Deliveries into intrastate and interstate pipelines near the Waha, Texas header system. Pipelines with interconnects near Waha include Oasis, Lone Star, Delhi, El Paso Natural Gas, Transwestern, NGPL and Northern Natural.

APPENDIX B

SUPPLEMENTARY RESULTS FOR CHAPTER II

Table B.1. Augmented Dickey-Fuller (ADF) and Kwiatkowski-Philips-Schmidt-Shin (KPSS) Test Statistics of Eight Natural Gas Spot Prices in the Natural Logarithms for Each Sub-Period

	ADF		KPSS		ADF		KPSS		ADF		KPSS	
Price Series	t-Stat	Lag	LM-Stat	Bandwidth	t-Stat	Lag	LM-Stat	Bandwidth	t-Stat	Lag	LM-Stat	Bandwidth
	First Sub-Period				Second Sub-Period				Third Sub-Period			
	Test in Level				Test in Level				Test in Level			
AECO Hub	-1.2691	4	3.5719	32	-2.5797	2	1.5425	39	-3.5097	0	0.5890	29
Chicago	-2.7112	6	1.3266	32	-2.8717	2	1.4469	39	-2.8474	17	0.5324	29
Dominion South	-3.0740	1	1.15097	32	-2.6529	3	1.4842	39	-2.6445	15	1.2519	29
Henry Hub	-2.4714	2	1.6783	32	-2.6032	2	1.5293	39	-2.8248	2	0.6706	29
Malin	-1.7557	2	3.1405	32	-3.1826	10	0.9357	39	-3.3492	3	0.6193	29
Oneok	-1.8692	5	2.3805	32	-3.0795	10	1.0429	39	-3.8175	2	0.6242	29
Opal	-2.3996	1	3.0527	32	-3.5530	7	0.8975	39	-3.4813	3	0.5560	29
Waha Hub	-2.3850	3	2.3978	32	-3.0890	10	1.2055	39	-3.6006	4	0.6431	29
	Test in First Difference				Test in First Difference				Test in First Difference			
AECO Hub	-26.3915	3	0.1868	37	-43.2179	1	0.0336	21	-23.7516	2	0.1198	39
Chicago	-20.8860	7	0.1825	347	-40.6725	1	0.0385	35	-10.3670	16	0.0188	2
Dominion South	-22.8337	4	0.0716	38	-35.5044	2	0.0300	21	-8.1350	14	0.0637	30
Henry Hub	-31.9556	1	0.0750	15	-40.8216	1	0.0367	24	-31.3230	1	0.1320	29
Malin	-32.0032	1	0.1385	32	-15.8705	9	0.0344	39	-22.7060	3	0.1366	59
Oneok	-23.0913	4	0.1012	46	-16.0367	9	0.0385	45	-33.7464	1	0.0973	50
Opal	-28.9428	1	0.1029	18	-24.7092	6	0.0384	117	-27.2281	2	0.1017	48
Waha Hub	-29.8175	2	0.1091	56	-13.2521	14	0.0423	53	-24.0708	3	0.1159	60

Note: Under the null hypothesis of non-stationarity (unit root), the ADF test critical value at 1% level is -3.430; the null is rejected when t-Stat is less than the critical value. Under the null hypothesis of stationarity, the KPSS test critical value at 1% level is 0.739; the null is rejected when LM-stat is greater than the critical value. Lag (k) is selected from 0 to 20 based on Schwarz information criteria. Bandwidth is estimated using the Newey-West (1994) method.

Table B.2. Forecast Error Variance Decompositions of Eight Natural Gas Spot Prices Using the Entire Period (May 3, 1994 - October 31, 2014)

Horizon	AECO	Chicago	Dominion South	Henry Hub	Malin	Oneok	Opal	Waha Hub
				AECO				
1	68.16	0.72	0.00	22.66	4.62	3.85	0.00	0.00
5	61.01	0.72	0.37	25.60	8.08	3.97	0.06	0.19
10	58.36	0.42	0.38	26.75	9.63	4.21	0.05	0.20
				Chicago				
1	0.00	62.16	0.00	37.84	0.00	0.00	0.00	0.00
5	0.64	45.81	0.09	51.94	0.56	0.44	0.08	0.45
10	1.21	34.95	0.07	61.59	0.84	0.53	0.11	0.70
				Dominion South				
1	0.00	1.85	43.83	54.32	0.00	0.00	0.00	0.00
5	0.94	3.08	22.10	72.56	0.22	0.20	0.05	0.87
10	1.47	1.90	19.14	75.68	0.34	0.34	0.03	1.09
				Henry Hub				
1	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
5	1.74	0.19	0.22	96.61	0.44	0.40	0.04	0.37
10	2.53	0.12	0.12	95.46	0.67	0.71	0.03	0.37
				Malin				
1	0.00	1.46	0.00	30.57	60.19	7.79	0.00	0.00
5	1.92	1.01	0.27	31.07	58.55	6.93	0.05	0.21
10	2.96	0.59	0.25	33.18	55.65	7.17	0.05	0.15
				Oneok				
1	0.00	7.73	0.00	51.02	0.00	41.25	0.00	0.00
5	0.93	6.65	0.26	59.24	0.82	31.79	0.13	0.20
10	1.58	4.68	0.20	61.99	1.21	30.01	0.13	0.20
				Opal				
1	0.00	0.20	0.00	4.10	8.07	1.04	86.59	0.00
5	1.36	0.39	0.37	10.73	11.58	4.38	70.79	0.40
10	2.00	0.28	0.42	12.33	12.96	4.85	66.37	0.80
				Waha Hub				
1	0.00	2.87	0.00	58.62	0.98	15.30	0.00	22.23
5	1.42	3.57	0.35	67.63	2.01	16.35	0.08	8.59
10	2.22	2.43	0.29	69.96	2.37	17.16	0.07	5.51

