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

September 27, 2016

Abstract

This paper proposes a new approach to the estimation of Markov regime-switching affine term structure models. I show how to price the time series and cross-section of the term structure of commodities in the case of regime switching. The maximum likelihood based methods common in the literature pose significant numerical challenges, which become especially severe when there is the possibility of changes in regime. My approach avoids these numerical difficulties and allows for computationally fast estimation. I apply my approach to a model of natural gas futures prices from 1994 to 2013, 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. In my model, one regime corresponds to higher difference between the longer term and the shorter term contracts (higher spread) while the other regime corresponds to a lower spread. I find that the market acts as if regimes are more persistent than they really are. This could be a result of hedging pressure or mispricing. I find evidence that in the high spread regime, commercial producers are trying to hedge their short positions in natural gas by selling 3-5 month futures contracts, while in the low spread regime, commercial producers are trying to hedge their short positions by selling 6-9 month contracts. My model implies that it is more profitable to be long in the 3-5 month contracts in the high spread regime, whereas it is better to be long in the 6-9 month contracts in the low spread regime. I also find that a positive one standard deviation shock to the factor, which represents a change in prices that is basically constant across maturities, decreases the expected monthly return on the 3-9 month contracts by about 0.88%.

*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, Rossen Valkanov, Allan Timmermann, Ivana Konunjer, and Ruth Williams for their invaluable guidance and support. This paper also benefited from helpful suggestions from Julie Cullen, Mark Machina, Asad Dossani, and seminar participants at UCSD.


1 Introduction

Affine term structure models are a popular tool for the analysis of the pricing of fixed income securities such as bonds and futures. Hamilton and Wu (2011) 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 diversification. The class of Gaussian affine term structure models was originally developed by Vasicek (1977), Duffie and Kan (1996), Dai and Singleton (2000), Duffee (2002), and Piazzesi (2010) to characterize the relation between yields on bonds of different maturities. Hamilton and Wu (2011) adapt this class of models to commodities. Schwartz (1997), Schwartz and Smith (2000), and Casassus and Collin-Dufresne (2006), among others, also develop related models to describe commodity futures prices.

The affine term structure framework is based on the assumptions that the pricing kernel is exponentially 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. Most methods for estimating Gaussian affine term structure models in the literature are based on maximum likelihood estimation 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, I propose a new approach to the estimation of Markov regime-switching affine term structure models. I use a regression based method to estimate the reduced-form parameters in the first stage, and then estimate the prices of risk via minimum-chi-square estimation in the second stage. In this way, I bypass the numerical difficulties encountered by other methods and have no problem achieving convergence. I show how to price the time series and cross-section of the term structure of commodities in the case of regime switching. I apply my method to a model of natural gas futures contracts. Another contribution of this paper is providing novel empirical estimates to characterize risk premia
and the term structure of natural gas futures.

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. 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.

Natural gas spot and futures prices tend to exhibit seasonal variations 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. However, in recent years summer consumption has increased as utility companies have increased their demand for natural gas for air conditioning. Thus, natural gas consumption follows a seasonal pattern. Other factors affecting supply and demand also affect prices. On the supply side, amount of gas in storage, pipeline capacity, and imperfect information about storage affect prices. Natural gas supplies that were put in storage during periods of lower demand may be used to cushion the impact on price during periods of high demand. Factors affecting demand include weather conditions and aggregate economic conditions. In general, in the winter demand for natural gas is at its peak and shocks can cause large price fluctuations, since winter demand is inelastic and since supply is often unable to react quickly to short-term increases in demand. Thus, contracts for delivery in the winter may exhibit high volatility. At the same time, the inventory stored for the winter can partially cushion the impact of shocks and thus reduce the volatility of winter contracts. Prices of contracts for delivery in the winter can start exhibiting high volatility as soon as the early fall when information on the future availability of natural gas is released and expectations are formed on whether storage would be able to meet future demand (Suenaga, Smith, and Williams 2008). In contrast, prices of contracts for delivery in the spring tend not to exhibit high volatility, since demand is low in the spring.

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, thereby putting downward pressure on prices and decreasing the volatility in the spot and futures markets.

Although these fluctuations have a seasonal component, they are far from determinis-
tic. In some years there is virtually no winter spike in gas prices, and in others the high and volatile prices extend well beyond the winter months. This uncertainty introduces the possibility that producers or commercial users may at times make significant use of natural gas futures contracts for purposes of hedging. Hamilton and Wu (2014) show how variation in hedging pressure could influence the term structure of commodity futures prices. In this paper I model the level and volatility of natural gas prices using a Markov regime-switching model and study the consequences of these changes in regime on the risk premium by generalizing the futures-pricing model in Hamilton and Wu (2014) to allow for changes in regime.

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 (2002) among others. Dai, Singleton, and Yang (2007) and Ang and Bekaert (2002) use maximum-likelihood based methods based on an iterative procedure developed by Hamilton (1989). Bansal and Zhou (2002) use a two-step efficient method of moments estimator. These methods are subject to the numerical issues mentioned earlier, which this paper resolves.

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. In this paper, I show how such an approach can be generalized to allow for regime switching, and produce empirical estimates to characterize risk premia and the term structure of natural gas futures contracts.

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 from applying my method to a one factor model of natural gas futures prices from 1994 to 2013, 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:

$$X_{t+1} = \mu_s t + \Phi X_t + \nu_{t+1}, \nu_{t+1} | s_t \sim N(0, \sigma^2_s)$$  \hspace{1cm} (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.$^1$ Thus, in my empirical application $X_t$ will be a scalar. I find that shorter term contracts have a

---

$^1$In section 5.2 I provide some evidence on why a one factor model is appropriate.
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. \(^2\)

The no-arbitrage assumption implies the existence of a pricing kernel \( M_t \) such that

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

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:

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

I also assume that the market price of risk \( \lambda_t \) is an affine function of the state variable:

\[
\lambda_{t,s_t} = \sigma_{s_t}^{-1} (\lambda_{0,s_t} + \lambda_1 X_t)
\]

Here the price of risk \( \lambda_{t,s_t} \) is time-varying and regime-dependent. I assume that \( \lambda_1 \) does not depend on the regime. \(^3\) No arbitrage implies the existence of an equivalent martingale measure - the risk neutral measure \( Q \). The historic measure \( P \) and the risk-neutral measure \( Q \) are related through the pricing kernel \( M_{t+1} \). I can compute the price \( P(X_t) \) of an asset with payoff \( g(X_{t+1}) \) in regime \( j \) as

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

Under the \( Q \)-measure, the factor \( X_t \) follows a Gaussian autoregression:

\[
X_{t+1} = \mu_{s_t}^Q + \Phi^Q X_t + v_{t+1}^Q
\]

where

\[
\mu_j^Q = \mu_j - \lambda_{0,j}
\]

\(^2\)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.

