Pricing of Risk in Natural Gas Futures Contracts: A New Approach to Affine Term Structure Models Subject to Changes in Regime

Irina Zhecheva*

Department of Economics
University of California, San Diego

November 18, 2016

Latest version available here

Abstract

Commodities alternate between periods of contango, when the near month futures trade at a discount to back month futures, and backwardation, when near month futures are priced higher than back month futures. Market participants (hedgers and speculators) are exposed to substantial risk due to the possibility of the market switching from one state to the other. I provide a new way to characterize these risks in commodities futures markets. I apply this framework to the natural gas futures market where these risks are particularly substantial, and study the consequences of changes in regime on the risk premium. Motivated by the historically observed switches of the natural gas market between states of contango and backwardation, I propose and estimate a Markov regime-switching Gaussian affine term structure model and estimate the states of the market from the data. I find very strong evidence that there are regimes in the data, using a dataset of prices from 1995 to 2014. Moreover, the regimes correspond precisely to historically observed periods of contango and backwardation. I find that the market acts as if regimes are more persistent than they really are, which could be a result of hedging pressure. Moreover, I find that regime switching risk is priced. Calculating expected returns conditional on each regime confirms the claim that agents face significant risks from the possibility of a change in the regime. The maximum likelihood based methods common in the literature for estimating Gaussian affine term structure models with regime switching pose significant numerical challenges. A separate contribution of this paper is to propose a new approach to estimating this type of models. My approach avoids these numerical difficulties and allows for computationally fast estimation.

*Department of Economics, UCSD, 9500 Gilman Drive, La Jolla CA 92093-0508. The author can be reached at izhechev@ucsd.edu. I thank my advisors James Hamilton, Allan Timmermann, Rossen Valkanov, Ivana Komunjer, and Ruth Williams for their invaluable guidance and support. This paper also benefited from helpful comments from Johannes Wieland, David Lagakos, Julie Cullen, Mark Machina, Yixiao Sun, Asad Dossani, Leland Farmer, and seminar participants at UCSD.
1 Introduction

Commodities alternate between periods when the spread between the longer term and shorter term futures contracts is positive and periods when it is negative. When the spread between the longer term and the shorter term contracts is positive, commodities futures are said to be in contango, whereas when that spread is negative, they are said to be in backwardation. Participants in commodities futures markets are exposed to substantial risks due to the possibility of the market switching from one state to the other. This paper provides a new way to characterize these risks in commodities futures markets using a regime switching approach.

As futures contracts approach expiry, investors who want to stay long in a given contract roll their position, i.e. sell their position and buy the next month contract. In backwardation, investors can often earn a positive return by rolling their position, i.e. selling out of the contract at a higher price than what they initially paid\(^1\), and reinvesting at a lower price. That is why a common trading strategy in backwardation is to be long in a futures contract and roll your position each month. In contango, investors sell out of the contract at a lower price than what they initially paid\(^2\). Hence, a common trading strategy in contango is to be short in a futures contract and roll your position each month. As long as the market continues to be in backwardation, an investor may be getting a positive return by being long, but if the market unexpectedly switches to contango, the investor would likely experience a significant loss. Therefore, investors in commodities futures face significant risk due to the possibility of the market switching from one state to the other, and it is crucial to be able to determine the turning points.

Natural gas futures prices are very volatile, and the risks faced by market participants due to the possibility of switching from one state to the other are particularly substantial. Thus, the natural gas market is a useful setting for studying this type of risk. Natural gas is one of the most heavily traded commodity futures contracts in the United States. The natural gas futures market is highly liquid with daily open interest of up to about 300,000 contracts for the one month contract and total open interest of up to 1,400,000 contracts. The second contribution of this paper is to characterize these risks in the natural gas futures market. The method I develop is particularly well suited for identifying these risks in this setting. Motivated by the historically observed switches of natural gas futures prices between states of contango and backwardation, I propose and estimate a Markov regime-switching

\(^1\)This is the case if the term structure of futures prices has approximately the same level and slope at time \(t\) when the contract was bought and time \(t + 1\) when investors sell out of the contract.

\(^2\)This is the case if the term structure of futures prices has approximately the same level and slope at time \(t\) when the contract was bought and time \(t + 1\) when investors sell out of the contract.
Gaussian affine term structure model and estimate the states of the market from the data. In my model, the level and volatility of natural gas futures are allowed to switch with the regime. The risk premium is also regime dependent. I choose a model with two regimes in order to be able to see how the regimes estimated from the data are related to the two states of contango and backwardation observed in the market. I find that the estimated regimes correspond exactly to periods of contango and backwardation observed in the data. Moreover, I find very strong statistical evidence that there are regimes in the data. I also find that regime shift risk is priced.

I provide empirical estimates to characterize risk premia and the term structure of natural gas futures. I find that expected futures returns change sign depending on which regime we are conditioning on. For example, I find that in the backwardation regime, an investor with a long position in the 9-month contract would on average earn a positive expected monthly holding return of 2.5%. However, in the contango regime, the investor would earn a negative expected monthly holding return of about -0.5% on average. Thus, as long as the market is in backwardation, the investor would on average profit from being long in the 9-month contract. However, if the regime switches to contango, the investor would experience a substantial loss. This is consistent with the claim that agents face significant risks from the possibility of a change in the regime.

The third contribution of this paper is methodological. I propose a new approach to the estimation of Markov regime-switching Gaussian affine term structure models, which solves the numerical issues encountered by other estimation methods in the literature. Most methods for estimating Gaussian affine term structure models in the literature are based on maximum likelihood and exploit both the distributional assumptions and the no-arbitrage restrictions. These methods estimate latent pricing factors jointly with the model parameters by explicitly imposing the cross-equation no-arbitrage restrictions. As a result, these methods rely on numerical optimization, are computationally intensive, and pose significant numerical challenges, which become especially severe when there is the possibility of changes in regime. Estimation difficulties commonly arise due to highly non-linear and badly behaved likelihood surfaces, which are flat along many directions of the parameter space, and it is hard to achieve convergence. These problems can make estimation by MLE very difficult or infeasible. These difficulties have been documented by multiple researchers such as Kim and Orphanides (2005), Duffee (2002), Ang and Piazzesi (2003), Kim (2008), Duffee and Stanton (2008), Duffee (2009), and Ang and Bekaert (2002). To facilitate estimation of my model, I propose a new estimation method for Markov regime-switching affine term structure models. I use a regression based method to estimate the reduced-form parameters in the first stage.

---

3This holds for all contracts except for the 6-month contract.
and then estimate the prices of risk and risk-neutral transition probabilities via minimum-
chi-square estimation in the second stage. The no-arbitrage restrictions are not used or
imposed in the first-stage reduced-form estimation, but are exploited in the second stage of
the estimation. The minimum-chi-square procedure chooses the values of the structural pa-
rameters so that the values for the recursive pricing parameters implied by the no-arbitrage
restrictions most closely fit the unrestricted first-stage estimates. This approach is asymp-
totically equivalent to full information maximum likelihood. The numerical component in
the second stage is far simpler to implement than the one associated with other maximum
likelihood based methods in the literature. In this way, I bypass the numerical difficulties
encountered by other methods and have no problem achieving convergence. Another advan-
tage of the minimum-chi-square approach is that the value of the objective function provides
a way to test whether the no-arbitrage restrictions are consistent with the data. I show how
to price the time series and cross-section of the term structure of commodities futures prices
in the case of regime switching, and apply my approach to estimate my model.

The class of Gaussian affine term structure models was originally developed by Vasicek
to characterize the relation between yields on bonds of different maturities. The Gaussian
affine term structure framework is based on the assumptions that the pricing kernel is expo-
nentially affine, prices of risk are affine in the state variables, and innovations to the state
variables are conditionally Gaussian. Under these assumptions, the price process is affine
in the state variables, and no-arbitrage restrictions constrain the coefficients on the state
variables.

Hamilton and Wu (2014) adapt this class of models to commodities. They show that an
affine factor structure of commodity futures prices can result from the interaction between
arbitrageurs and commercial producers seeking hedges or financial investors seeking diver-
sification. Schwartz (1997), Schwartz and Smith (2000), and Casassus and Collin-Dufresne
(2006), among others, also develop related models to describe commodity futures prices.

Regime-switching models for the term structure of interest rates have been proposed and
estimated by Dai, Singleton, and Yang (2007), Bansal and Zhou (2002), and Ang and Bekaert
maximum-likelihood based methods based on an iterative procedure developed by Hamilton
methods are subject to the numerical issues mentioned earlier, which this paper resolves.

Almansour (2016) models the futures term structure of crude oil and natural gas using a
convenience yield model.\(^4\) He extends the two-factor stochastic convenience yield model of

\(^4\)Convenience yield is the benefit derived from holding the underlying physical commodity rather than
Gibson and Schwartz (1990) to allow the convenience yield level as well as other parameters to be regime dependent. He uses a two-factor model with the log of the spot price and the convenience yield as latent factors. The main advantages of my framework over his is that my framework gives a clear way to test the underlying model assumptions, while Almansour’s does not. I provide results of a test of the underlying assumptions, and show that they are satisfied. Moreover, I provide tests of a number of hypotheses about the differences in regime, which he does not. Furthermore, Almansour does not allow the average values of the historic regime-switching transition probabilities (which he assumes to be time-varying) to differ from the risk-neutral transition probabilities, and I show that is key for risk pricing in the natural gas futures market. Another advantage of my paper is that I discuss the implications for investment strategy and hedging, whereas he does not.

Adrian, Crump and Moench (2013) and Diez de los Rios (forthcoming) have recently proposed regression based algorithms for estimation of single-regime affine term structure models that avoid the numerical difficulties associated with maximum likelihood estimation. The approach I propose in this paper allows for regime-switching and combines regression based and numerical calculations.

1.1 Background on the natural gas market

Natural gas accounts for 30% of U.S. electricity generation and is predicted to account for an even larger proportion in the future. This commodity has gained a lot of attention in recent years with the discovery of shale gas and the advancement of drilling technology.

Some of the potential reasons for the existence of the two market states of contango and backwardation are shortages, weather concerns, or geopolitical events. Contango can result from trader perceptions of future shortage or short-term supply glut, whereas backwardation can result from near term shortage or future supply glut. For example, if there is a hurricane that is expected to disrupt supply only for a short time, the near term futures may spike while the longer term futures can be relatively unaffected. This unanticipated weather event would induce a period of backwardation.

The transitions from one market state to another are in part due to seasonal changes in supply and demand conditions. Natural gas production is relatively constant throughout the year, but consumption tends to peak during the winter heating season (November through March) as home heating use rises and tends to be moderate in other seasons. In general, there is often a shortage of natural gas in the winter as supply is often unable to react quickly to short-term increases in demand, so winter months are often characterized by backwardation.
At the same time, the inventory stored for the winter can help to meet demand, and if there is no current shortage, the market may not be in backwardation in the winter. In the remaining seasons when supply and demand are in balance or there is short-term oversupply, the market can often be in contango. Other factors affecting supply and demand can also affect whether the market is in contango or backwardation. On the supply side, amount of gas in storage, pipeline capacity, and imperfect information about storage can affect the state of the market.

Most of U.S. natural gas consumption is from domestic production. U.S. dry production has been steadily increasing since 2006. It reached its highest recorded annual total in 2015 and is still rising during 2016. The increases in production were due to more efficient, cost-effective drilling techniques which have allowed for horizontal drilling in shale formations. This has led to an unprecedented surge in supply. As a result, since around 2009, the natural gas market has been in contango. This observation is captured by my model, which classifies the data sample post 2009 as being in the contango regime.

The fluctuations in the amount of natural gas in storage and the expectations of future oversupply or shortage can cause the market to switch between contango and backwardation. Although the fluctuations of prices between periods of contango and periods of backwardation have a seasonal component, they are far from deterministic. In some years there is virtually no shortage of natural gas in the winter, while in others there could be shortage well beyond the winter months. Thus, backwardation does not always occur in the winter, and other seasons are not always characterized by contango. The duration of episodes of contango and backwardation also varies from year to year. The type of nondeterministic seasonality observed in the natural gas market cannot be fully captured by seasonal dummy models or other methods for modeling deterministic seasonality. To capture these historically observed fluctuations, I model the level and volatility of natural gas futures using a Markov regime-switching affine term structure model and estimate the states of the market from the data.