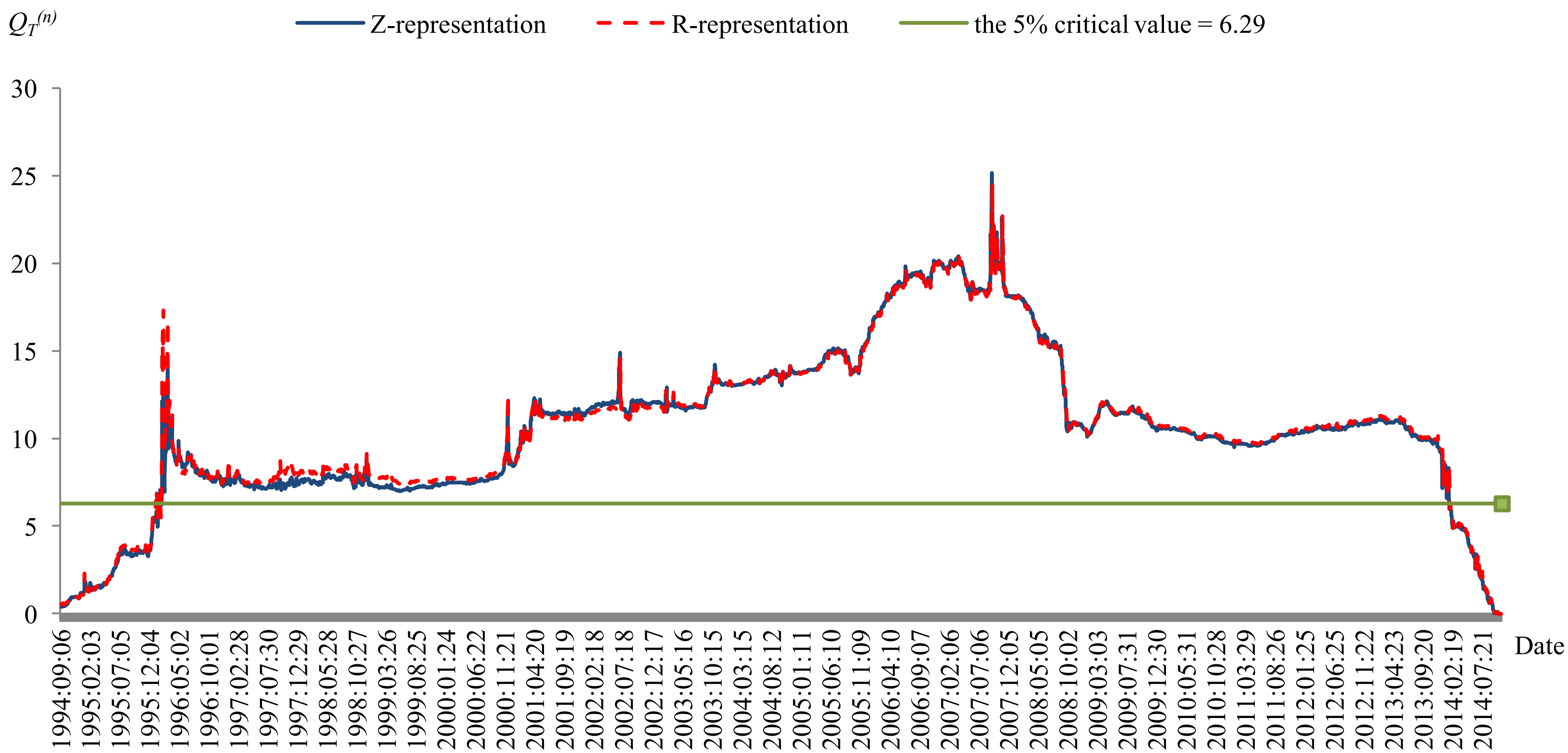

Figure B.1. Plots of sup $Q_T^{(n)}$ for the entire period (May 3, 1994 to October 31, 2014) with exogenous variables (daily degree-days)

