In the Nick of Time: A Heteroskedastic SVAR Model and Its Application to the Crude Oil Futures Market

Maastricht University

Hang Sun, Jaap W.B. Bos,
Zhuo Li

In the Nick of Time: A Heteroskedastic SVAR Model and Its Application to the Crude Oil Futures Market

RM/17/019

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

In the Nick of Time: A Heteroskedastic SVAR Model and Its Application to the Crude Oil Futures Market

By Hang Sun, Jaap W. B. Bos, Zhuo Li *

Many economic analyses revolve around the identification of shocks. However, this becomes difficult if we do not have enough information, for example because we do not observe the underlying process at a high enough frequency. As a result, if the response of one variable to a shock to another takes place 'in the nick of time' this shock remains unidentified. We introduce a structural vector-autoregression model with Markov-switching heteroskedasticity in the data generating process that allows us to study instantaneous impulse-response relationships with the proper selection of a supporting 'catalyst', which can be easier to find than an instrumental variable.

JEL: G13; C32; Q02

Keywords: SVAR, Identification, Markov-switching, Commodity prices, Index Trading

I. Introduction

In empirical analyses, we often observe data at a frequency that is not high enough to capture the speed of the data generating processes behind the causal effects we are interested in. In this paper, we look into this problem by proposing a novel identification

* Sun: Maastricht University School of Business and Economics, P.O. Box 616, 6200 MD, Maastricht, The Netherlands, h.sun@maastrichtuniversity.nl. Bos: Maastricht University School of Business and Economics, P.O. Box 616, 6200 MD, Maastricht, The Netherlands, j.bos@maastrichtuniversity.nl. Li: Corresponding author, Wuhan University Economics and Management School, Luojiashan Wuchang, Wuhan, China, lizhuo@whu.edu.cn, tel. +86-27-68753023, fax +86-27-68754916. Hang Sun gratefully acknowledges support from the China Scholarship Council. The usual disclaimer applies.

strategy based on a structural vector-autoregression (SVAR) model.

We subsequently demonstrate how our method can shed light on the relationship between commodity futures prices and commodity index trader flows. There are several reasons why this relationship is hotly debated and appropriate for our purpose. One reason is because commodity index traders (hereafter CITs) are seen by some as 'speculators' - a group of passive investors who are believed to have large-scale and overwhelmingly long-only positions in the markets (Masters, 2009). Another reason is because this relationship is typically studied using weekly data, since these are readily available, even though the impact of trades on returns is generally assumed to materialize almost instantaneously.[1]

One way of examining the impact of CITs on commodity prices is through an indirect experiment design (see, e.g., Tang and Xiong, 2012). Another way is by examining the correlation between CIT positions and commodity prices directly (e.g., Irwin and Sanders, 2012, Singleton, 2014, Hamilton and Wu, 2015). However, there is little agreement among authors. Singleton (2014) finds the 13-week changes of CIT positions can positively predict excess returns of crude oil futures. Irwin and Sanders (2012) and Hamilton and Wu (2015) find no evidence of a positive, significant impact of index trading on commodity futures prices.

One possible explanation for these inconsistent findings is the mismatch between the frequency of data employed and the speed of the response being studied. Since the available index trading data are reported weekly, all of the existing studies that attempt to examine the influence of index trading directly can only observe the delayed interaction between index trading and futures markets after one week. Indeed, if we regress returns directly on the 'instantaneous' changes of index trader positions, there will be an obvious reverse causality problem: index traders in commodity markets will adjust their positions over time depending on market conditions, even if they maintain a fixed portfolio based on the commodity index they follow. Indeed, using a proprietary

[1]Many studies also focus on the interrelationship between the real demand of crude oil and oil prices with different identification strategies, e.g. Kilian and Murphy (2013), Juvenal and Petrella (2014), and Anzuini et al. (2014).

database of the U.S. Commodity Futures Trading Commission (CFTC), Cheng et al. (2014) find that CIT positions are significantly affected by the market stress indicated by the VIX.