\(^3\)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 \( \lambda_{1,s_t} \). 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.
and
\[ \Phi^Q = \Phi - \lambda_1 \]  
\[ (8) \]
and \( v_{t+1}^Q|s_t = j \sim Q N(0, \sigma_j^2) \). 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:
\[ f_t^{(n)} = A_s^{(n)} + B^{(n)} X_t + u_t^{(n)} \]  
\[ (9) \]

From equation (8) and equation (A-8), 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.

Let \( r_{x_t^{(n-1)}} \) denote the one-month log holding return of a futures contract maturing in \( n \) periods:
\[ r_{x_t^{(n-1)}} = f_{t+1}^{(n-1)} - f_t^{(n)} \]  
\[ (10) \]
The holding return is the return on buying a futures contract with maturity \( n \) months in month \( t \) and then selling it as an \( (n - 1) \) period 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 \( \pi^Q_{jk}, 1 \leq j, k \leq 2 \), with \( \sum_{k=1}^2 \pi^Q_{jk} = 1 \), for \( j = 1, 2 \). I assume that the risk-neutral transition probabilities \( \pi^Q_{jk} \) and the real-world transition probabilities \( \pi^P_{jk} \) are regime independent. I allow \( \pi^P_{jk} \neq \pi^Q_{jk} \). Agents are presumed to know the history of the factor \( X_t \) and of 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 \( X_t \):
\[ f_t^{(n)} = A_s^{(n)} + B^{(n)} X_t + u_t^{(n)} \]  
\[ (11) \]
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:

\[
A_j^{(n)} = \log \left( \sum_{k=1}^{2} \pi^{Qjk} e^{A_k^{(n-1)}} \right) + B^{(n-1)}(\mu_j - \lambda_{0,j}) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2 \tag{12}
\]

for $j = 1, 2$ and

\[
B^{(n)} = B^{(n-1)}(\Phi - \lambda_1) \tag{13}
\]

Equations (12) and (13) are derived in Appendix B in equations (A-7) and (A-9). 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 of the logs 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:

\[
f_t^{(n)} = A_{st}^{(n)} + B^{(n)}X_t + u_t^{(n)}, n = 3, \ldots, 9 \tag{14}
\]
where

\[
\begin{pmatrix}
u_t^{(3)} \\
v_t^{(4)} \\
v_t^{(5)} \\
v_t^{(6)} \\
v_t^{(7)} \\
v_t^{(8)} \\
v_t^{(9)}
\end{pmatrix}
|s_t \sim N(0, \Omega)
\] (15)

and

\[
\Omega \equiv \begin{pmatrix}
\Omega^{(3)} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \Omega^{(4)} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \Omega^{(5)} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \Omega^{(6)} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \Omega^{(7)} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \Omega^{(8)} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \Omega^{(9)}
\end{pmatrix}
\] (16)

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

\[
X_{t+1} = \mu_{s_t} + \Phi_X + v_{t+1}, \quad v_{t+1}|s_t \sim N(0, \sigma_{s_t}^2)
\] (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 \(\lambda_{0,1}, \lambda_{0,2}, \text{ and } \lambda_{1}\). I choose the values of \(\lambda_{0,1}, \lambda_{0,2}, \text{ and } \lambda_{1}\) that most closely fit the no arbitrage restrictions in equations (12) and (13). Minimum-chi-square estimation 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 \(L(\pi; Y)\) denote the log likelihood for the entire sample, and let \(\hat{\pi} = \arg\max L(\pi; Y)\) de-
note the full-information maximum likelihood estimate. If $\hat{R}$ is a consistent estimate of the information matrix,

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

(18)

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

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

(19)

As noted by Hamilton and Wu (2012), the variance of $\hat{\theta}$ can be approximated with $T^{-1}(\hat{\Gamma}' \hat{\Gamma})^{-1}$ 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 $B^{(n)}$ (from the regime-switching regressions) and the values of $A_j^{(n)}$ and $B^{(n)}$ 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:

$$\hat{\pi} = \left(\hat{\mu}_1, \hat{\mu}_2, \hat{\Phi}, vec(\hat{A}'), vec(\hat{B}), \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\Omega}_{(3)}, \hat{\Omega}_{(4)}, \hat{\Omega}_{(5)}, \hat{\Omega}_{(6)}, \hat{\Omega}_{(7)}, \hat{\Omega}_{(8)}, \hat{\Omega}_{(9)}, \hat{\pi}^{F_{11}}, \hat{\pi}^{F_{22}}\right)'$$

(20)

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),

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

(21)

and $\hat{B} = \left(\hat{B}^{(3)}, \hat{B}^{(4)}, \hat{B}^{(5)}, \hat{B}^{(6)}, \hat{B}^{(7)}, \hat{B}^{(8)}, \hat{B}^{(9)}\right)'$ 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 $\hat{\pi}^{F_{11}}$ and $\hat{\pi}^{F_{22}}$ are the unrestricted maximum likelihood estimates of the regime switching probabilities $\pi^{F_{11}}$ and $\pi^{F_{22}}$ from the regime-switching VAR.
Let
\[ \theta = (\mu_1, \mu_2, \Phi, A_1^{(3)}, A_2^{(3)}, B^{(3)}, \sigma_1^2, \sigma_2^2, \Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5, \Omega_6, \Omega_7, \Omega_8, \Omega_9, \pi^{p_{11}}, \pi^{p_{22}}, \lambda_{0,1}, \lambda_{0,2}, \lambda_1, \pi^{Q_{11}}, \pi^{Q_{22}})' \] (22)
and
\[ g(\theta) = (\mu_1, \mu_2, \Phi, vec(A'(\theta)), vec(B(\theta)), \sigma_1^2, \sigma_2^2, \Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5, \Omega_6, \Omega_7, \Omega_8, \Omega_9, \pi^{p_{11}}, \pi^{p_{22}})' \] (23)