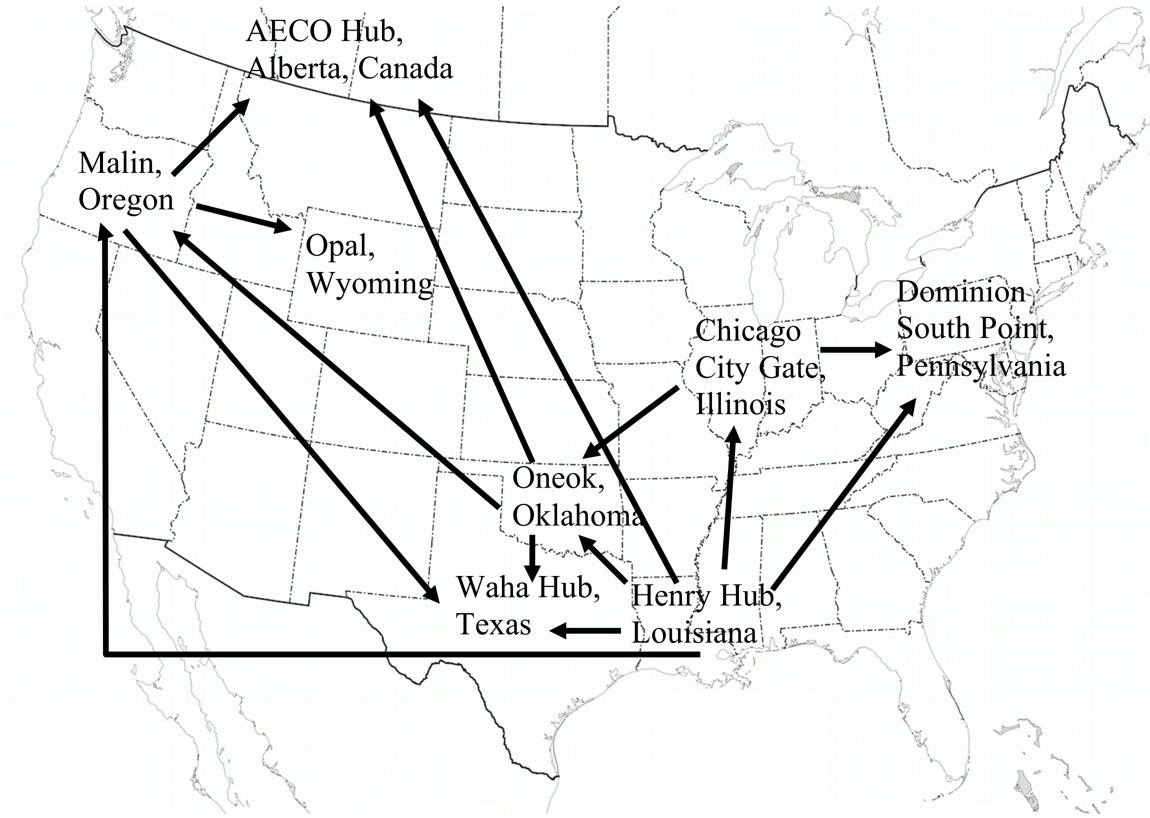

Figure B.2. Contemporaneous casual flows for the entire period (May 3, 1994 - October 31, 2014)

Figure B.3. Impulse response functions of eight natural gas spot prices for the entire period (May 3, 1994 - October 31, 2014)

APPENDIX C

LIST OF DATA CONSIDERED IN CHAPTER V

Table C.1. List of Variables considered with Their Transformation and Source

	Name	Transformation	Source
	Observable Variable		
1.	U.S. Natural Gas Gross Withdrawals (MMcf)	dln	U.S. EIA
	Carbon Dioxide Emissions		
2.	Petroleum Coke CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
3.	Residual Fuel Oil CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
4.	Other Petroleum Products CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
5.	Coal, Including Coal Coke Net Imports, CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
6.	Natural Gas, Excluding Supplemental Gaseous Fuels, CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
7.	Aviation Gasoline CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
8.	Distillate Fuel Oil, Excluding Biodiesel, CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
9.	Jet Fuel CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
10.	Kerosene CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
11.	LPG CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
12.	Lubricants CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA
13.	Motor Gasoline, Excluding Ethanol, CO2 Emissions (Million Metric Tons of Carbon Dioxide)*	ln	U.S. EIA