The uncertainty about gas prices introduces the possibility that commercial producers or commercial users may at times make significant use of natural gas futures contracts for purposes of hedging. If commercial producers believe that the price of natural gas may fall in the future, they can take a short position in futures to secure a higher selling price for the natural gas they produce. Conversely, if commercial users think that the price of natural gas may rise in the future, they can take a long position in futures to lock in a lower price for natural gas. Natural gas futures are also traded by speculators, who are willing to assume the price risk that hedgers try to avoid in return for a risk premium. I find evidence that commercial users and commercial producers use natural gas futures contracts for purposes of hedging. I find that in the contango regime, commercial producers may be
trying to hedge their short positions in natural gas by selling 4-6 month contracts, while in the backwardation regime, commercial producers may be trying to hedge their short positions by selling 6-9 month contracts. I obtain analogous implications for commercial users. Hamilton and Wu (2014) show how variation in hedging pressure could influence the term structure of commodity futures prices. I study the consequences of changes in regime on the risk premium by generalizing the futures-pricing model in Hamilton and Wu (2014) to allow for changes in regime.

The rest of the paper is organized as follows. Sections 2 and 3 present the model framework, Section 4 describes the estimation approach, Section 5 gives empirical results for a one factor model of natural gas futures prices from 1995 to 2014, and Section 6 concludes.

2 Model

Let \( F_t^{(n)} \) be the price of a futures contract with maturity \( n \) at time \( t \). I assume that the log of the price is a function of a factor \( X_t \) that follows a Gaussian autoregression:

\[
X_{t+1} = \mu_{s_t} + \Phi X_t + v_{t+1}, \quad v_{t+1} | s_t \sim N(0, \sigma_{s_t}^2)
\]  

(1)

I find that a one factor model using the first principal component of the log of the futures prices with maturities 3 months to 9 months as a factor describes the data well.\(^5\) Thus, in my empirical application \( X_t \) will be a scalar. I find that shorter term contracts have a systematically different behavior and do not fit well in my framework together with the long term contracts.

The intercept parameter is regime switching, and \( s_t \) denotes the regime at time \( t \), \( s_t \in \{1, 2\} \). I assume that the slope parameter \( \Phi \) is regime independent.\(^6\)

The no-arbitrage assumption implies the existence of a pricing kernel \( M_{t,t+1} \) such that

\[
F_t^{(n)} = E_t^p \left[ M_{t,t+1} F_{t+1}^{(n-1)} | s_t = j \right]
\]

(2)

if the regime at time \( t \) is \( j \). Following Dai, Singleton, and Yang (2007), I assume that the pricing kernel is exponentially affine and takes the following form:

\[
M_{t,t+1} = \exp \left[ -\Gamma_{s_t,s_{t+1}} - \frac{1}{2} \lambda_{t,s_t}^2 - \lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right]
\]

(3)

\(^5\)In section 5.2 I provide some evidence on why a one factor model is appropriate.

\(^6\)I also estimated a version of the model with \( \Phi \) allowed to vary with regime, but found that this led to only a trivial increase in the log likelihood for the system in equations (14)-(17) that I estimate. Since the equations are simpler and more intuitive with \( \Phi \) constant, I only discuss the simpler case in this paper.
where $\Gamma_{st, st+1}$ is the market price of regime shift risk, $\lambda_{t, st}$ is the market price of factor risk, and $v_{t+1}$ is the innovation from the factor equation (1). $\Gamma_{st, st+1}$ is referred to as the market price of regime shift risk because it can be interpreted as the log expected return per unit of regime shift risk exposure as I show in equation (29) in section (5.2). I also assume that the market price of factor risk $\lambda_{t, st}$ is an affine function of the state variable:

$$\lambda_{t, st} = \sigma_{st}^{-1} (\lambda_{0, st} + \lambda_1 X_t)$$  \hspace{1cm} (4)

Here the market price of factor risk $\lambda_{t, st}$ is time-varying and regime-dependent. I assume that $\lambda_1$ does not depend on the regime.\(^7\) No arbitrage implies the existence of an equivalent martingale measure - the risk neutral measure $\mathbb{Q}$. The historic measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$ are related through the pricing kernel $M_{t, t+1}$. The price $P(X_t)$ of an asset with payoff $g(X_{t+1})$ in regime $j$ can be computed as

$$P(X_t) = E_t^P [M_{t, t+1} g(X_{t+1}) | s_t = j] = E_t^Q [g(X_{t+1}) | s_t = j]$$  \hspace{1cm} (5)

Under the $\mathbb{Q}$-measure, the factor $X_t$ follows a Gaussian autoregression:

$$X_{t+1} = \mu_{st}^Q + \Phi^Q X_t + v_{t+1}^Q$$  \hspace{1cm} (6)

where

$$\mu_j^Q = \mu_j - \lambda_{0, j}$$  \hspace{1cm} (7)

and

$$\Phi^Q = \Phi - \lambda_1$$  \hspace{1cm} (8)

and $v_{t+1}^Q|s_t = j \sim \mathcal{N}(0, \sigma_j^2)$. The above relations are derived in Appendix A.

Let $f_t^{(n)} \equiv \ln F_t^{(n)}$. Equations (1), (2), and (3) together imply that the log of the futures price is affine in the state variable:

$$f_t^{(n)} = A_{st}^{(n)} + B^{(n)} X_t + u_t^{(n)}$$  \hspace{1cm} (9)

From equation (8) and equation (A-10), it follows that the factor loadings $B^{(n)}$ are regime independent. This ensures exact closed-form solutions for the futures prices, and is consistent with Dai, Singleton, and Yang (2007). The intercept term is allowed to change with the regime.

\(^7\)When I estimated a version of the model in which $\Phi$ varies with the regime, I also allowed $\lambda_1$ to vary with the regime, denoting it $\lambda_{1, st}$. I failed to reject the hypothesis that $\lambda_{1, 1} - \lambda_{1, 2} = 0$. Hence, I consider the simpler model in which $\lambda_1$ is regime independent.
Let $r_{x_t+1}^{(n-1)}$ denote the one-month log holding return of a futures contract maturing in $n$ months:

$$r_{x_t+1}^{(n-1)} = f_{t+1}^{(n-1)} - f_t^{(n)}$$

(10)

The holding return is the return on buying an $n$-month futures contract in month $t$ and then selling it as an $(n-1)$-month futures contract in month $t + 1$.

In my application, I assume that there are 2 regimes that govern the dynamic properties of the factor $X_t$. The unobserved regime variable $s_t$ is presumed to follow a 2-state Markov chain, with the risk-neutral probability of switching from regime $s_t = j$ to regime $s_{t+1} = k$ given by $\pi_{Qjk}, 1 \leq j, k \leq 2$, with $\sum_{k=1}^{2} \pi_{Qjk} = 1$, for $j = 1, 2$. I assume that the risk-neutral transition probabilities $\pi_{Qjk}$ and the real-world transition probabilities $\pi_{Pjk}$ are regime independent. I allow $\pi_{Qjk} \neq \pi_{Pjk}$. Agents are presumed to know the regime they are currently in, as well as the history of the factor $X_t$ and of the regime. The econometrician does not observe the regime. The Markov process governing regime changes is assumed to be conditionally independent of the $X_t$ process for tractability.

3 No arbitrage conditions for futures contract prices

Under my assumptions, the log of the futures price is affine in the factor $X_t$:

$$f_t^{(n)} = A_t^{(n)} + B^{(n)} X_t + u_t^{(n)}$$

(11)

My model implies the following cross-equation non-arbitrage restrictions on the parameters $A_j^{(n)}, j = 1, 2$ and $B^{(n)}$ characterizing the futures contract price:

$$A_j^{(n)} = \log \left( \sum_{k=1}^{2} \pi_{Qjk} e^{A_k^{(n-1)}} \right) + B^{(n-1)}(\mu_j - \lambda_{0,j}) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2$$

(12)

for $j = 1, 2$ and

$$B^{(n)} = B^{(n-1)}(\Phi - \lambda_1)$$

(13)

Equations (12) and (13) are derived in Appendix B in equations (A-9) and (A-11). They are very similar to the standard recursions for affine term structure models in the bond pricing literature (see for example Ang and Piazzesi (2003)). In bond pricing, the recursion for the intercept adds a term $\delta_0$ corresponding to the interest earned each period. For commodities, such a term does not appear since there is no initial capital investment. The recursions above represent non-linear cross-equation no arbitrage restrictions. These restrictions are not used or imposed in the initial reduced-form estimation, but are exploited in the second stage of
inference described below.

4 Estimation procedure

I assume that the factor $X_t$ is observed, and it is the first principal component extracted from the demeaned log prices of the futures contracts with maturities from 3 months to 9 months. Based on equations (9) and (1), I propose the following two-step method for estimating the parameters of the model.

4.1 Estimation of reduced-form parameters via regime-switching VAR’s

First, I estimate the following regime-switching regressions:

$$f_t^{(n)} = A_{st}^{(n)} + B^{(n)} X_t + u_t^{(n)}, n = 3, \ldots, 9$$  \hspace{1cm} (14)

where

$$
\begin{pmatrix}
  u^{(3)}_t \\
  u^{(4)}_t \\
  u^{(5)}_t \\
  u^{(6)}_t \\
  u^{(7)}_t \\
  u^{(8)}_t \\
  u^{(9)}_t \\
\end{pmatrix}
\sim N(0, \Omega)  \hspace{1cm} (15)
\]

I treat the first principal component extracted from futures prices as observed. Internal consistency between actual and model-implied principal components requires $h' A_j = -k$ for $j = 1, 2$ and $h' B = 1$ where $h$ is the $7 \times 1$ vector of principal component coefficients, $A_j$ is the $7 \times 1$ vector of model-implied values of the coefficients $A_j^{(n)}$ from equation (12) (for $n = 3, \ldots, 9$), $B$ is the $7 \times 1$ vector of model-implied coefficients $B^{(n)}$ from equation (13) (for $n = 3, \ldots, 9$), and $k = \frac{-h' \sum_{t=1}^{T} f_t}{h' f_T}$, where $f_T = (f_t^{(3)}, f_t^{(4)}, f_t^{(5)}, f_t^{(6)}, f_t^{(7)}, f_t^{(8)}, f_t^{(9)})'$. I do not impose these internal consistency conditions in the estimation. However, I find that they are satisfied to a very large degree of accuracy, and any inconsistency is negligible. Specifically, I find that $h' A_1 = 3.7861$, $h' A_2 = 3.7875$, $k = -3.7868$, $h' B = 1.0021$, so that $h' A_1 \approx -k$, $h' A_2 \approx -k$, and $h' B \approx 1$. 

---

8I treat the first principal component extracted from futures prices as observed. Internal consistency between actual and model-implied principal components requires $h' A_j = -k$ for $j = 1, 2$ and $h' B = 1$ where $h$ is the $7 \times 1$ vector of principal component coefficients, $A_j$ is the $7 \times 1$ vector of model-implied values of the coefficients $A_j^{(n)}$ from equation (12) (for $n = 3, \ldots, 9$), $B$ is the $7 \times 1$ vector of model-implied coefficients $B^{(n)}$ from equation (13) (for $n = 3, \ldots, 9$), and $k = \frac{-h' \sum_{t=1}^{T} f_t}{h' f_T}$, where $f_T = (f_t^{(3)}, f_t^{(4)}, f_t^{(5)}, f_t^{(6)}, f_t^{(7)}, f_t^{(8)}, f_t^{(9)})'$. I do not impose these internal consistency conditions in the estimation. However, I find that they are satisfied to a very large degree of accuracy, and any inconsistency is negligible. Specifically, I find that $h' A_1 = 3.7861$, $h' A_2 = 3.7875$, $k = -3.7868$, $h' B = 1.0021$, so that $h' A_1 \approx -k$, $h' A_2 \approx -k$, and $h' B \approx 1$. 

---
and

\[
\Omega \equiv \begin{pmatrix}
\Omega_{(3)} & 0 & 0 & 0 & 0 & 0 \\
0 & \Omega_{(4)} & 0 & 0 & 0 & 0 \\
0 & 0 & \Omega_{(5)} & 0 & 0 & 0 \\
0 & 0 & 0 & \Omega_{(6)} & 0 & 0 \\
0 & 0 & 0 & 0 & \Omega_{(7)} & 0 \\
0 & 0 & 0 & 0 & 0 & \Omega_{(8)} \\
0 & 0 & 0 & 0 & 0 & 0 & \Omega_{(9)} \\
\end{pmatrix}
\]

(16)

jointly with the regime-switching regression for \( X_t \):

\[
X_{t+1} = \mu_{s_t} + \Phi X_t + v_{t+1}, v_{t+1} | s_t \sim N(0, \sigma^2_{s_t})
\]

(17)
as a vector system of regime-switching equations. The time-series regressions in equation (14) estimate exposures of the futures prices with respect to the contemporaneous pricing factor. The regime-switching regression in equation (17) serves to decompose the pricing factor into a predictable component and a factor innovation by regressing the factor on its lagged level.

The estimation is done via the EM algorithm and is explained in detail in Appendix C. The general vector version of the EM algorithm is found in Hamilton (forthcoming).