Here
\[ A(\theta) = \begin{bmatrix} A_1^{(3)} & A_2^{(3)} \\ A_1^{(4)} & A_2^{(4)} \\ A_1^{(5)} & A_2^{(5)} \\ A_1^{(6)} & A_2^{(6)} \\ A_1^{(7)} & A_2^{(7)} \\ A_1^{(8)} & A_2^{(8)} \\ A_1^{(9)} & A_2^{(9)} \end{bmatrix} \] (24)

and \[ B(\theta) = (B^{(3)}, B^{(4)}, B^{(5)}, B^{(6)}, B^{(7)}, B^{(8)}, B^{(9)})' \]. \( A_j^{(n)} \) in \( vec(A'(\theta)) \) and \( B^{(n)} \) in \( vec(B(\theta)) \) for \( n = 4, \ldots, 9 \) are defined by the no arbitrage restrictions from equations (12) and (13):
\[ A_j^{(n)} = \log \left( \sum_{k=1}^{2} \pi^{Q_{jk}} e^{A_k^{(n-1)}} \right) + B^{(n-1)}(\mu_j - \lambda_{0,j}) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2 \] (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 \( \Omega_{(n)}, n = 3 \ldots 9 \). Equations (25) and (26) represent restrictions implied by the model. By using the estimators for \( \lambda_{0,1}, \lambda_{0,2}, \) and \( \lambda_1 \) outlined above, I obtain the price of risk parameters that most closely fit these restrictions.
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, 1994 to August, 2013. 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 \( \{ A_j, \mu_j, \sigma_j^2 \} \), \( j = 1, 2, B, \Omega, \) 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.

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) 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 \((Y'Y)/NT\):

\[
ER(k) \equiv \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}}, \quad k = 1, 2, \ldots, k_{\text{max}}
\]  

(28)

where \( k \) is the number of factors used, and \( k_{\text{max}} \) is a specified maximum number of factors.

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

\[
GR(k) \equiv \frac{\log(1 + \hat{\lambda}_k^*)}{\log(1 + \hat{\lambda}_{k+1}^*)}
\]  

(29)

where \( V(k) = \sum_{j=k+1}^{\infty} \hat{\lambda}_j \) 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 \( GR(k) \):

\[
\hat{r}_{ER} = \max_{1 \leq k \leq k_{\text{max}}} ER(k)
\]  

(30)

\[
\hat{r}_{GR} = \max_{1 \leq k \leq k_{\text{max}}} GR(k)
\]  

(31)
These estimators are called the ER and GR estimators, respectively. Both of these estimators 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 risk and the risk-neutral transition probabilities.

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. I observe that in regime 1 the spread is higher and tends to be positive, whereas in regime 2 the spread is lower and tends to be negative. Therefore, I refer to regime 1 as the high spread regime and regime 2 as the low spread regime. Figures 3 and 4 show the log of the observed 3 month futures price and the factor, respectively, with shaded areas representing the low spread regime.

I test the restrictions $\pi^{P11} = \pi^{Q11}$ and $\pi^{P22} = \pi^{Q22}$, and find that they are rejected. Thus, my results suggest that $\pi^P \neq \pi^Q$. By allowing for $\pi^P \neq \pi^Q$ in my model, I have another channel through which risk preferences can affect expected returns. I find that $\pi^{Q11}_{11} > \pi^{P11}_{11}$ and $\pi^{Q22}_{22} > \pi^{P22}_{22}$. This suggests that investors act as though regimes are more persistent than they really are. This could be a result of hedging pressure or mispricing.

The ratio $\frac{\pi^{P}_{jk}}{\pi^{Q}_{jk}}$ can be interpreted as related to the market price of regime shift risk. Consider a security which pays $1 if the regime changes next month. 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

\[
\log E_t^P[\mathbb{1}_{\{s_{t+1} = k\}}|s_t = j] = \log \left( \frac{\pi^P_{jk}}{\pi^Q_{jk}} \right)
\]

Thus, \( \log \left( \frac{\pi^P_{jk}}{\pi^Q_{jk}} \right) \) 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 from regime \( j \) to regime \( k \). Denote

\[
\Gamma_{jk} \equiv \log \left( \frac{\pi^P_{jk}}{\pi^Q_{jk}} \right)
\]

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. Using my estimates for \( \pi^P_{jk} \) and \( \pi^Q_{jk} \), I find that the estimated market prices of regime shift risk are \( \hat{\Gamma}_{11} = -0.0750 \), \( \hat{\Gamma}_{12} = 1.4235 \), \( \hat{\Gamma}_{21} = 2.3400 \), \( \hat{\Gamma}_{22} = -0.1972 \).

For instance, a security which pays off $1 if the regime switches from regime 2 (low spread) today to regime 1 (high spread) 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 the low spread regime to the high spread regime 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 premium investors are willing to pay to hedge against regime shift risk is very low, reflecting the fact that they think the low spread regime is considerably more persistent than it actually is.

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

\[
P_t^{(1)} = E_t^Q[\mathbb{1}_{\{s_{t+1} = 2\}}|s_t = 1] = \pi^Q_{12} = 1 - \pi^Q_{11} = $0.0224
\]

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

\[ E_t^p [\Pi(s_{t+1} = 2) | s_t = 1] = \pi^{F_{12}} = 1 - \pi^{F_{11}} = 0.0930 \]

i.e. about 9 cents. Once again, the premium 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 low spread regime.