	Coal Consumption		
14.	Coal Imports (Thousand Short Tons)*	dln	U.S. EIA
15.	Coal Consumption for Electricity Generation, All Sectors (Thousand Short Tons)*	ln	U.S. EIA
16.	Coal Consumption (Quadrillion Btu)*	dln	U.S. EIA
	Coal Prices		
17.	Bloomberg Mid Sulfur Illinois Basin Coal Spot Price Fob*	dln	Bloomberg L.P.
18.	Bloomberg Low Sulfur Compliance Coal Spot Price/Big Sandy Barge Fob*	ln	Bloomberg L.P.
19.	Bloomberg Pennsylvania Railcar Seam Coal Spot Price Fob*	dln	Bloomberg L.P.
20.	Bloomberg 1% Sulfur Coal Spot Price Fob/Utah Colorado*	dln	Bloomberg L.P.
21.	Bloomberg Powder River Basin 8800 Btu Coal Spot Price Fob/Gillette Wyoming*	dln	Bloomberg L.P.
22.	Cost of Coal Receipts at Electric Generating Plants (Dollars per Million Btu, Including Taxes)*	dln	U.S. EIA
	Coal Production		
23.	Coal Exports (Thousand Short Tons)*	ln	U.S. EIA
24.	Coal Production (Quadrillion Btu)*	dln	U.S. EIA
	Coal Stocks		
25.	Coal Stocks, Producers and Distributors (Thousand Short Tons)*	dln	U.S. EIA
26.	Coal Stocks, End-Use Sectors Total (Thousand Short Tons)*	dln	U.S. EIA
27.	Coal Stocks, Electric Power Sector (Thousand Short Tons)*	dln	U.S. EIA
	Crude Oil and Petroleum Product Consumption		
28.	Distillate Fuel Oil Consumption for Electricity Generation, All Sectors (Thousand Barrels)*	ln	U.S. EIA

29.	Residual Fuel Oil Consumption for Electricity Generation, All Sectors (Thousand Barrels)*	ln	U.S. EIA
30.	Other Petroleum Liquids Consumption for Electricity Generation, All Sectors (Thousand Barrels)*	ln	U.S. EIA
31.	Petroleum Coke Consumption for Electricity Generation, All Sectors (Thousand Short Tons)*	ln	U.S. EIA
32.	Petroleum Consumption (Excluding Biofuels) (Quadrillion Btu)*	dln	U.S. EIA
33.	U.S. Imports of Crude Oil (Thousand Barrels)*	ln	U.S. EIA
34.	U.S. Imports of Gasoline Blending Components (Thousand Barrels)*	ln	U.S. EIA
35.	U.S. Imports of Finished Motor Gasoline (Thousand Barrels)*	ln	U.S. EIA
36.	U.S. Imports of Distillate Fuel Oil (Thousand Barrels)*	dln	U.S. EIA
37.	U.S. Imports of Residual Fuel Oil (Thousand Barrels)*	dln	U.S. EIA
38.	U.S. Imports of Petroleum Coke (Thousand Barrels)*	dln	U.S. EIA
	Crude Oil and Petroleum Product Prices		
39.	Europe Brent Spot Price FOB (Dollars per Barrel)*	dln	U.S. EIA
40.	New York Harbor Conventional Gasoline Regular Spot Price FOB (Dollars per Gallon)*	dln	U.S. EIA
41.	U.S. Gulf Coast Conventional Gasoline Regular Spot Price FOB (Dollars per Gallon)*	dln	U.S. EIA
42.	New York Harbor No. 2 Heating Oil Spot Price FOB (Dollars per Gallon)*	dln	U.S. EIA
43.	U.S. Gulf Coast Kerosene-Type Jet Fuel Spot Price FOB (Dollars per Gallon)*	dln	U.S. EIA
44.	Cushing, OK Crude Oil Future Contract 1 (Dollars per Barrel)*	dln	U.S. EIA
45.	Cushing, OK Crude Oil Future Contract 2 (Dollars per Barrel)*	dln	U.S. EIA
46.	Cushing, OK Crude Oil Future Contract 3 (Dollars per Barrel)*	dln	U.S. EIA
47.	Cushing, OK Crude Oil Future Contract 4 (Dollars per Barrel)*	dln	U.S. EIA