4.2 Minimum-chi-square estimation of structural parameters

I use a minimum-chi-square approach to estimate the price of risk parameters \( \lambda_{0,1}, \lambda_{0,2}, \) and \( \lambda_1 \) and the risk-neutral probabilities \( \pi_{Q11} \) and \( \pi_{Q22} \). The procedure chooses the values of \( \lambda_{0,1}, \lambda_{0,2}, \lambda_1, \pi_{Q11}, \) and \( \pi_{Q22} \) so that the values for the recursive pricing parameters \( A_j^{(n)} \) and \( B^{(n)} \) implied by the no-arbitrage restrictions in equations (12) and (13) most closely fit the unrestricted first-stage estimates from equation (14). Minimum-chi-square estimation in the setting of single regime Gaussian affine term structure models is described in Hamilton and Wu (2012).

Let \( \pi \) denote the vector of reduced-form parameters (VAR coefficients, variance of the factor, measurement error variances, and \( P \)-measure regime-switching probabilities). Let \( \mathcal{L}(\pi; Y) \) denote the log likelihood for the entire sample, and let \( \hat{\pi} = \arg \max \mathcal{L}(\pi; Y) \) denote the full-information maximum likelihood estimate. If \( \hat{R} \) is a consistent estimate of the information matrix,

\[
R = -T^{-1} E \left[ \frac{\partial^2 \mathcal{L}(\pi; Y)}{\partial \pi \partial \pi'} \right]
\]

(18)
then $\theta$ can be estimated by minimizing the chi-square statistic

$$T [\hat{\pi} - g(\theta)]' \hat{R} [\hat{\pi} - g(\theta)]$$

As noted by Hamilton and Wu (2012), the variance of $\hat{\theta}$ can be approximated with $T^{-1}(\hat{\Gamma}' \hat{R} \hat{\Gamma})^{-1}$ for $\hat{\Gamma} = \frac{\partial g(\theta)}{\partial \theta} |_{\theta = \hat{\theta}}$.

In my case, I want to minimize the distance between the unrestricted maximum likelihood estimates of the coefficients $A_j^{(n)}$ and $B_j^{(n)}$ (from the regime-switching regressions) and the values of $A_j^{(n)}$ and $B_j^{(n)}$ implied by the no arbitrage restrictions. According to equations (12) and (13), these are predicted to be functions of $\theta$, a vector of structural parameters summarized in equation (22) below. Let $\hat{\pi}$ be the vector of the unrestricted maximum likelihood estimates from the regime-switching VAR:

$$\hat{\pi} = \left( \hat{\mu}_1, \hat{\mu}_2, \hat{\Phi}, \text{vec}(\tilde{A}'), \text{vec}(\tilde{B}), \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\Omega}_3, \hat{\Omega}_4, \hat{\Omega}_5, \hat{\Omega}_6, \hat{\Omega}_7, \hat{\Omega}_8, \hat{\Omega}_9, \hat{\pi}_{P11}, \hat{\pi}_{P22} \right)'$$

where $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\Phi}$, $\hat{\sigma}_1^2$, and $\hat{\sigma}_2^2$ are the unrestricted maximum likelihood estimates of the parameters $\mu_1$, $\mu_2$, $\Phi$, $\sigma_1^2$, and $\sigma_2^2$ from the regime-switching autoregression for the factor in equation (17),

$$\tilde{A} = \begin{bmatrix} \tilde{A}_1^{(3)} & \tilde{A}_2^{(3)} \\ \tilde{A}_1^{(4)} & \tilde{A}_2^{(4)} \\ \tilde{A}_1^{(5)} & \tilde{A}_2^{(5)} \\ \tilde{A}_1^{(6)} & \tilde{A}_2^{(6)} \\ \tilde{A}_1^{(7)} & \tilde{A}_2^{(7)} \\ \tilde{A}_1^{(8)} & \tilde{A}_2^{(8)} \\ \tilde{A}_1^{(9)} & \tilde{A}_2^{(9)} \end{bmatrix}$$

and $\tilde{B} = \left( \tilde{B}^{(3)}, \tilde{B}^{(4)}, \tilde{B}^{(5)}, \tilde{B}^{(6)}, \tilde{B}^{(7)}, \tilde{B}^{(8)}, \tilde{B}^{(9)} \right)'$ are the unrestricted maximum likelihood estimates of the coefficients of the regime-switching regressions for the futures prices in equation (14), $\hat{\Omega}_{(n)}$, $n = 3 \ldots 9$ are the unrestricted maximum likelihood estimates of the measurement error variances $\Omega_{(n)}$, $n = 3 \ldots 9$ in equation (16), and $\hat{\pi}_{P11}$ and $\hat{\pi}_{P22}$ are the unrestricted maximum likelihood estimates of the regime switching probabilities $\pi_{P11}$ and $\pi_{P22}$ from the regime-switching VAR.

Let

$$\theta = (\mu_1, \mu_2, \Phi, A_1^{(3)}, A_2^{(3)}, B_1^{(3)}, B_2^{(3)}, \sigma_1^2, \sigma_2^2, \Omega_3, \Omega_4, \Omega_5, \Omega_6, \Omega_7, \Omega_8, \Omega_9, \pi_{P11}, \pi_{P22}, \lambda_{0,1}, \lambda_{0,2}, \lambda_1, \pi_{Q11}, \pi_{Q22})'$$

(22)
and

\[ g(\theta) = (\mu_1, \mu_2, \Phi, \text{vec}(A'(\theta)), \text{vec}(B(\theta)), \sigma_1^2, \sigma_2^2, \Omega_{(3)}, \Omega_{(4)}, \Omega_{(5)}, \Omega_{(6)}, \Omega_{(7)}, \Omega_{(8)}, \Omega_{(9)}; \pi^{p_{11}}, \pi^{p_{22}})' \]

(23)

Here

\[ A(\theta) = \begin{bmatrix}
A_{1}^{(3)} & A_{2}^{(3)} \\
A_{1}^{(4)} & A_{2}^{(4)} \\
A_{1}^{(5)} & A_{2}^{(5)} \\
A_{1}^{(6)} & A_{2}^{(6)} \\
A_{1}^{(7)} & A_{2}^{(7)} \\
A_{1}^{(8)} & A_{2}^{(8)} \\
A_{1}^{(9)} & A_{2}^{(9)}
\end{bmatrix} \]  \hspace{1cm} (24)

and

\[ B(\theta) = (B_{(3)}, B_{(4)}, B_{(5)}, B_{(6)}, B_{(7)}, B_{(8)}, B_{(9)})' \]. \hspace{1cm} \text{vec}(A'(\theta)) \text{ and } \text{vec}(B(\theta)) \]

for \( n = 4, \ldots, 9 \) are defined by the no arbitrage restrictions from equations (12) and (13):

\[ A_j^{(n)} = \log \left( \sum_{k=1}^{2} \pi_{Qjk} e^{A_k^{(n-1)}} \right) + B^{(n-1)}(\mu_j - \lambda_{0,j}) + \frac{1}{2} B^{(n-1)^2} \sigma_2^2 \]

(25)

for \( j = 1, 2 \) and

\[ B^{(n)} = B^{(n-1)}(\Phi - \lambda_1) \]

(26)

For \( n = 3 \), \( g(A_{1}^{(3)}) = A_{1}^{(3)} \), \( g(A_{2}^{(3)}) = A_{2}^{(3)} \), and \( g(B^{(3)}) = B^{(3)} \). Then \( \hat{\theta} \) is obtained as

\[ \hat{\theta} \equiv \text{argmin}_{\theta} \{ T [\hat{\pi} - g(\theta)]' \hat{R} [\hat{\pi} - g(\theta)] \} \]

(27)

In this way I obtain estimates of the prices of risk \( \lambda_{0,1}, \lambda_{0,2}, \) and \( \lambda_1 \) and of the risk-neutral transition probabilities \( \pi^{Q_{11}} \) and \( \pi^{Q_{22}} \) as part of the vector \( \hat{\theta} \). I also obtain second-stage estimates of \( \mu_1, \mu_2, \Phi, \sigma_1^2, \sigma_2^2, \pi^{p_{11}}, \pi^{p_{22}}, A_{1}^{(3)}, A_{2}^{(3)}, B^{(3)} \), and \( \Omega_{(n)} \), \( n = 3 \ldots 9 \). By using the estimators for \( \lambda_{0,1}, \lambda_{0,2}, \lambda_1, \pi^{Q_{11}} \) and \( \pi^{Q_{22}} \) outlined above, I obtain the price of risk parameters and risk-neutral transition probabilities so that the values of \( A_j^{(n)} \) and \( B^{(n)} \) implied by equations (25) and (26) most closely fit the unrestricted estimates from equation (14).
5 Empirical results

5.1 Data

I estimate a one factor model using data on prices of natural gas futures contracts traded on NYMEX with maturities 3 months to 9 months for the period from January, 1995 to June, 2014. The factor is constructed as the first principal component extracted from the demeaned log prices of these contracts. The data is obtained from Datastream. Natural gas contracts expire three business days prior to the first calendar day of the delivery month. Figure 1 shows the log of the observed 3 month futures price. I use a cross-section of $N = 7$ maturities in my estimation. I estimate the reduced-form parameters $\{A_j, \mu_j, \sigma_j^2\}$ for $j = 1, 2, B, \Omega$, and $\Phi$, and the probabilities $\pi_{P11}, \pi_{P22}$, and $\rho_1$ in the first step of the estimation procedure, and then estimate the prices of risk $\lambda_{0,j}$ and $\lambda_1$ and the risk-neutral probabilities $\pi_{Q11}$ and $\pi_{Q22}$ in the second step. Here $\rho_1$ is the probability that the initial state is regime 1. Since the longest maturity contract I use in my estimation is the 9-month contract, whereas the shortest term contract I use is the 3-month contract, I define the market as being in contango if the price of the 9-month contract is higher than the price of the 3-month contract. Conversely, I define the market as being in backwardation if the price of the 9-month contract is lower than the price of the 3-month contract.

5.2 Estimation results

The first principal component used as factor in my model captures 98.73% of the variation in futures prices. I use the Eigenvalue Ratio and Growth Ratio tests proposed in Ahn and Horenstein (2013) (described in Appendix F) in order to estimate the number of factors in my model. Both of these tests yield 1 as the number of factors that need to be used. This justifies my use of a one factor model.

Table 1 shows the estimates of the reduced-form parameters and the historical transition probabilities from the first stage of the estimation. Table 2 shows second stage estimates, including the estimates of the market prices of factor risk and the risk-neutral transition probabilities. The value of the chi-squared objective function from the second stage estimation has an asymptotic $\chi^2(q)$ distribution under the null hypothesis that the model-implied no-arbitrage restrictions are satisfied by the data. Here $q$ is the number of reduced-form parameters in the first stage, which is equal to 35. In my estimation, the value of the objective function is 14.5270, which indicates that we fail to reject the null hypothesis. Hence, I conclude that the no-arbitrage restrictions are satisfied by the data.

I find that the factor is very persistent, with $\hat{\Phi} = 0.9785$. The factor is a stationary
stochastic process under the $\mathbb{P}$-measure. The loadings of the futures prices on the first principal component are basically constant across maturities. Thus, the factor essentially represents a parallel change in prices. Because of its effect on price levels, I refer to this factor as the level factor. This is commonly done in the literature using principal component factor models.

Table 3 shows Wald t-statistics for the hypothesis tests testing whether there is regime switching in the various parameters. I find very strong evidence that there are regimes in the data. The null hypotheses that the level of the factor $\mu$ and the levels $A^{(3)}$, $A^{(4)}$, $A^{(5)}$, $A^{(7)}$, $A^{(8)}$, $A^{(9)}$ of the contracts are not regime switching are strongly rejected. The level $\mu$ of the factor is higher in regime 1 than in regime 2, and $\mu_1 - \mu_2$ is statistically significantly different from 0. The levels of the contracts with maturity 3-5 months are statistically significantly lower in regime 1, while the levels of the contracts with maturity 7-9 months are statistically significantly higher in regime 1. The estimated variances in the two regimes are not statistically significantly different. The coefficients $\hat{B}$ and the variance of the measurement errors $\hat{\Omega}$ are regime independent by assumption. Figure 2 shows the spread between the 9 month contract and the 3 month contract plotted against the smoothed probability of regime 2. It can be seen that in regime 1 the spread is higher and is almost always positive, whereas in regime 2 the spread is lower and is almost always negative. Thus, I find that the estimated regimes correspond to the previously defined market states of contango and backwardation. Regime 1 (the positive spread regime) represents contango, while regime 2 (the negative spread regime) represents backwardation. Figures 3 and 4 show the log of the observed 3 month futures price and the factor, respectively, with shaded areas representing the backwardation regime.