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

\[
E_t[F_{t+1}^{(n-1)}|s_t=j] = \left[ \sum_{k=1}^{2} \pi^{w_{jk}} e^{A_{k}^{(n-1)}} \right] e^{B^{(n-1)\sigma_j} \lambda_{t,j}} \tag{34}
\]

This expression is derived in Appendix B in equation (A-12).

Figure 5 shows the expected returns for each contract in the high spread regime and in the low spread regime, averaged over time. In the high spread regime, buying the 3-5 month contracts today and selling them next month on average yields a profit. So does shorting the 6-9 month contracts today and closing out the position next month. In the low spread regime, shorting the 3-5 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, it is more profitable to be long in the 3-5 month contracts in the high spread regime, whereas it is better to be long in the 6-9 month contracts in the low spread regime.

The fact that there is a positive return to a long position in the 3-5 month contracts in the high spread regime suggests that there is demand for short positions in futures. This could mean that in the high spread regime, commercial producers are trying to hedge their short positions in natural gas by selling 3-5 month futures contracts. Moreover, in the high spread regime there is a positive return to a short position in the 6-9 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 6-9 month contracts. Similarly, in the low spread 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 low spread regime there is a positive return to a short position in the 3-5 month contracts, which could be a result of commercial users trying to hedge their long positions by buying 3-5 month contracts.

The expected log return \( E_t^p [r_x^{(n-1)}_{t+1}] \) is related to the risk premium investors demand
for holding a futures contract maturing in $n$ months for 1 month. The expressions for the expected log returns conditional on the regime as a function of model parameters are derived in Appendix B.

$$
E_t^p \left[ r_{x_{t+1}}^{(n-1)} | s_t = j \right] = \sum_{k=1}^{2} \pi_{jk} A_k^{(n-1)} - \log \left( \sum_{k=1}^{2} \pi_{jk} e^{A_k^{(n-1)}} \right) + B^{(n-1)} \lambda_0, j + B^{(n-1)} \lambda_1 X_t
$$

The price of risk $\lambda_1$ is statistically significant and negative. This implies that an increase in the level of futures prices decreases the expected log returns on the 3-9 month contracts. The expected log return loading for the $n$-month contract is $B^{(n)} \lambda_1$. According to my estimates, a positive one standard deviation shock to the level factor reduces the expected log return on the 3-9 month contracts 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 risk in the two regimes.

### 6 Conclusion

In this paper I have proposed a new method for estimating regime-switching affine term structure models. 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. I use my approach to estimate a regime-switching affine term structure model for natural gas futures prices from 1994 to 2013, and find very strong evidence that there are regimes in the data. I find that one regime corresponds to a higher difference between the longer term and the shorter term contracts (higher spread) while the other regime corresponds to a lower difference between the longer term and the shorter term contracts (lower spread). My results show that the market acts as if regimes are more persistent than they really are. This could be a result of hedging pressure or mispricing. I find evidence that commercial users and commercial producers use natural gas futures contracts for purposes of hedging. In the high spread regime, commercial producers may be trying to hedge their short positions in natural gas by selling 3-5 month contracts, while in the low spread regime, commercial producers may be trying to hedge their short positions by selling 6-9 month contracts. I obtain analogous implications for commercial users. Moreover, I find that an increase in the level of futures prices decreases the expected returns on the
3-9 month contracts. According to my estimates, a positive one standard deviation shock to the level factor, which represents a change in prices that is essentially constant across maturities, reduces the expected return on the 3-9 month contracts by about 0.88%.
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

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

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

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

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

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

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

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

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

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

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

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

$$ = E_t^Q(g(X_{t+1})|s_t = j) $$

Therefore, under the $\mathbb{Q}$-measure,

$$ X_{t+1}|s_t = j \sim Q((\mu_j - \lambda_{0,j}) + (\Phi - 1)X_t, \sigma_j^2) $$  \hspace{1cm} (A-1)
or, equivalently,

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

where

\[ \mu_j^Q \equiv \mu_j - \lambda_{0,j} \]  
(A-3)

and

\[ \Phi^Q \equiv \Phi - \lambda_1 \]  
(A-4)

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

\[ X_{t+1} = \mu_j^Q + \Phi^Q X_t + v_{t+1}^Q \]  
(A-5)

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

B  Calculating expected returns

\[
E_t^P \left[ r x_{t+1}^{(n-1)} \mid s_t = j \right] = E_t^P \left[ f_{t+1}^{(n-1)} - f_t^{(n)} \mid s_t = j \right] \\
= E_t^P \left[ A_{n+1}^{(n-1)} + B^{(n-1)} X_{t+1} - A_{s_t}^{(n)} - B^{(n)} X_t \mid s_t = j \right] \\
= E_t^P \left[ A_{s_{t+1}}^{(n-1)} + B^{(n-1)} (\mu_s + \Phi X_t + v_{t+1}) - A_s - B^{(n)} X_t \mid s_t = j \right] \\
= \pi_{P1} A_1^{(n-1)} + \pi_{P2} A_2^{(n-1)} + B^{(n-1)} \mu_j + (B^{(n-1)} \Phi - B^{(n)}) X_t - A_j^{(n)} \\
\]  
(A-6)

The futures price is

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

Therefore,

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

The above equation implies the following recursions:

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

or equivalently

\[ A_j^{(n)} = \log \left( \sum_{k=1}^{2} \pi^{Q} jk e^{A_k^{(n-1)}} \right) + B^{(n-1)} (\mu_j - \lambda_{0,j}) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2 \] \hspace{1cm} (A-7)

and

\[ B^{(n)} = B^{(n-1)} \Phi^Q \] \hspace{1cm} (A-8)

or equivalently

\[ B^{(n)} = B^{(n-1)} (\Phi - \lambda_1) \] \hspace{1cm} (A-9)
$$E_t^p \left[ f_{t+1}^{(n-1)}|s_t = j \right] = \sum_{k=1}^{2} \pi^p_{jk} E_t \left[ f_{t+1}^{(n-1)}|s_t = j \right]$$