48.	New York Harbor No. 2 Heating Oil Future Contract 1 (Dollars per Gallon)*	dln	U.S. EIA
49.	New York Harbor No. 2 Heating Oil Future Contract 2 (Dollars per Gallon)*	dln	U.S. EIA
50.	New York Harbor No. 2 Heating Oil Future Contract 3 (Dollars per Gallon)*	dln	U.S. EIA
51.	New York Harbor No. 2 Heating Oil Future Contract 4 (Dollars per Gallon)*	dln	U.S. EIA
52.	Cushing, OK WTI Spot Price FOB (Dollars per Barrel)*	dln	U.S. EIA
53.	Cost of Distillate Fuel Receipts at Electric Generating Plants (Dollars per Million Btu, Including Taxes)*	dln	U.S. EIA
54.	Cost of Residual Fuel Receipts at Electric Generating Plants (Dollars per Million Btu, Including Taxes)*	dln	U.S. EIA
55.	Costs of Petroleum Coke Receipts at Electric Generating Plants (Dollars per Million Btu, Including Taxes)*	dln	U.S. EIA
	Crude Oil and Petroleum Product Production		
56.	Crude Oil Production, Total OPEC (Thousand Barrels per Day)*	dln	U.S. EIA
57.	U.S. Exports of Crude Oil (Thousand Barrels)*	ln	U.S. EIA
58.	U.S. Exports of Gasoline Blending Components (Thousand Barrels)*	dln	U.S. EIA
59.	U.S. Exports of Finished Motor Gasoline (Thousand Barrels)*	ln	U.S. EIA
60.	U.S. Exports of Distillate Fuel Oil (Thousand Barrels)*	ln	U.S. EIA
61.	U.S. Exports of Residual Fuel Oil (Thousand Barrels)*	ln	U.S. EIA
62.	U.S. Exports of Petroleum Coke (Thousand Barrels)*	ln	U.S. EIA
63.	Active Well Service Rig Count (Number of Rigs)*	ln	U.S. EIA
64.	Crude Oil Rotary Rigs in Operation (Number of Rigs)*	dln	U.S. EIA
65.	Crude Oil Production (Quadrillion Btu)*	dln	U.S. EIA

	Crude Oil and Petroleum Product Stocks		
66.	U.S. Ending Stocks of Crude Oil (Thousand Barrels)*	dln	U.S. EIA
67.	U.S. Ending Stocks of Total Gasoline (Thousand Barrels)*	ln	U.S. EIA
68.	U.S. Ending Stocks of Gasoline Blending Components (Thousand Barrels)*	ln	U.S. EIA
69.	U.S. Ending Stocks of Fuel Ethanol (Thousand Barrels)*	dln	U.S. EIA
70.	U.S. Ending Stocks of Distillate Fuel Oil (Thousand Barrels)*	dln	U.S. EIA
71.	U.S. Ending Stocks of Residual Fuel Oil (Thousand Barrels)*	ln	U.S. EIA
	Electricity Consumption		
72.	Electricity Retail Sales to the Residential Sector (Million Kilowatthours)*	ln	U.S. EIA
73.	Electricity Retail Sales to the Commercial Sector (Million Kilowatthours)*	ln	U.S. EIA
74.	Electricity Retail Sales to the Industrial Sector (Million Kilowatthours)*	ln	U.S. EIA
75.	Electricity Retail Sales to the Transportation Sector (Million Kilowatthours)*	dln	U.S. EIA
76.	Electricity Direct Use (Million Kilowatthours)*	dln	U.S. EIA
77.	Nuclear Electric Power Consumption (Quadrillion Btu)*	ln	U.S. EIA
	Electricity Net generation		
78.	Electricity Net Generation From Geothermal, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
79.	Electricity Net Generation From Solar/PV, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
80.	Electricity Net Generation From Wind, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
81.	Electricity Net Generation From Coal, All Sectors (Million Kilowatthours)*	ln	U.S. EIA

82.	Electricity Net Generation From Petroleum, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
83.	Electricity Net Generation From Natural Gas, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
84.	Electricity Net Generation From Other Gases, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
85.	Electricity Net Generation From Nuclear Electric Power, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
86.	Electricity Net Generation From Conventional Hydroelectric Power, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
87.	Electricity Net Generation From Wood, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
88.	Electricity Net Generation From Waste, All Sectors (Million Kilowatthours)*	ln	U.S. EIA
89.	Other Consumption for Electricity Generation, All Sectors (Trillion Btu)*	ln	U.S. EIA
90.	Other Gases Consumption for Electricity Generation, All Sectors (Trillion Btu)*	dln	U.S. EIA
91.	Nuclear Electric Power Production (Quadrillion Btu)*	ln	U.S. EIA
	Electricity Prices		
92.	Wholesale electricity spot price- Mid-C	ln	Bloomberg L.P.
93.	Wholesale electricity spot price- Southwest- Pola Verde	ln	Bloomberg L.P.
94.	Wholesale electricity spot price- Northern California (NP15)	ln	Bloomberg L.P.
95.	Average Retail Price of Electricity, Commercial (Cents per Kilowatthour, Including Taxes)*	ln	U.S. EIA
96.	Average Retail Price of Electricity, Industrial (Cents per Kilowatthour, Including Taxes)*	ln	U.S. EIA