Figure 5 shows the expected monthly holding returns for each contract conditional on the contango regime and the backwardation regime, respectively, averaged over each regime. Here I define the expected monthly holding return on the $n$-month contract conditional on being in regime $j$ at time $t$ as $E_{F^{(n)}_t}\left[ F^{(n)}_{t+1} - F^{(n)}_t | s_t = j \right].$ The fact that the expected holding returns for all contracts (except the 6-month contract) change sign depending on the regime and are large in magnitude in each regime shows that agents face very substantial risk due to the possibility of changes in regime. For instance, in the backwardation regime, an investor holding a long position in the 9-month contract would on average earn a positive expected monthly holding return of 2.5%. On the other hand, in the contango regime, the investor would earn a negative expected monthly holding return of about -0.5% on average. Thus,
as long as the market is in backwardation, the investor would on average earn a profit, but if the regime switches, he would experience a significant loss.

From Figure 5, we can also see that in the contango regime, buying the 4-6 month contracts today and selling them next month on average yields a profit. So does shorting the 7-9 month contracts today and closing out the position next month. In the backwardation regime, shorting the 4-5 contracts today and closing out the position next month on average yields a profit. So does going long on the 6-9 month contracts today and selling them next month. Thus, on average it is profitable to be long in the 4-6 month contracts in the contango regime, whereas it is profitable to be long in the 6-9 month contracts in the backwardation regime.

The fact that there is a positive return to a long position in the 4-6 month contracts in the contango regime suggests that there is demand for short positions in futures. This could mean that in the contango regime, commercial producers are trying to hedge their short positions in natural gas by selling 4-6 month futures contracts. Moreover, in the contango regime there is a positive return to a short position in the 7-9 month contracts, which could indicate demand for long positions in these contracts. In turn, this could indicate that commercial users are trying to hedge their long positions in natural gas by buying 7-9 month contracts. Similarly, in the backwardation regime there is a positive return to a long position in the 6-9 month contracts, which could be a result of commercial producers trying to hedge their short positions by selling 6-9 month contracts. Moreover, in the backwardation regime there is a positive return to a short position in the 4-5 month contracts, which could be a result of commercial users trying to hedge their long positions by buying 4-5 month contracts.

By allowing for $\pi^P \neq \pi^Q$ in my model, I have another channel through which risk preferences can affect expected returns. I test the restrictions $\pi^{P11} = \pi^{Q11}$ and $\pi^{P22} = \pi^{Q22}$, and find that they are rejected. Thus, my results suggest that $\pi^P \neq \pi^Q$. I find that $\pi^{Q11}_1 > \pi^{P11}_1$ and $\pi^{Q22}_2 > \pi^{P22}_2$. This suggests that investors act as though regimes are more persistent than they really are. This is consistent with what we see in reality, as market participants often just act as if the current state of the market will continue next period. This overestimation of the persistence of the regimes by investors could be a result of hedging pressure, which I discuss later in this section.

The expected log holding return $E_t^P \left[ r_{t+1}^{(n-1)} | s_t = j \right]$ is related to the risk premium investors demand in regime $j$ for holding a futures contract maturing in $n$ months for 1 month. The expression for the expected log holding return of the $n$-month contract conditional on
the regime $E_t^P [rx_{t+1}^{(n-1)} | s_t = j]$, derived in equation (A-12) in Appendix B, is

$$E_t^P [rx_{t+1}^{(n-1)} | s_t = j] = \sum_{k=1}^{2} \pi_{P, jk} A_k^{(n-1)} - \log \left( \sum_{k=1}^{2} \pi_{Q, jk} e^{A_k^{(n-1)}} \right) + B^{(n-1)} \lambda_{0, j} + B^{(n-1)} \lambda_1 X_t - \frac{1}{2} B^{(n-1)^2} \sigma_j^2 \tag{28}$$

From the formula, we can see that nonzero expected returns arise in part due to the difference between the historic probabilities $\pi_{P, jk}$ and the risk-neutral probabilities $\pi_{Q, jk}$. The difference between $\pi_{P, jk}$ and $\pi_{Q, jk}$ accounts for a large proportion of expected returns.

It can be shown that the market price of regime shift risk $\Gamma_{j,k} = \log \left( \frac{\pi_{P, jk}}{\pi_{Q, jk}} \right)$ (see equation (A-2) in Appendix A for derivation). Since $\pi_{P, jk}$ and $\pi_{Q, jk}$ are statistically significantly different, the market price of regime shift risk is nonzero, i.e. regime shift risk is priced. To motivate the labeling of $\Gamma_{s_t, s_{t+1}}$, as the market price of regime shift risk, consider a security which pays $1 if the regime switches from regime $s_t = j$ in month $t$ to regime $s_{k+1} = k$ in month $t + 1$. This security has payoff $\mathbb{1}_{\{s_{t+1} = k\}}$ and has exposure only to the risk of shifting from regime $j$ in month $t$ to regime $k$ in month $t + 1$. Conditional on the current regime $s_t = j$, its current price is

$$P_t^j = E_t^Q [\mathbb{1}_{\{s_{t+1} = k\}} | s_t = j] = \pi_{Q, jk}$$

Therefore, its log expected return is

$$\log \frac{E_t^P [\mathbb{1}_{\{s_{t+1} = k\}} | s_t = j]}{P_t^j} = \log \left( \frac{\pi_{P, jk}}{\pi_{Q, jk}} \right) = \Gamma_{j,k} \tag{29}$$

Thus, $\Gamma_{j,k}$ gives the log expected return per unit of regime shift risk exposure, and can therefore be interpreted as the market price of regime shift risk associated with switching from regime $j$ to regime $k$.

For instance, a security which pays off $1 if the regime switches from regime 2 (backwardation) today to regime 1 (contango) next month is priced at

$$P_t^{(2)} = E_t^Q [\mathbb{1}_{\{s_{t+1} = 1\}} | s_t = 2] = \pi_{Q, 21} = 1 - \pi_{Q, 22} = $0.0187

Thus, investors are willing to pay only about 2 cents to hedge against the regime switching from backwardation to contango next month. The expected payoff of the security is

$$E_t^P [\mathbb{1}_{\{s_{t+1} = 1\}} | s_t = 2] = \pi_{P, 21} = 1 - \pi_{P, 22} = $0.1944.
So the security pays off about 19 cents on average. The price investors are willing to pay to hedge against regime shift risk is very low, reflecting the fact that they think the backwardation regime is considerably more persistent than it actually is.

Similarly, a security which pays off $1 if the regime changes from regime 1 (contango) today to regime 2 (backwardation) next month is priced at

$$P_t^{(1)} = E_t^Q [\mathbb{1}_{s_{t+1} = 2} | s_t = 1] = \pi_t^{Q12} = 1 - \pi_t^{Q11} = 0.0224$$

Thus, investors are willing to pay about 2 cents to hedge against the risk of the regime switching from contango today to backwardation next month. On average, the security pays off

$$E_t^P [\mathbb{1}_{s_{t+1} = 2} | s_t = 1] = \pi_t^{P12} = 1 - \pi_t^{P11} = 0.0930$$

i.e. about 9 cents. Once again, the price investors are willing to pay to hedge against the risk of the regime changing is low, but it is closer to the actual expected payoff of the security. To summarize, I find that agents act as if both regimes are more persistent than they are, with the perceived overestimation of the regime persistence being even higher for the backwardation regime.

Why is the security which pays off $1 if the regime switches from regime 2 (backwardation) today to regime 1 (contango) next month so cheap? The potential seller of this security could be using it as a hedge against risk, if the state of the world when the seller would have to pay $1 is one in which the seller will profit from other sources, and the state of the world when the seller keeps the dollar is one when even a little more money would be helpful. For example, a commercial user who buys natural gas with 3-month contracts would benefit from selling this security. He would have to pay on the contract if the regime shifts from backwardation to contango. But if the regime shifts, he would profit by from then on being able to buy futures at a lower price. Hedging pressure could cause the price to fall below the 19-cent valuation.

The price of risk $\lambda_1$ is statistically significant and negative, and the loadings $B^{(n)}$ are positive. From equation (28), we can see that this implies that an increase in the level of futures prices (as measured by the factor $X_t$) decreases the expected log holding returns on the 4-9 month contracts in each regime. The expected log holding return loading for the $n$-month contract is $B^{(n-1)} \lambda_1$. According to my estimates, a positive one standard deviation shock to the level factor reduces the expected log holding return on the 4-9 month contracts in each regime by about 0.88%.

The estimates of the prices of risk $\lambda_{0,1}$ and $\lambda_{0,2}$ are not statistically significant. Moreover, using a Wald test I find that $\lambda_{0,1} - \lambda_{0,2}$ is not statistically significantly different from 0. Thus,
I do not find considerable differences in the market pricing of factor risk in the two regimes.

6 Conclusion

In this paper I have provided a new way of characterizing risk in commodities futures markets, which tend to switch between periods of contango when the spread between the longer term futures and the shorter term futures is positive, and backwardation when the spread between the longer term futures and the shorter term futures is negative. I apply my framework to the natural gas futures market, where the risk agents face due to the possibility of switching between the two states of the market is particularly substantial. Motivated by the historically observed switches of natural gas futures prices between states of contango and backwardation, I propose and estimate a Markov regime-switching Gaussian affine term structure model with two regimes and estimate the states of the market from the data. In my model, the level and volatility of natural gas futures, as well as the risk premium, are regime dependent. I study the consequences of changes in regime on the risk premium, and produce novel empirical estimates to characterize risk premia and the term structure of natural gas futures contracts. I find very strong evidence that there are regimes in the data. Moreover, I find that the regimes in my model correspond precisely to historically observed periods of contango and backwardation. I also find that regime switching risk is priced. I find that expected futures returns for most contracts change sign depending on which regime we are conditioning on. This is consistent with the claim that agents face significant risks from the possibility of a change in the regime. My results show that the market acts as if regimes are more persistent than they really are, which could be a result of hedging pressure. I find evidence that commercial users and commercial producers use natural gas futures contracts for purposes of hedging. In the contango regime, commercial producers may be trying to hedge their short positions in natural gas by selling 4-6 month contracts, while in the backwardation regime, commercial producers may be trying to hedge their short positions by selling 6-9 month contracts. I obtain analogous implications for commercial users. A separate contribution of the paper is to propose a new method for estimating Gaussian affine term structure models subject to regime-switching. My approach allows for computationally fast estimation and avoids the numerical difficulties that are common when using other maximum likelihood based methods in the literature.
APPENDIX

A Relation between $\mathbb{P}$-dynamics and $\mathbb{Q}$-dynamics

By no arbitrage, an asset with payoff $g(X_{t+1})$ has a price in regime $j$ equal to

$$P(X_t) = E_t^P[M_{t,t+1}g(X_{t+1})|s_t = j] = E_t^Q[g(X_{t+1})|s_t = j]$$  \hspace{1cm} (A-1)

$$\pi_{Qjk} = E_t^Q[1_{\{s_{t+1}=k\}}|s_t = j] = E_t^P[1_{\{s_{t+1}=k\}}M_{t,t+1}|s_t = j]$$

$$= E_t^P[1_{\{s_{t+1}=k\}}\exp\left(-\Gamma_{st,s_{t+1}} - \frac{1}{2}\lambda_{t,ts}^2 - \lambda_{t,ts}\sigma_{st}^{-1}v_{t+1}\right)|s_t = j]$$

$$= E_t^P[\exp\left(-\frac{1}{2}\lambda_{t,ts}^2 - \lambda_{t,ts}\sigma_{st}^{-1}v_{t+1}\right)|s_t = j] E_t^P[1_{\{s_{t+1}=k\}}\exp\left(-\Gamma_{st,s_{t+1}}\right)|s_t = j]$$

$$= \exp\left(-\frac{1}{2}\lambda_{t,j}^2\right) E_t^P[\exp\left(-\lambda_{t,ts}\sigma_{st}^{-1}v_{t+1}\right)|s_t = j] \pi_{Pjk} \exp\left(-\Gamma_{j,k}\right)$$

$$= \exp\left(-\frac{1}{2}\lambda_{t,j}^2\right) \exp\left(\frac{1}{2}Var_t(-\lambda_{t,ts}\sigma_{st}^{-1}v_{t+1}|s_t = j)\right) \pi_{Pjk} \exp\left(-\Gamma_{j,k}\right)$$

$$= \exp\left(-\frac{1}{2}\lambda_{t,j}^2\right) \exp\left(\frac{1}{2}\lambda_{t,j}^2\sigma_j^{-2}\sigma_{t,j}^2\right) \pi_{Pjk} \exp\left(-\Gamma_{j,k}\right)$$