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

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

$$= \sum_{k=1}^{2} \pi^p_{jk} \left( A_k^{(n-1)} + B^{(n-1)} \mu_{st} + \Phi X_t|s_t = j \right)$$

$$= \sum_{k=1}^{2} \pi^p_{jk} \left( A_k^{(n-1)} + B^{(n-1)} (\mu_j + \Phi X_t) \right)$$

$$= \sum_{k=1}^{2} \pi^p_{jk} A_k^{(n-1)} + B^{(n-1)} (\mu_j + \Phi X_t)$$

$$= \sum_{k=1}^{2} \pi^p_{jk} A_k^{(n-1)} + B^{(n-1)} (\mu_j + \Phi X_t)$$

$$E_t^p \left[ r_{x_{t+1}}^{(n-1)}|s_t = j \right] = E_t^p \left[ f_{t+1}^{(n-1)} - f_t^{(n)}|s_t = j \right] = E_t^p \left[ f_{t+1}^{(n-1)}|s_t = j \right] - f_t^{(n)}$$

$$= \sum_{k=1}^{2} \pi^p_{jk} A_k^{(n-1)} + B^{(n-1)} (\mu_j + \Phi X_t) - \log \left( \sum_{k=1}^{2} \pi^p_{jk} e^{A_k^{(n-1)}} \right) -$$

$$B^{(n-1)} (\mu_j^Q + \Phi^Q X_t) - \frac{1}{2} B^{(n-1)^2} \sigma_j^2$$

$$= \sum_{k=1}^{2} \pi^p_{jk} A_k^{(n-1)} + B^{(n-1)} (\mu_j - \mu_j^Q) + B^{(n-1)} (\Phi - \Phi^Q) X_t -$$

$$\log \left( \sum_{k=1}^{2} \pi^Q_{jk} e^{A_k^{(n-1)}} \right) - \frac{1}{2} B^{(n-1)^2} \sigma_j^2$$

$$= \sum_{k=1}^{2} \pi^p_{jk} A_k^{(n-1)} - \log \left( \sum_{k=1}^{2} \pi^Q_{jk} e^{A_k^{(n-1)}} \right) + B^{(n-1)} \lambda_{0,j} + B^{(n-1)} \lambda_1 X_t -$$

$$\frac{1}{2} B^{(n-1)^2} \sigma_j^2$$

20
\[ E^P_t \left[ r_{x_{t+1}^{(n-1)}} \mid F_t \right] = \sum_{j=1}^{2} E^P_t \left[ r_{x_{t+1}^{(n-1)}} \mid s_t = j \right] P(s_t = j \mid F_t) \]

\[ = \sum_{j=1}^{2} \left[ 2 \pi P_{jk} A_k^{(n-1)} - \log \left( \sum_{k=1}^{2} \pi Q_{jk} e^A_k^{(n-1)} \right) + B^{(n-1)} \lambda_0 j + B^{(n-1)} \lambda_1 X_t - \frac{1}{2} B^{(n-1)^2} \sigma_j^2 \right] \times P(s_t = j \mid F_t) \]

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

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

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

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

\[ = \sum_{k=1}^{2} \pi Q_{jk} e^{A_k^{(n-1)}} \left[ e^{B^{(n-1)} (\mu^Q_j + \Phi^Q X_t + v_{t+1})} \mid s_t = j \right] \]

\[ = \sum_{k=1}^{2} \pi Q_{jk} e^{A_k^{(n-1)}} e^{B^{(n-1)} (\mu^Q_j + \Phi^Q X_t) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2} \]

(A-10)

\[ E^P_t \left[ F_{t+1}^{(n-1)} \mid s_t = j \right] = \sum_{k=1}^{2} \pi P_{jk} E^P_t \left[ F_{t+1}^{(n-1)k} \mid s_t = j \right] \]

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

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

\[ = \sum_{k=1}^{2} \pi P_{jk} e^{A_k^{(n-1)}} \left[ e^{B^{(n-1)} (\mu^P_j + \Phi^P X_t + v_{t+1})} \mid s_t = j \right] \]

\[ = \sum_{k=1}^{2} \pi P_{jk} e^{A_k^{(n-1)}} e^{B^{(n-1)} (\mu^P_j + \Phi^P X_t) + \frac{1}{2} B^{(n-1)^2} \sigma_j^2} \]

(A-11)
Then

$$E_t[F_{t+1}^{(n-1)}| s_t = j] = \frac{\sum_{j=1}^{n} \pi_j P_{jk} e^{A_k(n-1)}}{\sum_{j=1}^{n} \pi_j e^{A_k(n-1)}} e^{B(n-1)(\mu_j^p + \Phi \theta X_t)} + \frac{1}{2} B(n-1)^2 \sigma_j^2$$

$$= \frac{\sum_{j=1}^{n} \pi_j P_{jk} e^{A_k(n-1)}}{\sum_{j=1}^{n} \pi_j e^{A_k(n-1)}} e^{B(n-1)(\mu_j^q + \Phi \theta X_t)} + \frac{1}{2} B(n-1)^2 \sigma_j^2$$

$$= \frac{\sum_{j=1}^{n} \pi_j P_{jk} e^{A_k(n-1)}}{\sum_{j=1}^{n} \pi_j e^{A_k(n-1)}} e^{B(n-1)(\mu_j + \Phi \theta X_t)}$$

$$= \frac{\sum_{j=1}^{n} \pi_j e^{A_k(n-1)}}{\sum_{j=1}^{n} \pi_j e^{A_k(n-1)}} e^{B(n-1)\lambda_{0,j} + \lambda_1 X_t}$$

$$E_t[F_{t+1}^{(n-1)}| s_t = j] = \frac{\sum_{j=1}^{n} \pi_j P_{jk} e^{A_k(n-1)}}{\sum_{j=1}^{n} \pi_j e^{A_k(n-1)}} e^{B(n-1)\sigma_j \lambda_{t,j}}$$

$$\text{(A-12)}$$

## 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 $Pr(s_t = j|\mathcal{F}_T)$ 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, $\theta \equiv \{\text{vec}(A'), \text{vec}(B), \Phi, \Omega, \{\mu_j, \sigma_j\}_{j=1}^n\}$ 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 $Pr(s_t|Y, \theta)$ are replaced with the smoothed probabilities $Pr(s_t = j|Y, \theta^{(l-1)})$ computed in the previous iteration, for $Y = \{Y_{t1}, Y_{t2}, X_t\}_{t=1}^T$ 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 $\theta$.  