97.	Average Retail Price of Electricity, Residential (Cents per Kilowatthour, Including Taxes)*	ln	U.S. EIA
98.	Wholesale electricity spot price - New England Mass Hub	ln	U.S. EIA
99.	Wholesale electricity spot price - PJM West	ln	U.S. EIA
100.	Wholesale electricity spot price- Southern California (SP15)	ln	U.S. EIA
101.	Uranium, u3o8 restricted price, Nuexco exchange spot, US Dollars per Pound*	dln	www.indexmundi.com
	Macroeconomic Variables		
102.	1-Year Treasury Constant Maturity Rate (GS1)*	dln	Federal Reserve Bank of St. Louis
103.	3-Month Treasury Bill: Secondary Market Rate (TB3MS)*	dln	Federal Reserve Bank of St. Louis
104.	Moody's Seasoned Aaa Corporate Bond Yield©*	dln	Federal Reserve Bank of St. Louis
105.	Moody's Seasoned Baa Corporate Bond Yield©*	dln	Federal Reserve Bank of St. Louis
106.	Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL)*	dln	Federal Reserve Bank of St. Louis
107.	Consumer Price Index for All Urban Consumers: Energy (CPIENGSL)*	dln	Federal Reserve Bank of St. Louis
108.	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy (CPILFESL)*	dln	Federal Reserve Bank of St. Louis
109.	Canada / U.S. Foreign Exchange Rate (EXCAUS)*	dln	Federal Reserve Bank of St. Louis
110.	U.S. / Euro Foreign Exchange Rate*	dln	Federal Reserve Bank of St. Louis
111.	U.S. / U.K. Foreign Exchange Rate*	dln	Federal Reserve Bank of St. Louis
112.	Effective Federal Funds Rate (FEDFUNDS)*	dln	Federal Reserve

			Bank of St. Louis
113.	Monthly Real GDP Index*	dln	Federal Reserve Bank of St. Louis
114.	Industrial Production Index (INDPRO)*	dln	Federal Reserve Bank of St. Louis
115.	New York Stock Exchange Composite Index*	dln	Federal Reserve Bank of St. Louis
116.	Real Disposable Personal Income: Per capita*	dln	Federal Reserve Bank of St. Louis
117.	S&P500 Index - Last Price*	dln	Federal Reserve Bank of St. Louis
118.	S&P500 index - Volumn*	dln	Federal Reserve Bank of St. Louis
119.	Civilian Unemployment Rate (UNRATE)*	dln	Federal Reserve Bank of St. Louis
	Natural Gas Consumption		
120.	Natural Gas Consumption for Electricity Generation, All Sectors (Billion Cubic Feet)*	ln	U.S. EIA
121.	U.S. Natural Gas Lease and Plant Fuel Consumption (MMcf)*	dln	U.S. EIA
122.	U.S. Natural Gas Pipeline & Distribution Use (MMcf)*	ln	U.S. EIA
123.	U.S. Natural Gas Residential Consumption (MMcf)*	ln	U.S. EIA
124.	Natural Gas Deliveries to Commercial Consumers (Including Vehicle Fuel through 1996) in the U.S. (MMcf)*	ln	U.S. EIA
125.	U.S. Natural Gas Industrial Consumption (MMcf)*	ln	U.S. EIA
126.	U.S. Natural Gas Vehicle Fuel Consumption (MMcf)*	dln	U.S. EIA
127.	U.S. Natural Gas Deliveries to Electric Power Consumers (MMcf)*	ln	U.S. EIA
128.	U.S. Liquefied Natural Gas Imports (MMcf)*	dln	U.S. EIA
129.	U.S. Natural Gas Pipeline Imports (MMcf)	dln	U.S. EIA

130.	Natural Gas Consumption (Excluding Supplemental Gaseous Fuels) (Quadrillion Btu)*	ln	U.S. EIA
	Natural Gas Prices		
131.	Natural Gas Spot Market- Algonquin	ln	Bloomberg L.P.
132.	Natural Gas Spot Market- AECO	ln	Bloomberg L.P.
133.	Natural Gas Spot Market- Chicago	ln	Bloomberg L.P.
134.	Natural Gas Spot Market- Dominion South	dln	Bloomberg L.P.
135.	Natural Gas Spot Market- Oneok	ln	Bloomberg L.P.
136.	Natural Gas Spot Market- Waha	ln	Bloomberg L.P.
137.	Natural Gas Spot Market- Malin	ln	Bloomberg L.P.
138.	Natural Gas Spot Market- Opal	ln	Bloomberg L.P.
139.	Natural Gas Spot Market-TETCO M3	ln	Bloomberg L.P.
140.	Cost of Natural Gas Receipts at Electric Generating Plants (Dollars per Million Btu, Including Taxes)*	dln	U.S. EIA
141.	Natural Gas Futures Contract 1 (Dollars per Million Btu)	dln	U.S. EIA
142.	Natural Gas Futures Contract 2 (Dollars per Million Btu)	dln	U.S. EIA
143.	Natural Gas Futures Contract 3 (Dollars per Million Btu)	dln	U.S. EIA
144.	Natural Gas Futures Contract 4 (Dollars per Million Btu)	dln	U.S. EIA
145.	U.S. Price of Natural Gas Delivered to Residential Consumers (Dollars per Thousand Cubic Feet)	dln	U.S. EIA
146.	U.S. Price of Natural Gas Sold to Commercial Consumers (Dollars per Thousand Cubic Feet)	dln	U.S. EIA
147.	United States Natural Gas Industrial Price (Dollars per Thousand Cubic Feet)	dln	U.S. EIA
148.	U.S. Natural Gas Citygate Price (Dollars per Thousand Cubic Feet)	dln	U.S. EIA
149.	Price of U.S. Natural Gas LNG Imports (Dollars per Thousand Cubic Feet)	dln	U.S. EIA