Therefore,

$$\Gamma_{j,k} = \log\left(\frac{\pi_{Qjk}}{\pi_{Pjk}}\right)$$  \hspace{1cm} (A-2)
By no arbitrage, an asset with payoff $g(X_{t+1})$ has a price in regime $j$ equal to

$$P(X_t) = E_t [M_{t,t+1}g(X_{t+1}) | s_t = j]$$

$$= E_t \left[ \exp \left( -\Gamma_{s_t,s_{t+1}} - \frac{1}{2} \lambda_{t,s_t}^2 + \lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left( -\frac{1}{2} \lambda_{t,j}^2 \right) E_t^p \left[ \exp \left( -\Gamma_{s_t,s_{t+1}} - \lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left( -\frac{1}{2} \lambda_{t,j}^2 \right) \sum_{k=1}^{2} \pi_{t,j}^{p,jk} \exp \left( -\Gamma_{j,k} \right) E_t^p \left[ \exp \left( -\lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left( -\frac{1}{2} \lambda_{t,j}^2 \right) \sum_{k=1}^{2} \pi_{t,j}^{Q,jk} \pi_{t,j}^{p,jk} E_t^p \left[ \exp \left( -\lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left( -\frac{1}{2} \lambda_{t,j}^2 \right) \sum_{k=1}^{2} \pi_{t,j}^{Q,jk} \pi_{t,j}^{p,jk} E_t^p \left[ \exp \left( -\lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left( -\frac{1}{2} \lambda_{t,j}^2 \right) \int \pi_{t,j}^{p,jk} \pi_{t,j}^{Q,jk} \pi_{t,j}^{p,jk} E_t^p \left[ \exp \left( -\lambda_{t,s_t} \sigma_{s_t}^{-1} v_{t+1} \right) g(X_{t+1}) | s_t = j \right]$$

$$= \exp \left( -\frac{1}{2} \lambda_{t,j}^2 \right) \int g(X_{t+1}) \exp \left( -\lambda_{t,j} \sigma_{s_t}^{-1} (X_{t+1} - \mu_j - \Phi X_t) \right) (2\pi)^{-1/2} \sigma_j^{-1}$$

$$\exp \left( -\frac{1}{2} \sigma_j^{-1} (X_{t+1} - \mu_j - \Phi X_t)^2 \right) dX_{t+1}$$

$$= (2\pi)^{-1/2} \sigma_j^{-1} \int g(X_{t+1}) \exp \left( -\frac{1}{2} \frac{1}{\sigma_j^2} (X_{t+1} - \mu_j - \Phi X_t)^2 \right.$$

$$+ 2\lambda_{t,j} \sigma_{s_t}^{-1} (X_{t+1} - \mu_j - \Phi X_t) + \lambda_{t,j}^2 \right) dX_{t+1}$$

$21$
\[
= (2\pi)^{-1/2} \sigma_j^{-1} \int g(X_{t+1}) \exp \left( -\frac{1}{2} \left[ \frac{1}{\sigma_j} (X_{t+1} - \mu_j - \Phi X_t + \lambda_{t,j})^2 \right] \right) dX_{t+1}
\]
\[
= (2\pi)^{-1/2} \sigma_j^{-1} \int g(X_{t+1}) \exp \left( -\frac{1}{2} \left[ X_{t+1} - \mu_j - \Phi X_t + \sigma_j \lambda_{t,j} \right]^2 \sigma_j \right) dX_{t+1}
\]
\[
= (2\pi)^{-1/2} \sigma_j^{-1} \int g(X_{t+1}) \exp \left( -\frac{1}{2} \left[ X_{t+1} - \mu_j - \Phi X_t + \sigma_j \lambda_{t,j} \right]^2 \sigma_j \right) dX_{t+1}
\]
\[
= (2\pi)^{-1/2} \sigma_j^{-1} \int g(X_{t+1}) \exp \left( -\frac{1}{2} \left[ X_{t+1} - \mu_j - \Phi X_t + \sigma_j \lambda_{t,j} \right]^2 \sigma_j \right) dX_{t+1} =
\]
\[
= (2\pi)^{-1/2} \sigma_j^{-1} \int g(X_{t+1}) \exp \left( -\frac{1}{2} \sigma_j^2 \left[ X_{t+1} - \left( \mu_j - \lambda_{0,j} \right) - \left( \Phi - \lambda_1 \right) X_t \right]^2 \right) dX_{t+1} =
\]
\[
= E_t^Q (g(X_{t+1})|s_t = j)
\]

where in the last line I used Equation (A-1).

Therefore, under the Q-measure,

\[
X_{t+1}|s_t = j \sim^Q N \left( (\mu_j - \lambda_{0,j}) + (\Phi - \lambda_1) X_t, \sigma_j^2 \right) \quad (A-3)
\]

or, equivalently,

\[
X_{t+1}|s_t = j \sim^Q N \left( \mu_j^Q + \Phi^Q X_t, \sigma_j^2 \right) \quad (A-4)
\]

where

\[
\mu_j^Q \equiv \mu_j - \lambda_{0,j} \quad (A-5)
\]

and

\[
\Phi^Q \equiv \Phi - \lambda_1 \quad (A-6)
\]

Hence, under the Q-measure, \( X_{t+1} \) follows the dynamics

\[
X_{t+1} = \mu_j^Q + \Phi^Q X_t + v_{t+1}^Q \quad (A-7)
\]

where \( v_{t+1}^Q|s_t = j \sim^Q N(0, \sigma_j^2) \) under the Q-measure.
B Calculating expected returns

\[
E_t^P \left[ r_{t+1}^{(n-1)} | s_t = j \right] = E_t^P \left[ f_{t+1}^{(n-1)} - f_t^{(n)} | s_t = j \right] = E_t^P \left[ A^{(n-1)}_{st+1} + B^{(n-1)} X_{t+1} - A^{(n)}_{st} - B^{(n)} X_t | s_t = j \right] = E_t^P \left[ A^{(n-1)}_{st+1} + B^{(n-1)} (\mu_{st} + \Phi X_t + \nu_{t+1}) - A^{(n)}_{st} - B^{(n)} X_t | s_t = j \right] = \pi_{P1} A_1^{(n-1)} + \pi_{P2} A_2^{(n-1)} + B^{(n-1)} \mu_j + (B^{(n-1)} \Phi - B^{(n)}) X_t - A_j^{(n)}
\] (A-8)

The futures price is

\[
F_t^{(n)} = e^{A_j^{(n)} + B^{(n)} X_t}
\]

\[
f_t^{(n)} = \log F_t^{(n)} = \log E_t^Q \left[ F_{t+1}^{(n-1)} | s_t = j \right] = \log \left( \sum_{k=1}^{2} \pi_{Q,k} E_t^Q \left[ F_{t+1}^{(n-1)} | s_t = j \right] \right) = \log \left( \sum_{k=1}^{2} \pi_{Q,k} e^{A_k^{(n-1)}} \right) + \log E_t^Q \left[ e^{B^{(n-1)} X_{t+1}} | s_t = j \right] = \log \left( \sum_{k=1}^{2} \pi_{Q,k} e^{A_k^{(n-1)}} \right) + \log E_t^Q \left[ e^{B^{(n-1)} (\mu_j^{Q} + \Phi^Q X_t + \nu_{t+1})} | s_t = j \right] = \log \left( \sum_{k=1}^{2} \pi_{Q,k} e^{A_k^{(n-1)}} \right) + \log \left[ e^{B^{(n-1)} (\mu_j^{Q} + \Phi^Q X_t) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2} \right] = \log \left( \sum_{k=1}^{2} \pi_{Q,k} e^{A_k^{(n-1)}} \right) + B^{(n-1)} (\mu_j^{Q} + \Phi^Q X_t) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2
\]

Therefore,

\[
f_t^{(n)} = A_j^{(n)} + B^{(n)} X_t = \log \left( \sum_{k=1}^{2} \pi_{Q,k} e^{A_k^{(n-1)}} \right) + B^{(n-1)} \mu_j^{Q} + \frac{1}{2} B^{(n-1)^2} \sigma_j^2 + B^{(n-1)} \Phi^Q X_t
\]

The above equation implies the following recursions:

\[
A_j^{(n)} = \log \left( \sum_{k=1}^{2} \pi_{Q,k} e^{A_k^{(n-1)}} \right) + B^{(n-1)} \mu_j^{Q} + \frac{1}{2} B^{(n-1)^2} \sigma_j^2
\]

23
or equivalently

\[ A_j^{(n)} = \log \left( \sum_{k=1}^{2} \pi Q_{jk} e^{A_k^{(n-1)}} \right) + B^{(n-1)}(\mu_j - \lambda_{0,j}) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2 \]  
(A-9)

and

\[ B^{(n)} = B^{(n-1)} \Phi^Q \]  
(A-10)

or equivalently

\[ B^{(n)} = B^{(n-1)}(\Phi - \lambda_1) \]  
(A-11)

\[
E_t^p \left[ f_{t+1}^{(n-1)} | s_t = j \right] = \sum_{k=1}^{2} \pi_{jk}^p E_t^p \left[ f_{t+1}^{(n-1)k} | s_t = j \right] \\
= \sum_{k=1}^{2} \pi_{jk}^p \left( A_k^{(n-1)} + B^{(n-1)} X_{t+1} | s_t = j \right) \\
= \sum_{k=1}^{2} \pi_{jk}^p \left( A_k^{(n-1)} + B^{(n-1)} E_t^p [\mu_{s_t} + \Phi X_t | s_t = j] \right) \\
= \sum_{k=1}^{2} \pi_{jk}^p \left( A_k^{(n-1)} + B^{(n-1)} (\mu_j + \Phi X_t) \right) \\
= \sum_{k=1}^{2} \pi_{jk}^p A_k^{(n-1)} + \left( \sum_{k=1}^{2} \pi_{jk}^p \right) B^{(n-1)} (\mu_j + \Phi X_t) \\
= \sum_{k=1}^{2} \pi_{jk}^p A_k^{(n-1)} + B^{(n-1)} (\mu_j + \Phi X_t)
\]

24
We can also derive an expression for \( E^p_t \left[ x_{t+1}^{(n-1)} | s_t = j \right] \) where \( f_t^{(n-1)} = f_t^{(n)} | s_t = j \):

\[
E^p_t \left[ f_t^{(n-1)} - f_t^{(n)} | s_t = j \right] = E^p_t \left[ f_t^{(n-1)} | s_t = j \right] - f_t^{(n)j}
\]

\[
= \sum_{k=1}^{2} \pi^p_{jk} A_k^{(n-1)} + B^{(n-1)}(\mu_j + \Phi X_t) - \log \left( \sum_{k=1}^{2} \pi^Q_{jk} e^{A_k^{(n-1)}} \right) - B^{(n-1)}(\mu_j^Q + \Phi^Q X_t) - \frac{1}{2} B^{(n-1)^2} \sigma^2_j
\]

\[
= \sum_{k=1}^{2} \pi^p_{jk} A_k^{(n-1)} + B^{(n-1)}(\mu_j - \mu_j^Q) + B^{(n-1)}(\Phi - \Phi^Q) X_t - \log \left( \sum_{k=1}^{2} \pi^Q_{jk} e^{A_k^{(n-1)}} \right) - \frac{1}{2} B^{(n-1)^2} \sigma^2_j
\]

We can also derive an expression for \( E^p_t \left[ x_{t+1}^{(n-1)} | s_t = j \right] \).

\[
E^p_t \left[ x_{t+1}^{(n-1)} | s_t = j \right] = \sum_{j=1}^{2} E^p_t \left[ x_{t+1}^{(n-1)} | s_t = j \right] P(s_t = j | \mathcal{F}_t)
\]

\[
= \sum_{j=1}^{2} \left[ \sum_{k=1}^{2} \pi^p_{jk} A_k^{(n-1)} - \log \left( \sum_{k=1}^{2} \pi^Q_{jk} e^{A_k^{(n-1)}} \right) + B^{(n-1)} \lambda_{0,j} + B^{(n-1)} \lambda_1 X_t \right] - \frac{1}{2} B^{(n-1)^2} \sigma^2_j \times P(s_t = j | \mathcal{F}_t)
\]

\[
F_t^{(n)j} = E^Q_t \left[ F_t^{(n-1)} | s_t = j \right] = \sum_{k=1}^{2} \pi^Q_{jk} E^Q_t \left[ F_t^{(n-1)k} | s_t = j \right]
\]

\[
= \sum_{k=1}^{2} \pi^Q_{jk} E^Q_t \left[ e^{A_k^{(n-1)} + B^{(n-1)} X_{t+1}} | s_t = j \right]
\]

\[
= \sum_{k=1}^{2} \pi^Q_{jk} e^{A_k^{(n-1)}} \left[ e^{B^{(n-1)} X_{t+1}} | s_t = j \right]
\]

\[
= \sum_{k=1}^{2} \pi^Q_{jk} e^{A_k^{(n-1)}} \left[ e^{B^{(n-1)}(\mu_j^Q + \Phi^Q X_t + \sigma_j^2)} | s_t = j \right]
\]

\[
= \sum_{k=1}^{2} \pi^Q_{jk} e^{A_k^{(n-1)}} e^{B^{(n-1)}(\mu_j^Q + \Phi^Q X_t) + \frac{1}{2} B^{(n-1)^2} \sigma^2_j}
\]
\[E_t^P[F_{t+1}^{(n-1)}|s_t = j] = \sum_{k=1}^{2} \pi_{Pj} E_t P[e^{A_k^{(n-1)} + B^{(n-1)}X_t} | s_t = j]\]

\[= \sum_{k=1}^{2} \pi_{Pj} E_t P[e^{A_k^{(n-1)}} | s_t = j]\]

\[= \left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] E_t P[e^{B^{(n-1)}X_t} | s_t = j]\]

\[= \left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] E_t P[e^{B^{(n-1)}(\mu_j^P + \Phi X_t + \nu_t + 1)} | s_t = j]\]

\[= \left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\mu_j^P + \Phi X_t) + \frac{1}{2} B^{(n-1)}^2 \sigma_j^2} \quad (A-14)\]