$^4\mathcal{F}_T$ represents information available up to time $T$
The regime-switching system I estimate is of the form

\[
Y_{t1} = \mu_{s_t} + \Phi X_t + \varepsilon_t \quad \varepsilon_t|s_t \sim N(0, \sigma^2_{s_t}) \quad (A-13)
\]

\[
Y_{t2} = A_{s_t} + B X_t + u_t \quad u_t|s_t \sim N(0, \Omega) \quad (A-14)
\]

Equation (14) when stacked across all maturities \( n = 3, \ldots, 9 \) is of the form of the above equation (A-14) with \( Y_{t2} = (f^{(3)}_t, f^{(4)}_t, f^{(5)}_t, f^{(6)}_t, f^{(7)}_t, f^{(8)}_t, f^{(9)}_t)' \), while equation (17) is of the form of equation (A-13) 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 \( \theta^{(\ell)} \) and \( \Pr(s_t = j|\theta^{(\ell)}, Y) \) for \( Y = \{Y_{t1}, Y_{t2}, X_t\}_{t=1}^T \).

**Iteration \( (\ell + 1) \)** of the algorithm works as follows.

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

1a) Taking each \( n = 1, \ldots, 7 \) one at a time starting with \( n = 1 \), construct

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

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

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

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

\( A_{n}^{(\xi)} = n^{th} \text{ element of } A_{n}^{(\xi)} \)

\( B_{n}^{(\xi)} = n^{th} \text{ element of } B_{n}^{(\xi)} \)

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

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

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

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

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

and

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

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

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

\[
\hat{u}_{nt2}^{(\ell+1)} = \tilde{y}_{nt2}^{(\ell)} - \hat{A}_{2n}^{(\ell)} - \hat{B}_{2n}^{(\ell)} X_t
\]

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

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

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

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

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

Specifically,

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

\[
\hat{\pi}_{21}^{(\ell+1)} = \frac{\sum_{t=2}^{T} \hat{\pi}_{21}^{(\ell)} \Pr(s_t = 2|Y_T, \theta^{(\ell)}) \Pr(s_{t-1} = 2|Y_{t-1}, \theta^{(\ell)})}{\sum_{t=2}^{T} \Pr(s_{t-1} = 2|Y_T, \theta^{(\ell)})}
\]
Step 4. Update the inference about smoothed probabilities. This step calculates the smoothed probabilities \( P(s_t = j | Y, \theta^{(l+1)}) \) using the Hamilton filtering and smoothing algorithms, which are described in Appendix E. The initial probability vector \( \rho \) is updated as

\[
\rho_j^{(l+1)} = Pr(s_1 = j | Y_T, \theta^{(l)})
\]

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

\[
y_t = x_t' \beta + z_t' c_{s_t} + \sigma_{s_t} v_t
\]

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

\[
\eta_t = \begin{bmatrix}
\frac{1}{\sqrt{2\pi \sigma_1^2}} \exp \left\{ \frac{-(y_t - x_t' \beta - z_t' c_1)^2}{2\sigma_1^2} \right\} \\
\frac{1}{\sqrt{2\pi \sigma_2^2}} \exp \left\{ \frac{-(y_t - x_t' \beta - z_t' c_2)^2}{2\sigma_2^2} \right\}
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial \beta'} = \begin{bmatrix}
\frac{y_t - x_t' \beta - z_t' c_1}{\sigma_1^2} \\
\frac{y_t - x_t' \beta - z_t' c_2}{\sigma_2^2}
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial c_1'} = \begin{bmatrix}
\frac{(y_t - x_t' \beta - z_t' c_1) x_t'}{\sigma_1^2} \\
0
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial c_2'} = \begin{bmatrix}
0 \\
\frac{(y_t - x_t' \beta - z_t' c_2) x_t'}{\sigma_2^2}
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial \sigma_1^2} = \begin{bmatrix}
- \frac{1}{2\sigma_1^2} + \frac{(y_t - x_t' \beta - z_t' c_1)^2}{2\sigma_1^4} \\
0
\end{bmatrix}
\]

\[
\frac{\partial \log \eta_t}{\partial \sigma_2^2} = \begin{bmatrix}
- \frac{1}{2\sigma_2^2} + \frac{(y_t - x_t' \beta - z_t' c_2)^2}{2\sigma_2^4} \\
0
\end{bmatrix}
\]

The MLE for \( \theta = (\beta', c_1', c_2', \sigma_1^2, \sigma_2^2)' \) 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-15)) can be written
\[
\left( \sum_{t=1}^T x_t y_t \lambda_{t1}^2 + \sum_{t=1}^T x_t y_t \lambda_{t2}^2 \right) = \left( \sum_{t=1}^T x_t x'_t \lambda_{t1}^2 + \sum_{t=1}^T x_t x'_t \lambda_{t2}^2 \right) \beta + \left( \sum_{t=1}^T x_t z'_t \lambda_{t1}^2 \right) c_1 + \left( \sum_{t=1}^T x_t z'_t \lambda_{t2}^2 \right) c_2.
\]

Take the analogous FOC for choice of \( c_1 \) and \( c_2 \) and stack the three equations together:
\[
\begin{bmatrix}
\sum_{t=1}^T x_t y_t \lambda_{t1}^2 + \sum_{t=1}^T x_t y_t \lambda_{t2}^2 \\
\sum_{t=1}^T x_t x'_t \lambda_{t1}^2 + \sum_{t=1}^T x_t x'_t \lambda_{t2}^2 \\
\sum_{t=1}^T x_t z'_t \lambda_{t1}^2 + \sum_{t=1}^T x_t z'_t \lambda_{t2}^2
\end{bmatrix} = \begin{bmatrix}
\beta \\
c_1 \\
c_2
\end{bmatrix}.
\]  

(A-16)

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
\[
\begin{align*}
\tilde{y}_t &= y_t \lambda_{t1} \\
\tilde{x}_t &= x_t \lambda_{t1} \\
\tilde{z}_{t1} &= z_t \lambda_{t1} \\
\tilde{z}_{t2} &= 0
\end{align*}
\]
whereas the next $T$ observations (denoted $T + t$ for $t = 1, \ldots, T$) are from

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

The OLS coefficients from this artificial system are given by

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

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

which will be recognized as a closed-form solution to the FOC for the MLE as given in equation (A-16).