150.	U.S. Natural Gas Pipeline Imports Price (Dollars per Thousand Cubic Feet)	dln	U.S. EIA
151.	Price of Liquefied U.S. Natural Gas Exports (Dollars per Thousand Cubic Feet)	ln	U.S. EIA
152.	Price of U.S. Natural Gas Pipeline Exports (Dollars per Thousand Cubic Feet)	ln	U.S. EIA
153.	Henry Hub Natural Gas Spot Price (Dollars per Million Btu)	dln	U.S. EIA
	Natural Gas Production		
154.	Natural Gas Rotary Rigs in Operation (Number of Rigs)*	dln	U.S. EIA
155.	Liquefied U.S. Natural Gas Exports (MMcf)*	dln	U.S. EIA
156.	U.S. Natural Gas Pipeline Exports (MMcf)	dln	U.S. EIA
157.	Natural Gas Plant Liquids Production (Quadrillion Btu)*	dln	U.S. EIA
158.	Capacity Utilization % ANR Pipeline Co-Delivery	none	Velocity Suite Online
159.	Capacity Utilization % Columbia Gas Transmission Corp-Delivery	none	Velocity Suite Online
160.	Capacity Utilization % ANR Pipeline Co-Receipt	dln	Velocity Suite Online
161.	Capacity Utilization % Columbia Gas Transmission Corp-Receipt	none	Velocity Suite Online
	Natural Gas Storage		
162.	AGA Producing Region Natural Gas Underground Storage Volume (MMcf)	ln	U.S. EIA
163.	AGA Eastern Consuming Region Natural Gas Underground Storage Volume (MMcf)	ln	U.S. EIA
164.	AGA Western Consuming Region Natural Gas Underground Storage Volume (MMcf)	ln	U.S. EIA
165.	Total Natural Gas Underground Storage Capacity (MMcf)	dln	U.S. EIA

	Renewable Energy Consumption		
166.	Wood Consumption for Electricity Generation, All Sectors (Trillion Btu)*	dln	U.S. EIA
167.	Hydroelectric Power Consumption (Quadrillion Btu)*	ln	U.S. EIA
168.	Geothermal Energy Consumption (Quadrillion Btu)*	ln	U.S. EIA
169.	Solar/PV Energy Consumption (Quadrillion Btu)*	dln	U.S. EIA
170.	Wind Energy Consumption (Quadrillion Btu)*	ln	U.S. EIA
171.	Biomass Energy Consumption (Quadrillion Btu)*	ln	U.S. EIA
	Renewable Energy Production		
172.	Waste Consumption for Electricity Generation, All Sectors (Trillion Btu)*	dln	U.S. EIA
173.	Biomass Energy Production (Quadrillion Btu)*	dln	U.S. EIA
174.	Hydroelectric Power Production (Quadrillion Btu)*	dln	U.S. EIA
175.	Geothermal Energy Production (Quadrillion Btu)*	dln	U.S. EIA
176.	Solar/PV Energy Production (Quadrillion Btu)*	dln	U.S. EIA
177.	Wind Energy Production (Quadrillion Btu)*	ln	U.S. EIA
	Weather		
178.	Cooling Degree Day*	ln	U.S. EIA
179.	Heating Degree Day*	ln	U.S. EIA

Note: The asterisk '*' indicates slow-moving variables. 'ln' refers to natural logarithms. 'dln' refers to first difference of natural logarithms. 'none' refers to no transformation.
Energy-related data are retrieved from the U.S. EIA data browser varying on energy sources and the Velocity Suite Online database.
Data of macroeconomics variables are obtained from the Federal Reserve Bank of St. Louis.