Then

\[\frac{E_t^P[F_{t+1}^{(n-1)}|s_t = j]}{F_t^{(n)j}} = \frac{\left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\mu_j^P + \Phi X_t) + \frac{1}{2} B^{(n-1)}^2 \sigma_j^2}}{\left[ \sum_{k=1}^{2} \pi_{Qj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\mu_j^P + \Phi X_t) + \frac{1}{2} B^{(n-1)}^2 \sigma_j^2}}\]

\[= \frac{\left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\mu_j^P + \Phi X_t)}}{\left[ \sum_{k=1}^{2} \pi_{Qj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\mu_j^Q + \Phi X_t)}}\]

\[= \frac{\left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\mu_j^P - \mu_j^Q)}}{\left[ \sum_{k=1}^{2} \pi_{Qj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\mu_j^Q - \mu_j^P)}}\]

\[= \frac{\left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\lambda_{0,j} + \lambda_1 X_t)}}{\left[ \sum_{k=1}^{2} \pi_{Qj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}(\lambda_{0,j} + \lambda_1 X_t)}}\]

\[= \frac{\left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}\lambda_{1,j}}}{\left[ \sum_{k=1}^{2} \pi_{Qj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}\lambda_{1,j}}} \quad (A-15)\]

Therefore,

\[\frac{E_t^P[F_{t+1}^{(n-1)} - F_t^{(n)j}|s_t = j]}{F_t^{(n)j}} = \frac{E_t^P[F_{t+1}^{(n-1)}|s_t = j] - F_t^{(n)j}}{F_t^{(n)j}}\]

\[= \frac{E_t^P[F_{t+1}^{(n-1)}|s_t = j]}{F_t^{(n)j}} - 1\]

\[= \frac{\left[ \sum_{k=1}^{2} \pi_{Pj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}\lambda_{1,j}}}{\left[ \sum_{k=1}^{2} \pi_{Qj} e^{A_k^{(n-1)}} \right] e^{B^{(n-1)}\lambda_{1,j}}} - 1 \quad (A-16)\]
C  EM algorithm for first stage estimation

In the first stage I estimate the system of equations (14) and (17). It is known that in the absence of regime-switching, maximum likelihood estimation of this system is equivalent to OLS estimation equation by equation. Conditional on parameters, the inference about the regime \( \Pr(s_t = j | \mathcal{F}_T)^{10} \) can be obtained using the Hamilton filtering and smoothing algorithm. This suggests estimation via the EM algorithm. Let \( \theta \) denote the vector of parameters to be estimated, \( \theta \equiv \{ \text{vec}(A'), \text{vec}(B), \Phi, \Omega, \{ \mu_j, \sigma_j \}_{j=1}^2 \} \) where \( \text{vec}(A') \) and \( \text{vec}(B) \) are as defined in Section 4.2. First, I initialize the algorithm with an initial guess for the vector of parameters, and compute the corresponding smoothed probabilities. Then each iteration \( l \) of the algorithm proceeds as follows. First, I update inference for the regression parameters equation by equation. An updated estimate \( \hat{\theta}^{(l)} \) is derived as a solution to the first-order conditions for maximization of the likelihood function, where the conditional regime probabilities \( \Pr(s_t | Y, \theta^{(l-1)}) \) are replaced with the smoothed probabilities \( \Pr(s_t = j | Y, \theta^{(l-1)}) \) computed in the previous iteration, for \( Y = \{ Y_{t1}, Y_{t2}, X_t \}_t^{T=1} \) defined below. Conditional on knowing the smoothed probabilities, a closed form solution for the regression parameters of each regime-switching equation can be obtained by linear regression in which the observations are weighted by the smoothed probability that they came from the corresponding regime. Details are shown below. Next, I update inference about the smoothed probabilities \( \Pr(s_t = j | Y, \theta^{(l)}) \), where I am conditioning on the parameter vector estimate obtained in the current iteration instead of the unknown parameter vector \( \theta \).

The regime-switching system I estimate is of the form

\[
Y_{t1} = \mu_{s_t} + \Phi X_t + \varepsilon_t | s_t \sim N(0, \sigma_{s_t}^2) \quad (A-17)
\]

\[
Y_{t2} = A_{s_t} + B X_t + u_t | s_t \sim N(0, \Omega) \quad (A-18)
\]

Equation (14) when stacked across all maturities \( n = 3, \ldots, 9 \) is of the form of the above equation (A-18) with \( Y_{t2} = (f_t^{(3)}, f_t^{(4)}, f_t^{(5)}, f_t^{(6)}, f_t^{(7)}, f_t^{(8)}, f_t^{(9)})' \), while equation (17) is of the form of equation (A-17) with \( Y_{t1} = X_{t+1} \).

I estimate the vector system using a partially restricted algorithm equation by equation. The algorithm for a single equation is described in Appendix D. Suppose at the previous iteration of the algorithm I have estimates \( \theta^{(l)} \) and \( \Pr(s_t = j | \theta^{(l)}, Y) \) for \( Y = \{ Y_{t1}, Y_{t2}, X_t \}_t^{T=1} \). Iteration \( (l + 1) \) of the algorithm works as follows.

**Step 1.** Update inference for \( Y_{t2} \) regression parameters.

\[ \mathcal{F}_T \] represents information available up to time \( T \).
1a) Taking each \( n = 1, \ldots, 7 \) one at a time starting with \( n = 1 \), construct

\[
\omega_n^{2(\ell)} = \text{row } n, \text{ col. } n \text{ element of } \Omega^{(\ell)}
\]

\[
\lambda_{nt1}^{(\ell)} = \frac{\sqrt{\Pr(s_t = 1|Y, \theta^{(\ell)})}}{\omega_n^{(\ell)}}
\]

\[
\lambda_{nt2}^{(\ell)} = \frac{\sqrt{\Pr(s_t = 2|Y, \theta^{(\ell)})}}{\omega_n^{(\ell)}}
\]

\( Y_{t2}^{(n)} \) = \( n \)th element of \( Y_{t2} \)

\( A_{tn}^{(\ell)} \) = \( n \)th element of \( A_i^{(\ell)} \)

\( B_n^{(\ell)} \) = \( n \)th element of \( B^{(\ell)} \)

For \( t = 1, \ldots, T \), define

\[
\tilde{y}_n^{(\ell)} = \lambda_{nt1}^{(\ell)} Y_{t2}^{(n)}
\]

\[
\tilde{x}_n = \lambda_{nt1}^{(\ell)} X_t
\]

\[
\tilde{z}_{n1t}^{(\ell)} = \lambda_{nt1}^{(\ell)}
\]

\[
\tilde{z}_{n2t}^{(\ell)} = 0
\]

and

\[
\tilde{y}_{n,T+t}^{(\ell)} = \lambda_{nt2}^{(\ell)} Y_{t2}^{(n)}
\]

\[
\tilde{x}_{n,T+t} = \lambda_{nt2}^{(\ell)} X_t
\]

\[
\tilde{z}_{n1,T+t}^{(\ell)} = 0
\]

\[
\tilde{z}_{n2,T+t}^{(\ell)} = \lambda_{nt2}^{(\ell)}
\]

Conditional on knowing \( \lambda_{nt1}^{(\ell)} \) and \( \lambda_{nt2}^{(\ell)} \), a closed-form solution for \( (\hat{A}_1^{(n)}, \hat{A}_2^{(n)}, \hat{B}^{(n)})' \) can be found by performing an OLS regression on an artificial sample of size \( 2T \),

\[
\tilde{y}_n^{(\ell)} = A_1^{(n)} \tilde{z}_{n1t}^{(\ell)} + A_2^{(n)} \tilde{z}_{n2t}^{(\ell)} + B^{(n)} \tilde{x}_n + \tilde{u}_t, \ t = 1, 2, \ldots, 2T
\]

Construct

\[
\tilde{u}_{nt1}^{(\ell+1)} = Y_{t2}^{(n)} - \hat{A}_{1n}^{(\ell+1)} - \hat{B}_n^{(\ell+1)} X_t
\]

\[
\tilde{u}_{nt2}^{(\ell+1)} = Y_{t2}^{(n)} - \hat{A}_{2n}^{(\ell+1)} - \hat{B}_n^{(\ell+1)} X_t
\]
1b) For each \( n = 1, \ldots, 7 \) calculate

\[
\omega_n^{(\ell+1)} = \left\{ \frac{1}{T} \left( \sum_{t=1}^{T} \hat{u}_{nt1}^{2(\ell+1)} Pr(s_t = 1|Y, \theta^{(\ell)}) + \sum_{t=1}^{T} \hat{u}_{nt2}^{2(\ell+1)} Pr(s_t = 2|Y, \theta^{(\ell)}) \right) \right\}^{1/2}
\]

**Step 2.** Update the inference for the \( Y_{t1} \) parameters. This involves the analogous steps to those above using the partially restricted algorithm for a single equation as described in Appendix D. The factor variance is updated as

\[
\sigma_1^{2(\ell+1)} = \frac{\sum_{t=1}^{T} Pr(s_t = 1|Y, \theta^{(\ell)}) (Y_{t1} - \mu_1^{(\ell+1)} - \Phi_1^{(\ell+1)} X_t)^2}{\sum_{t=1}^{T} Pr(s_t = 1|Y, \theta^{(\ell)})}
\]

\[
\sigma_2^{2(\ell+1)} = \frac{\sum_{t=1}^{T} Pr(s_t = 2|Y, \theta^{(\ell)}) (Y_{t1} - \mu_2^{(\ell+1)} - \Phi_2^{(\ell+1)} X_t)^2}{\sum_{t=1}^{T} Pr(s_t = 2|Y, \theta^{(\ell)})}
\]

**Step 3.** Update the inference about the transition probabilities. The transition probabilities are updated as

\[
\hat{\pi}^{P_{ij}(\ell+1)} = \frac{\sum_{t=2}^{T} Pr(s_t = j, s_{t-1} = i|Y_T, \theta^{(\ell)})}{\sum_{t=2}^{T} Pr(s_{t-1} = i|Y_T, \theta^{(\ell)})}
\]

Specifically,

\[
\hat{\pi}^{P_{11}(\ell+1)} = \frac{\sum_{t=2}^{T} \pi_{11}(\ell) Pr(s_t=1|Y_{t+1}, \theta^{(\ell)}) Pr(s_{t-1}=1|Y_{t-1}, \theta^{(\ell)})}{\sum_{t=2}^{T} Pr(s_{t-1} = 1|Y_T, \theta^{(\ell)})}
\]

\[
\hat{\pi}^{P_{21}(\ell+1)} = \frac{\sum_{t=2}^{T} \pi_{21}(\ell) Pr(s_t=1|Y_{t+1}, \theta^{(\ell)}) Pr(s_{t-1}=2|Y_{t-1}, \theta^{(\ell)})}{\sum_{t=2}^{T} Pr(s_{t-1} = 2|Y_T, \theta^{(\ell)})}
\]

**Step 4.** Update the inference about smoothed probabilities. This step calculates the smoothed probabilities \( P(s_t = j|Y, \theta^{(\ell+1)}) \) using the Hamilton filtering and smoothing algorithms, which are described in Appendix E. The initial probability vector \( \rho \) is updated as

\[
\rho^{(\ell+1)}_j = Pr(s_1 = j|Y_T, \theta^{(\ell)})
\]

**D EM algorithm for scalar regression**

Here I present the general form of the EM algorithm I use for estimation of each equation from my regime-switching vector system. Suppose the variances and some but not all of the
parameters change with the regime, that is
\[ y_t = x_t'\beta + z_t'c_{st} + \sigma_{st}v_t \]
for \( y_t \) a scalar, \( x_t \) an \((m \times 1)\) vector, \( z_t \) an \((r \times 1)\) vector, and \( v_t \sim N(0, 1)\). Thus
\[
\eta_t = \begin{bmatrix}
\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left\{ \frac{-(y_t-x_t'\beta-z_t'c_{s1})^2}{2\sigma_1^2} \right\} \\
\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp \left\{ \frac{-(y_t-x_t'\beta-z_t'c_{s2})^2}{2\sigma_2^2} \right\}
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial \beta'} = \begin{bmatrix}
(y_t-x_t'\beta-z_t'c_{s1})x_t' \\
(y_t-x_t'\beta-z_t'c_{s2})x_t'
\end{bmatrix}
\]