Thus an EM algorithm would work as follows. At the previous step I have calculated estimates $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{c}_{t|T}$, 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 $\beta, c_1, c_2$.

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

\[
\hat{\sigma}_1^2 = \frac{\sum_{t=1}^{T} (y_t - x_t' \hat{\beta} - z_t' \hat{c}_1)^2 \Pr(s_t = 1|Y)}{\sum_{t=1}^{T} \Pr(s_t = 1|Y)}
\]
\[
\hat{\sigma}_2^2 = \frac{\sum_{t=1}^{T} (y_t - x_t' \hat{\beta} - z_t' \hat{c}_2)^2 \Pr(s_t = 2|Y)}{\sum_{t=1}^{T} \Pr(s_t = 2|Y)}
\]
E Filtering and smoothing algorithm

Let
\[ \xi_t = \begin{bmatrix} \mathbb{1}\{s_t = 1\} \\ \mathbb{1}\{s_t = 2\} \end{bmatrix} \]  \hspace{1cm} (A-17)

Let \( \hat{\xi}_{t|\tau} = E(\xi_t|Y_{\tau}) \). Then
\[ \hat{\xi}_{t|\tau} = \begin{bmatrix} Pr(\xi_t = e_1|Y_{\tau}) \\ Pr(\xi_t = e_2|Y_{\tau}) \end{bmatrix} \]  \hspace{1cm} (A-18)

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 \( Y_{t-1} \):
\[ \eta_t = \begin{bmatrix} p(y_t|\theta_1, Y_{t-1}) \\ p(y_t|\theta_2, Y_{t-1}) \end{bmatrix} = \begin{bmatrix} p(y_t|\xi_t = e_1, Y_{t-1}) \\ p(y_t|\xi_t = e_2, Y_{t-1}) \end{bmatrix} \]  \hspace{1cm} (A-19)

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

In my model,
\[ \eta_{t1} = (2\pi)^{-(K+N)/2}|\Psi_{t1}|^{-1/2} \exp \left[ -\frac{1}{2} \begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} - \begin{bmatrix} \mu_1 + \Phi X_t \\ A_1 + BX_t \end{bmatrix} \right]' \Psi_{t1}^{-1} \begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} - \begin{bmatrix} \mu_1 + \Phi X_t \\ A_1 + BX_t \end{bmatrix} \]  \hspace{1cm} (A-20)

\[ \eta_{t2} = (2\pi)^{-(K+N)/2}|\Psi_{t2}|^{-1/2} \exp \left[ -\frac{1}{2} \begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} - \begin{bmatrix} \mu_2 + \Phi X_t \\ A_2 + BX_t \end{bmatrix} \right]' \Psi_{t2}^{-1} \begin{bmatrix} Y_{t1} \\ Y_{t2} \end{bmatrix} - \begin{bmatrix} \mu_2 + \Phi X_t \\ A_2 + BX_t \end{bmatrix} \]  \hspace{1cm} (A-21)

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 \( p(y_t|Y_{t-1}) = \eta_t' \hat{\xi}_{t|t-1} = \mathbb{1}_2(\eta_t \odot \hat{\xi}_{t|t-1}) \) 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
\[ \hat{\xi}_{t|t} = \frac{\eta_t \odot \hat{\xi}_{t|t-1}}{\mathbb{1}_2(\eta_t \odot \hat{\xi}_{t|t-1})} \]  \hspace{1cm} (A-20)

\[ \hat{\xi}_{t+1|t} = P \cdot \hat{\xi}_{t|t} \]  \hspace{1cm} (A-21)
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

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \left( P'(\hat{\xi}_{t+1|T}(\div)\hat{\xi}_{t+1|t}) \right) \quad (A-22)$$

where the sign ($\div$) denotes element-by-element division. The smoothed probabilities $\hat{\xi}_{t|T}$ are found by iteration on equation (A-22) backward for $t = T - 1, T - 2, \cdots, 1$. This iteration is started with $\hat{\xi}_{T|T}$, which is obtained from equation (A-20) for $t = T$. 


F References


Working paper.