APPENDIX D

SUPPLEMENTARY RESULTS FOR CHAPTER V

Table D.1. Test Statistics on Probability Forecasts from the FAVAR(5F) Model

Series	Chi-Squared	Root Mean-Squared Error	Brier Score	*Var(d)*	*MinVar*	*Scat(f)*	*Bias*^{2}	*Cov(f, d)*	Ranked Probability Score
Factor 1	21.3478	0.0311	0.4722	0.5104	0.0364	0.1105	0.0197	0.1025	0.0540
Factor 2	16.0000	0.0587	0.8912	0.8194	0.0018	0.0580	0.0585	0.0233	0.1447
Factor 3	18.0000	0.1025	0.9100	0.8646	0.0165	0.1186	0.0215	0.0556	0.1883
Factor 4	18.0000	0.1115	0.9333	0.8611	0.0015	0.0788	0.0246	0.0164	0.2026
Factor 5	28.3043	0.0858	1.0086	0.8368	0.0080	0.1622	0.0619	0.0301	0.2140
Natural Gas Gross Withdrawals	10.9130	0.7660	0.8237	0.8819	0.0463	0.1211	0.0184	0.1220	0.1418
System		0.3218	0.8398	0.8798	0.0321	0.1349	0.0067	0.1068	0.1576

Note: The null hypothesis of well calibration cannot be rejected if the chi-squared test statistic is less than the 5% critical value of $\chi^2(19)$ =30.144.

Brier Score = *Var(d)* + *MinVar(f)* + *Scat(f)* + *Bias*^{2} – 2*Cov*(f, d).

Table D.2. Test Statistics on Probability Forecasts from the FAVAR(10F) Model

Series	Chi-Squared	Root Mean-Squared Error	Brier Score	*Var(d)*	*MinVar*	*Scat(f)*	*Bias*^{2}	*Cov(f, d)*	Ranked Probability Score
Factor 1	12.6522	0.0384	0.5510	0.5104	0.0102	0.1341	0.0174	0.0606	0.0669
Factor 2	19.3333	0.0686	0.9526	0.8264	0.0010	0.0796	0.0524	0.0034	0.1623
Factor 3	43.4545	0.1034	0.9155	0.8576	0.0091	0.1061	0.0270	0.0421	0.1814
Factor 4	39.8182	0.1347	0.9402	0.8681	0.0080	0.1590	0.0212	0.0580	0.2450
Factor 5	18.0000	0.0894	1.0478	0.8160	0.0054	0.1769	0.0855	0.0179	0.2288
Factor 6	14.3913	0.0657	0.8384	0.8368	0.0226	0.1352	0.0604	0.1083	0.1217
Factor 7	9.1739	0.0754	0.9642	0.8715	0.0007	0.0703	0.0311	0.0047	0.1844
Factor 8	10.9130	0.0925	0.9552	0.8854	0.0029	0.0537	0.0193	0.0031	0.2168
Factor 9	11.0000	0.1312	0.9789	0.8819	0.0104	0.1277	0.0386	0.0399	0.2099
Factor 10	17.8696	0.1160	1.0216	0.8715	0.0040	0.0878	0.0458	-0.0062	0.2465
Natural Gas Gross Withdrawals	25.2727	1.0699	0.9105	0.8819	0.0372	0.1795	0.0211	0.1046	0.1879
System		0.3353	0.9160	0.8916	0.0140	0.1389	0.0076	0.0680	0.1865

Note: The null hypothesis of well calibration cannot be rejected if the chi-squared test statistic is less than the 5% critical value of $\chi^2(19) = 30.144$.
Brier Score = *Var(d)* + *MinVar(f)* + *Scat(f)* + *Bias*^{2} – 2*Cov*(f, d).

Table D.3. Test Statistics on Probability Forecasts from the VAR(5) Model

Series	Chi-Squared	Root Mean-Squared Error	Brier Score	*Var(d)*	*MinVar*	*Scat(f)*	$Bias^2$	*Cov(f, d)*	Ranked Probability Score
Natural Gas Gross Withdrawals	17.6667	0.6956	0.8358	0.8819	0.0265	0.0710	0.0279	0.0858	0.1237
Henry Hub Natural Gas Spot Prices	11.0000	0.7181	0.8893	0.8472	0.0001	0.0047	0.0243	-0.0064	0.1662
Natural Gas Consumption	19.3333	0.4123	0.8423	0.7535	0.0419	0.1909	0.0599	0.1019	0.1754
WTI Crude Oil Spot price	17.8696	0.8369	0.8830	0.8576	0.0023	0.0204	0.0298	0.0136	0.1424
S&P500	19.3333	0.6431	0.8956	0.8368	0.0001	0.0156	0.0441	0.0006	0.1274
System		0.6758	0.8692	0.8790	0.0138	0.0734	0.0093	0.0532	0.1470

Note: The null hypothesis of well calibration cannot be rejected if the chi-squared test statistic is less than the 5% critical value of $\chi^2(19) = 30.144$.

Brier Score = $Var(d) + MinVar(f) + Scat(f) + Bias^2 - 2Cov(\mathrm{f}, \mathrm{d})$.