(A-19)

\[
\frac{\partial \log \eta_t}{\partial c_{s1}'} = \begin{bmatrix}
\frac{(y_t-x_t'\beta-z_t'c_{s1})z_t'}{\sigma_1^2} \\
0
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial c_{s2}'} = \begin{bmatrix}
0 \\
\frac{(y_t-x_t'\beta-z_t'c_{s2})z_t'}{\sigma_2^2}
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial \sigma_1^2} = \begin{bmatrix}
\frac{1}{2\sigma_1^2} + \frac{(y_t-x_t'\beta-z_t'c_{s1})^2}{2\sigma_1^2} \\
0
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial \sigma_2^2} = \begin{bmatrix}
0 \\
\frac{1}{2\sigma_2^2} + \frac{(y_t-x_t'\beta-z_t'c_{s2})^2}{2\sigma_2^2}
\end{bmatrix}
\]

The MLE for \( \theta = (\beta', c_{s1}', c_{s2}', \sigma_1^2, \sigma_2^2)' \) satisfies
\[
\sum_{t=1}^{T} \left( \frac{\partial \log \eta_t}{\partial \theta'} \right)' \hat{\xi}_{t|T} = 0.
\]

Define
\[
\begin{bmatrix}
\lambda_{t1} \\
\lambda_{t2}
\end{bmatrix} = \begin{bmatrix}
\sigma_1^{-1}\sqrt{\Pr(s_t = 1|Y)} \\
\sigma_2^{-1}\sqrt{\Pr(s_t = 2|Y)}
\end{bmatrix}
\]

for \( Y = \{y_t, x_t, z_t\}_{t=1}^{T} \) the full set of observed data. Then the FOC associated with choice of \( \beta \) (using equation (A-19)) can be written

30
\[
\left( \sum_{t=1}^{T} x_t y_t \lambda_{t1}^2 + \sum_{t=1}^{T} x_t y_t \lambda_{t2}^2 \right) = \left( \sum_{t=1}^{T} x_t x_t' \lambda_{t1}^2 + \sum_{t=1}^{T} x_t x_t' \lambda_{t2}^2 \right) \beta \\
+ \left( \sum_{t=1}^{T} x_t z_t' \lambda_{t1}^2 \right) c_1 + \left( \sum_{t=1}^{T} x_t z_t' \lambda_{t2}^2 \right) c_2.
\]

Take the analogous FOC for choice of \( c_1 \) and \( c_2 \) and stack the three equations together:

\[
\begin{bmatrix}
\left( \sum_{t=1}^{T} x_t y_t \lambda_{t1}^2 + \sum_{t=1}^{T} x_t y_t \lambda_{t2}^2 \right) \\
\left( \sum_{t=1}^{T} x_t x_t' \lambda_{t1}^2 \right) \\
\left( \sum_{t=1}^{T} x_t x_t' \lambda_{t2}^2 \right) \\
\left( \sum_{t=1}^{T} z_t x_t' \lambda_{t1}^2 \right) \\
\left( \sum_{t=1}^{T} z_t x_t' \lambda_{t2}^2 \right) \\
\left( \sum_{t=1}^{T} z_t z_t' \lambda_{t1}^2 \right) \\
\left( \sum_{t=1}^{T} z_t z_t' \lambda_{t2}^2 \right)
\end{bmatrix} = \begin{bmatrix} \beta \\ c_1 \\ c_2 \end{bmatrix}.
\]

Conditional on knowing \( \lambda_{t1} \) and \( \lambda_{t2} \), a closed-form solution for \( (\hat{\beta}', \hat{c}_1', \hat{c}_2')' \) can be found by performing a single OLS regression on an artificial sample of size \( 2T \),

\[
\tilde{y}_t = \tilde{x}_t' \beta + \tilde{z}_{t1}' c_1 + \tilde{z}_{t2}' c_2 + \tilde{v}_t \quad t = 1, 2, ..., 2T,
\]

where for \( t = 1, 2, ..., T \) I have defined

\[
\tilde{y}_t = y_t \lambda_{t1} \\
\tilde{x}_t = x_t \lambda_{t1} \\
\tilde{z}_{t1} = z_t \lambda_{t1} \\
\tilde{z}_{t2} = 0
\]

whereas the next \( T \) observations (denoted \( T + t \) for \( t = 1, ..., T \)) are from

\[
\tilde{y}_{T+t} = y_t \lambda_{t2} \\
\tilde{x}_{T+t} = x_t \lambda_{t2} \\
\tilde{z}_{T+t,1} = 0
\]
\[ \hat{z}_{T+t,2} = z_t\lambda_2. \]

The OLS coefficients from this artificial system are given by

\[
\begin{bmatrix}
\hat{\beta} \\
\hat{c}_1 \\
\hat{c}_2
\end{bmatrix} = 
\begin{bmatrix}
\sum_{t=1}^{2T} \tilde{x}_t \tilde{x}_t' \\
\sum_{t=1}^{2T} \tilde{z}_{t1} \tilde{x}_t' \\
\sum_{t=1}^{2T} \tilde{z}_{t2} \tilde{x}_t'
\end{bmatrix}^{-1}
\begin{bmatrix}
\sum_{t=1}^{2T} \tilde{x}_t \tilde{y}_t \\
\sum_{t=1}^{2T} \tilde{z}_{t1} \tilde{y}_t \\
\sum_{t=1}^{2T} \tilde{z}_{t2} \tilde{y}_t
\end{bmatrix}
\]

\[
= 
\begin{bmatrix}
\left( \sum_{t=1}^{T} x_t x_t' + \sum_{t=1}^{T} x_t z_{t1}' \lambda_1^2 \right) & \left( \sum_{t=1}^{T} x_t z_{t1}' \lambda_1^2 \right) & 0 \\
\left( \sum_{t=1}^{T} z_{t1} z_{t1}' \lambda_1^2 \right) & 0 & \left( \sum_{t=1}^{T} z_{t2} z_{t2}' \lambda_2^2 \right)
\end{bmatrix}^{-1}
\times
\begin{bmatrix}
\left( \sum_{t=1}^{T} x_t y_t \lambda_1^2 \right) & \left( \sum_{t=1}^{T} x_t y_t \lambda_2^2 \right) \\
\left( \sum_{t=1}^{T} z_{t1} y_t \lambda_1^2 \right) & \left( \sum_{t=1}^{T} z_{t2} y_t \lambda_2^2 \right)
\end{bmatrix}
\]

which will be recognized as a closed-form solution to the FOC for the MLE as given in equation (A-20).

Thus an EM algorithm would work as follows. At the previous step I have calculated estimates \( \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\xi}_{t|T} \), from which I can construct \( \lambda_{t1} \) and \( \lambda_{t2} \). I then use these to construct \( \{\tilde{y}_t, \tilde{x}_t, \tilde{z}_{t1}, \tilde{z}_{t2}\}_{t=1}^{2T} \) and do an OLS regression of \( \tilde{y}_t \) on \( \tilde{x}_t, \tilde{z}_{t1}, \tilde{z}_{t2} \) to get new estimates of \( \hat{\beta}, c_1, c_2 \).

Taking first order conditions for \( \sigma_1^2 \) and \( \sigma_2^2 \) results in the following expressions for the next step estimates:

\[
\hat{\sigma}_1^2 = \frac{\sum_{t=1}^{T} (y_t - x_t' \hat{\beta} - z_{t1}' \hat{c}_1)^2 \Pr(s_t = 1|Y)}{\sum_{t=1}^{T} \Pr(s_t = 1|Y)}
\]

\[
\hat{\sigma}_2^2 = \frac{\sum_{t=1}^{T} (y_t - x_t' \hat{\beta} - z_{t2}' \hat{c}_2)^2 \Pr(s_t = 2|Y)}{\sum_{t=1}^{T} \Pr(s_t = 2|Y)}
\]

### E Filtering and smoothing algorithm

Let

\[ \xi_t = \begin{bmatrix} 1 \{s_t = 1\} \\ 1 \{s_t = 2\} \end{bmatrix} \quad (A-21) \]

Let \( \hat{\xi}_{t|T} = E(\xi_t|Y_t) \). Then

\[ \hat{\xi}_{t|T} = \begin{bmatrix} \Pr(\xi_t = e_1|Y_t) \\ \Pr(\xi_t = e_2|Y_t) \end{bmatrix} \quad (A-22) \]

32
where $Y_\tau$ consists of information available up to time $\tau$, $e_1 = (1, 0)'$, $e_2 = (0, 1)'$. Let $y_t$ be the vector of dependent variables of all the equations. Let $\eta_t$ be the vector of densities of $y_t$ conditional on $\xi_t$ and $Y_{t-1}$:

$$
\eta_t = \begin{bmatrix}
    p(y_t|\theta_1, Y_{t-1}) \\
    p(y_t|\theta_2, Y_{t-1}) \\
    p(y_t|\xi_t = e_1, Y_{t-1}) \\
    p(y_t|\xi_t = e_2, Y_{t-1})
\end{bmatrix}
$$  \hspace{1cm} (A-23)

where $\theta$ has been dropped on the right hand side for brevity.

In my model,

$$
\eta_{t1} = (2\pi)^{-(K+N)/2} |\Psi_1|^{-1/2} \exp \left[ -\frac{1}{2} \left( \begin{array}{c}
    Y_{1t} \\
    Y_{2t}
\end{array} \right) - \left( \begin{array}{c}
    \mu_1 + \Phi X_t \\
    A_1 + BX_t
\end{array} \right) \right] \Psi_1^{-1} \left[ \begin{array}{c}
    \left( Y_{1t} \\
    Y_{2t}
\end{array} \right) - \left( \begin{array}{c}
    \mu_1 + \Phi X_t \\
    A_1 + BX_t
\end{array} \right) \right]
$$

$$
\eta_{t2} = (2\pi)^{-(K+N)/2} |\Psi_2|^{-1/2} \exp \left[ -\frac{1}{2} \left( \begin{array}{c}
    Y_{1t} \\
    Y_{2t}
\end{array} \right) - \left( \begin{array}{c}
    \mu_2 + \Phi X_t \\
    A_2 + BX_t
\end{array} \right) \right] \Psi_2^{-1} \left[ \begin{array}{c}
    \left( Y_{1t} \\
    Y_{2t}
\end{array} \right) - \left( \begin{array}{c}
    \mu_2 + \Phi X_t \\
    A_2 + BX_t
\end{array} \right) \right]
$$

where

$$
\Psi_j = \begin{bmatrix}
    \sigma_j^2 & 0_{K \times N} \\
    0_{N \times K} & \Omega_j
\end{bmatrix}
$$

and $K = 1$ and $N = 7$.

The density of $y_t$ conditional on $Y_{t-1}$ is given by $p(y_t|Y_{t-1}) = \eta_t \hat{\xi}_{t|t-1} = I_2(\eta_t \odot \hat{\xi}_{t|t-1})$ where $\odot$ signifies element-wise matrix multiplication. The contemporaneous inference $\hat{\xi}_{t|t}$ about the unobserved state vector $\xi_t$ is given in matrix notation by the filtering recursions

$$
\hat{\xi}_{t|t} = \frac{\eta_t \odot \hat{\xi}_{t|t-1}}{I_2(\eta_t \odot \hat{\xi}_{t|t-1})} \hspace{1cm} (A-24)
$$

$$
\hat{\xi}_{t+1|t} = P \cdot \hat{\xi}_{t|t} \hspace{1cm} (A-25)
$$

where $P$ is the matrix of transition probabilities. The recursion is initialized with

$$
\hat{\xi}_{1|0} = \rho
$$

The smoothed inference about the unobserved state vector $\xi_t$ is given by

$$
\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \left( P' (\hat{\xi}_{t+1|T} (\div) \hat{\xi}_{t+1|t}) \right) \hspace{1cm} (A-26)
$$

where the sign $(\div)$ denotes element-by-element division. The smoothed probabilities $\hat{\xi}_{t|T}$ are found by iteration on equation (A-26) backward for $t = T - 1, T - 2, \cdots, 1$. This iteration
is started with $\hat{\xi}_{T\mid T}$, which is obtained from equation (A-24) for $t = T$.