Figure 1: Log of the observed three month natural gas futures price
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 the low spread regime (in red). Shaded areas represent the low spread regime.
Figure 3: Log of the observed three month natural gas futures price. Shaded areas represent the low spread regime.
Figure 4: First principal component (in blue) vs. smoothed probability of the low spread regime (in red). Shaded areas represent the low spread regime.
Figure 5: Expected returns conditional on the high spread regime (regime 1, in blue) and conditional on the low spread regime (regime 2, in red)
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>1.4891 (0.0097)</td>
</tr>
<tr>
<td>$A_{(3)}$</td>
<td>1.3803 (0.0058)</td>
<td>1.4743 (0.0058)</td>
</tr>
<tr>
<td>$A_{(4)}$</td>
<td>1.3926 (0.0037)</td>
<td>1.4555 (0.0044)</td>
</tr>
<tr>
<td>$A_{(5)}$</td>
<td>1.4120 (0.0031)</td>
<td>1.4309 (0.0045)</td>
</tr>
<tr>
<td>$A_{(6)}$</td>
<td>1.4328 (0.0030)</td>
<td>1.4064 (0.0050)</td>
</tr>
<tr>
<td>$A_{(7)}$</td>
<td>1.4524 (0.0031)</td>
<td>1.3863 (0.0055)</td>
</tr>
<tr>
<td>$A_{(8)}$</td>
<td>1.4692 (0.0036)</td>
<td>1.3778 (0.0070)</td>
</tr>
<tr>
<td>$A_{(9)}$</td>
<td>1.4786 (0.0049)</td>
<td></td>
</tr>
<tr>
<td>$B_{(3)}$</td>
<td>0.3724 (0.0035)</td>
<td></td>
</tr>
<tr>
<td>$B_{(4)}$</td>
<td>0.3760 (0.0022)</td>
<td></td>
</tr>
<tr>
<td>$B_{(5)}$</td>
<td>0.3789 (0.0018)</td>
<td></td>
</tr>
<tr>
<td>$B_{(6)}$</td>
<td>0.3808 (0.0017)</td>
<td></td>
</tr>
<tr>
<td>$B_{(7)}$</td>
<td>0.3809 (0.0019)</td>
<td></td>
</tr>
<tr>
<td>$B_{(8)}$</td>
<td>0.3797 (0.0021)</td>
<td></td>
</tr>
<tr>
<td>$B_{(9)}$</td>
<td>0.3770 (0.0028)</td>
<td></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>0.0048 (0.0005)</td>
<td></td>
</tr>
<tr>
<td>$\Omega_{(4)}$</td>
<td>0.0018 (0.0002)</td>
<td></td>
</tr>
<tr>
<td>$\Omega_{(5)}$</td>
<td>0.0013 (0.0001)</td>
<td></td>
</tr>
</tbody>
</table>
Table 1: First stage reduced form parameter estimates (continued)

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

*Asymptotic standard errors are in parentheses*
<table>
<thead>
<tr>
<th></th>
<th>Regime 1</th>
<th></th>
<th>Regime 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.0357</td>
<td>(0.0189)</td>
<td>-0.0395</td>
<td>(0.0292)</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>0.9785</td>
<td>(0.0113)</td>
<td>1.4959</td>
<td>(0.0083)</td>
</tr>
<tr>
<td>$A_{(3)}$</td>
<td>1.3743</td>
<td>(0.0046)</td>
<td>1.4959</td>
<td>(0.0083)</td>
</tr>
<tr>
<td>$B_{(3)}$</td>
<td>0.3767</td>
<td>(0.0020)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.0528</td>
<td>(0.0061)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\Omega_{(3)}$</td>
<td>0.0048</td>
<td>(0.0005)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\Omega_{(4)}$</td>
<td>0.0018</td>
<td>(0.0002)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\Omega_{(5)}$</td>
<td>0.0013</td>
<td>(0.0001)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\Omega_{(6)}$</td>
<td>0.0013</td>
<td>(0.0001)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\Omega_{(7)}$</td>
<td>0.0014</td>
<td>(0.0001)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\Omega_{(8)}$</td>
<td>0.0017</td>
<td>(0.0002)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\Omega_{(9)}$</td>
<td>0.0029</td>
<td>(0.0003)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\pi^{P_{11}}$</td>
<td>0.9070</td>
<td>(0.0252)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\pi^{P_{22}}$</td>
<td>0.8057</td>
<td>(0.0479)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\pi^{Q_{11}}$</td>
<td>0.9776</td>
<td>(0.0231)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\pi^{Q_{22}}$</td>
<td>0.9813</td>
<td>(0.0383)</td>
<td>0.586</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\lambda_0$</td>
<td>-0.0009</td>
<td>(0.0189)</td>
<td>0.0257</td>
<td>(0.0295)</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.0233</td>
<td>(0.0114)</td>
<td>0.0257</td>
<td>(0.0295)</td>
</tr>
</tbody>
</table>
Table 3: Wald t-statistics

<table>
<thead>
<tr>
<th>$H_0$</th>
<th>Standard error</th>
<th>Wald t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced-form parameters (first stage)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_{(4),1} - A_{(4),2} = 0$</td>
<td>0.0063</td>
<td>-13.0459</td>
</tr>
<tr>
<td>$A_{(5),1} - A_{(5),2} = 0$</td>
<td>0.0053</td>
<td>-8.1938</td>
</tr>
<tr>
<td>$A_{(6),1} - A_{(6),2} = 0$</td>
<td>0.0056</td>
<td>0.3393</td>
</tr>
<tr>
<td>$A_{(7),1} - A_{(7),2} = 0$</td>
<td>0.0057</td>
<td>8.0412</td>
</tr>
<tr>
<td>$A_{(8),1} - A_{(8),2} = 0$</td>
<td>0.0059</td>
<td>14.0888</td>
</tr>
<tr>
<td>$A_{(9),1} - A_{(9),2} = 0$</td>
<td>0.0079</td>
<td>12.7908</td>
</tr>
<tr>
<td>Reduced-form parameters (second stage)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_1 - \mu_2 = 0$</td>
<td>0.0357</td>
<td>2.1048</td>
</tr>
<tr>
<td>$\sigma_1^2 - \sigma_2^2 = 0$</td>
<td>0.0117</td>
<td>-0.4903</td>
</tr>
<tr>
<td>$A_{(3),1} - A_{(3),2} = 0$</td>
<td>0.0087</td>
<td>-13.9026</td>
</tr>
<tr>
<td>Structural parameters (second stage)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{0,1} - \lambda_{0,2} = 0$</td>
<td>0.0360</td>
<td>-0.7375</td>
</tr>
<tr>
<td>$\pi_{11}^{p} - \pi_{11}^{q} = 0$</td>
<td>0.0351</td>
<td>-2.0095</td>
</tr>
<tr>
<td>$\pi_{22}^{p} - \pi_{22}^{q} = 0$</td>
<td>0.0605</td>
<td>-2.9047</td>
</tr>
<tr>
<td></td>
<td>First stage estimates</td>
<td>Model implied values</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------</td>
<td>----------------------</td>
</tr>
<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>