F Estimation of number of factors

I use the Eigenvalue Ratio and Growth Ratio tests proposed in Ahn and Horenstein (2013) in order to estimate the number of factors in my model. Let $Y$ be the $T \times N$ matrix containing the demeaned futures price data, with $T = 234$ and $N = 7$, and let $\hat{\lambda}_k$ denote the $k^{th}$ largest eigenvalue of the covariance matrix $(Y'Y)/NT$. The Eigenvalue Ratio criterion function $ER(k)$ is the ratio of two adjacent eigenvalues of $(Y'Y)/NT$:

$$ER(k) \equiv \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}}, k = 1, 2, \ldots, k_{max}$$  \hspace{1cm} (A-27)

where $k$ is the number of factors used, and $k_{max}$ is a specified maximum number of factors.

The Growth Ratio criterion function $GR(k)$ is given by

$$GR(k) \equiv \frac{\log(1 + \hat{\lambda}_k^*)}{\log(1 + \hat{\lambda}_{k+1}^*)}$$  \hspace{1cm} (A-28)

where $V(k) = \sum_{j=k+1}^{m} \hat{\lambda}_j$ and $\hat{\lambda}_k^* = \hat{\lambda}_k/V(k)$.

The estimators of the true number of factors $r$ are the maximizers of $ER(k)$ and $GR(k)$:

$$\hat{r}_{ER} = \max_{1 \leq k \leq k_{max}} ER(k)$$  \hspace{1cm} (A-29)

$$\hat{r}_{GR} = \max_{1 \leq k \leq k_{max}} GR(k)$$  \hspace{1cm} (A-30)

These estimators are called the ER and GR estimators, respectively.
G References


Figure 1: Log of the observed three month natural gas futures price. The natural gas futures price is measured in dollars per million Btu.
Figure 2: The spread between the log of the nine month futures price and the log of the three month futures price (in blue) vs. smoothed probability of regime 2 (the backwardation regime) (in red). Shaded areas represent the backwardation regime.
Figure 3: Log of the observed three month natural gas futures price. Shaded areas represent the backwardation regime.
Figure 4: First principal component (in blue) vs. smoothed probability of the backwardation regime (in red). Shaded areas represent the backwardation regime.
Figure 5: Expected monthly holding returns (%) conditional on the contango regime (regime 1, in blue) and conditional on the backwardation regime (regime 2, in red), averaged over each regime.
Table 1: First stage reduced form parameter estimates

<table>
<thead>
<tr>
<th></th>
<th>Regime 1</th>
<th>Regime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>0.0365 (0.0189)</td>
<td>-0.0412 (0.0292)</td>
</tr>
<tr>
<td>( \Phi )</td>
<td>0.9785 (0.0113)</td>
<td></td>
</tr>
<tr>
<td>( A_{(3)} )</td>
<td>1.3803 (0.0058)</td>
<td>1.4891 (0.0097)</td>
</tr>
<tr>
<td>( A_{(4)} )</td>
<td>1.3926 (0.0037)</td>
<td>1.4743 (0.0058)</td>
</tr>
<tr>
<td>( A_{(5)} )</td>
<td>1.4120 (0.0031)</td>
<td>1.4555 (0.0044)</td>
</tr>
<tr>
<td>( A_{(6)} )</td>
<td>1.4328 (0.0030)</td>
<td>1.4309 (0.0045)</td>
</tr>
<tr>
<td>( A_{(7)} )</td>
<td>1.4524 (0.0031)</td>
<td>1.4064 (0.0050)</td>
</tr>
<tr>
<td>( A_{(8)} )</td>
<td>1.4692 (0.0036)</td>
<td>1.3863 (0.0055)</td>
</tr>
<tr>
<td>( A_{(9)} )</td>
<td>1.4786 (0.0049)</td>
<td>1.3778 (0.0070)</td>
</tr>
<tr>
<td>( B_{(3)} )</td>
<td></td>
<td>0.3724 (0.0035)</td>
</tr>
<tr>
<td>( B_{(4)} )</td>
<td></td>
<td>0.3760 (0.0022)</td>
</tr>
<tr>
<td>( B_{(5)} )</td>
<td></td>
<td>0.3789 (0.0018)</td>
</tr>
<tr>
<td>( B_{(6)} )</td>
<td></td>
<td>0.3808 (0.0017)</td>
</tr>
<tr>
<td>( B_{(7)} )</td>
<td></td>
<td>0.3809 (0.0019)</td>
</tr>
<tr>
<td>( B_{(8)} )</td>
<td></td>
<td>0.3797 (0.0021)</td>
</tr>
<tr>
<td>( B_{(9)} )</td>
<td></td>
<td>0.3770 (0.0028)</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.0528 (0.0061)</td>
<td>0.0584 (0.0098)</td>
</tr>
<tr>
<td>( \Omega_{(3)} )</td>
<td></td>
<td>0.0048 (0.0005)</td>
</tr>
<tr>
<td>( \Omega_{(4)} )</td>
<td></td>
<td>0.0018 (0.0002)</td>
</tr>
<tr>
<td>( \Omega_{(5)} )</td>
<td></td>
<td>0.0013 (0.0001)</td>
</tr>
</tbody>
</table>
Table 1: First stage reduced form parameter estimates (continued)

<table>
<thead>
<tr>
<th>Regime 1</th>
<th>Regime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega_{(6)}$</td>
<td>0.0013 (0.0001)</td>
</tr>
<tr>
<td>$\Omega_{(7)}$</td>
<td>0.0014 (0.0001)</td>
</tr>
<tr>
<td>$\Omega_{(8)}$</td>
<td>0.0017 (0.0002)</td>
</tr>
<tr>
<td>$\Omega_{(9)}$</td>
<td>0.0029 (0.0003)</td>
</tr>
<tr>
<td>$\pi^{P_{11}}$</td>
<td>0.9053 (0.0252)</td>
</tr>
<tr>
<td>$\pi^{P_{22}}$</td>
<td>0.8023 (0.0480)</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>1.0000</td>
</tr>
<tr>
<td>logL</td>
<td>2721.6</td>
</tr>
</tbody>
</table>

Asymptotic standard errors are in parentheses
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Regime 1</th>
<th>Regime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.0357</td>
<td>-0.0396</td>
</tr>
<tr>
<td></td>
<td>(0.0189)</td>
<td>(0.0292)</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>0.9785</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0113)</td>
<td></td>
</tr>
<tr>
<td>$A_{(3)}$</td>
<td>1.3743</td>
<td>1.4959</td>
</tr>
<tr>
<td></td>
<td>(0.0046)</td>
<td>(0.0083)</td>
</tr>
<tr>
<td>$B_{(3)}$</td>
<td>0.3767</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0528</td>
<td>0.0586</td>
</tr>
<tr>
<td></td>
<td>(0.0061)</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\Omega_{(3)}$</td>
<td>0.0048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td></td>
</tr>
<tr>
<td>$\Omega_{(4)}$</td>
<td>0.0018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td></td>
</tr>
<tr>
<td>$\Omega_{(5)}$</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>$\Omega_{(6)}$</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>$\Omega_{(7)}$</td>
<td>0.0014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>$\Omega_{(8)}$</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td></td>
</tr>
<tr>
<td>$\Omega_{(9)}$</td>
<td>0.0029</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td></td>
</tr>
<tr>
<td>$\pi_{P_{11}}$</td>
<td>0.9070</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0252)</td>
<td></td>
</tr>
<tr>
<td>$\pi_{P_{22}}$</td>
<td>0.8060</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0479)</td>
<td></td>
</tr>
<tr>
<td>$\pi_{Q_{11}}$</td>
<td>0.9776</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0231)</td>
<td></td>
</tr>
<tr>
<td>$\pi_{Q_{22}}$</td>
<td>0.9813</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0383)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_0$</td>
<td>-0.0008</td>
<td>0.0257</td>
</tr>
<tr>
<td></td>
<td>(0.0189)</td>
<td>(0.0295)</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.0233</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0114)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Wald t-statistics

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>Estimate</th>
<th>Standard error</th>
<th>Wald t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reduced-form parameters (1st stage)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_{(4),1} - A_{(4),2} = 0$</td>
<td>-0.0817</td>
<td>0.0063</td>
<td>-13.0459</td>
</tr>
<tr>
<td>$A_{(5),1} - A_{(5),2} = 0$</td>
<td>-0.0435</td>
<td>0.0053</td>
<td>-8.1938</td>
</tr>
<tr>
<td>$A_{(6),1} - A_{(6),2} = 0$</td>
<td>0.0019</td>
<td>0.0056</td>
<td>0.3393</td>
</tr>
<tr>
<td>$A_{(7),1} - A_{(7),2} = 0$</td>
<td>0.0460</td>
<td>0.0057</td>
<td>8.0412</td>
</tr>
<tr>
<td>$A_{(8),1} - A_{(8),2} = 0$</td>
<td>0.0829</td>
<td>0.0059</td>
<td>14.0888</td>
</tr>
<tr>
<td>$A_{(9),1} - A_{(9),2} = 0$</td>
<td>0.1008</td>
<td>0.0079</td>
<td>12.7908</td>
</tr>
<tr>
<td><strong>Reduced-form parameters (2nd stage)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1 - \mu_2 = 0$</td>
<td>0.0753</td>
<td>0.0357</td>
<td>2.1076</td>
</tr>
<tr>
<td>$\sigma_1^2 - \sigma_2^2 = 0$</td>
<td>-0.0057</td>
<td>0.0117</td>
<td>-0.4903</td>
</tr>
<tr>
<td>$A_{(3),1} - A_{(3),2} = 0$</td>
<td>-0.1216</td>
<td>0.0087</td>
<td>-13.9022</td>
</tr>
<tr>
<td><strong>Structural parameters (2nd stage)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{0.1} - \lambda_{0.2} = 0$</td>
<td>0.0489</td>
<td>0.0360</td>
<td>1.3575</td>
</tr>
<tr>
<td>$\pi_{P11} - \pi_{Q11} = 0$</td>
<td>-0.0706</td>
<td>0.0351</td>
<td>-2.0078</td>
</tr>
<tr>
<td>$\pi_{P22} - \pi_{Q22} = 0$</td>
<td>-0.1753</td>
<td>0.0605</td>
<td>-2.8996</td>
</tr>
</tbody>
</table>
Table 4: Reduced-form versus model implied values for $A_{(n),j}$ and $B_{(n)}$

<table>
<thead>
<tr>
<th></th>
<th>First stage estimates</th>
<th>Model implied values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{(3),1}$</td>
<td>1.3803</td>
<td>1.3743</td>
</tr>
<tr>
<td>$A_{(3),2}$</td>
<td>1.4891</td>
<td>1.4959</td>
</tr>
<tr>
<td>$A_{(4),1}$</td>
<td>1.3926</td>
<td>1.3948</td>
</tr>
<tr>
<td>$A_{(4),2}$</td>
<td>1.4743</td>
<td>1.4734</td>
</tr>
<tr>
<td>$A_{(5),1}$</td>
<td>1.4120</td>
<td>1.4142</td>
</tr>
<tr>
<td>$A_{(5),2}$</td>
<td>1.4555</td>
<td>1.4515</td>
</tr>
<tr>
<td>$A_{(6),1}$</td>
<td>1.4328</td>
<td>1.4326</td>
</tr>
<tr>
<td>$A_{(6),2}$</td>
<td>1.4309</td>
<td>1.4304</td>
</tr>
<tr>
<td>$A_{(7),1}$</td>
<td>1.4524</td>
<td>1.4502</td>
</tr>
<tr>
<td>$A_{(7),2}$</td>
<td>1.4064</td>
<td>1.4099</td>
</tr>
<tr>
<td>$A_{(8),1}$</td>
<td>1.4692</td>
<td>1.4670</td>
</tr>
<tr>
<td>$A_{(8),2}$</td>
<td>1.3863</td>
<td>1.3901</td>
</tr>
<tr>
<td>$A_{(9),1}$</td>
<td>1.4786</td>
<td>1.4831</td>
</tr>
<tr>
<td>$A_{(9),2}$</td>
<td>1.3778</td>
<td>1.3711</td>
</tr>
<tr>
<td>$B_{(3)}$</td>
<td>0.3724</td>
<td>0.3767</td>
</tr>
<tr>
<td>$B_{(4)}$</td>
<td>0.3760</td>
<td>0.3774</td>
</tr>
<tr>
<td>$B_{(5)}$</td>
<td>0.3789</td>
<td>0.3781</td>
</tr>
<tr>
<td>$B_{(6)}$</td>
<td>0.3808</td>
<td>0.3788</td>
</tr>
<tr>
<td>$B_{(7)}$</td>
<td>0.3809</td>
<td>0.3794</td>
</tr>
<tr>
<td>$B_{(8)}$</td>
<td>0.3797</td>
<td>0.3801</td>
</tr>
<tr>
<td>$B_{(9)}$</td>
<td>0.3770</td>
<td>0.3808</td>
</tr>
</tbody>
</table>