If we had daily or even intra-day data, the relationship between prices and trader flows would be easily observed. Using a daily dataset of commodity-linked notes, Henderson et al. (2014) show these notes reflect the participation of financial investors in commodity markets and contain information about future returns of referenced futures contracts; moreover, the information dissolves largely with one week. But can we still capture the effects found by Henderson et al. (2014) if we use our method and apply it to weekly data?

Indeed, in line with Singleton (2014) and Henderson et al. (2014), we also find positive instantaneous responses of crude oil excess returns to CIT inflow shocks. In addition, our approach allows us to demonstrate that the response of crude oil excess returns to CIT inflow shocks decays at an exceptionally low speed. The response remains significant in several following weeks, suggesting that the information contained in the position changes of CITs is absorbed very slowly. Hence, we complement Singleton (2014), who studies the effects of CIT inflow after a week, as we have reasons to believe that his findings only show part of a larger impact as revealed by this paper.

The remainder of this paper continues as follows. In Section III, we introduce a method for local identification using heteroskedasticity, before arriving at a more general identification strategy. Section IV contains our baseline results as well as a description of the data employed in our empirical analysis. We describe the mechanism behind our impulse response analyses and introduce some further robustness tests. In Section V, we enrich our model with additional controls based on the existing literature. We conclude in Section VI.

II. Identification

In principle, instantaneous impulse responses can be described by an SVAR model. Compared with reduced-form vector-autoregression (VAR) models, in an SVAR model

the residual covariance already has more structure since some variables are allowed to respond to shocks instantly. However, since we often believe that data do not contain enough information for the full identification of the structure of shocks, we typically impose some additional constraints on the structure of the VAR model.[2]

The most common form of constraints are triangular constraints. They require the variables in the SVAR to be ordered as a chain where shocks to any variable can only affect the variables ranked *after* the shocked variable directly. This is often a very strong requirement, since there does not exist such an order for many economic systems. Other, similar identification strategies face similar problems in deciding which variables, either in the short or long run, are immune to which shocks.

A few different identification strategies have been proposed. Among them, a popular strategy tries to identify an SVAR with sign restrictions, which is first proposed by Faust (2015). With sign restrictions, we constrain the direction of the responses to shocks, but not their timing. For example, a negative shock to the money supply will lead to an increase in the federal funds rate. In practice, however, this approach still suffers from two issues. First, for many economic systems, we do not have enough prior knowledge to establish convincing sign restrictions. For instance, the relationship between CIT flows and crude oil futures returns is still debated and it is therefore not possible to assert whether a certain response should be positive or negative. Second, unlike traditional short or long-run restrictions, the feasible shock structures identified by sign restrictions are not unique. As a result, in order to arrive at a single (set of) possible impulse response functions, we need to introduce additional restrictions, which are often difficult to interpret based on economic theory. Stock and Watson (2008) and following literature use proxy variables to identify an SVAR, but usually we still have to impose additional restrictions on the structures of reduced-form residuals (e.g. Mertens and Ravn, 2013). A machine-learning based algorithm called 'PC algorithm' is also used to identify an SVAR (see Spirtes et al., 2000), but it requires that there does not exist a causal loop in the system, which our case of CIT flows and oil futures returns

[2]For a review of common SVAR identification strategies, see Kilian (2013).

apparently does not satisfy, since the causality goes in both directions. There are also some other identification stategies that are not useful in a data-scarse environment, for example, Beetsma et al. (2009) rely on higher-frenquency data, which are often not available like the issue in this paper.

In this paper, we try to fully identify the structure of our SVAR model with a novel strategy based on heteroskedasticity. We follow Lanne et al. (2010) and Lütkepohl and Netšunajev (2014) and identify the SVAR model by assuming the data are generated by a heteroskedastic process which can be captured by Markov-switching volatility states.

The idea to identify an SVAR through heteroskedasticity is pioneered by Rigobon (2003). He assumes that if we believe the underlying economic mechanism behind an SVAR model is constant over the observed period, then we can try to split the observations into sub-periods, and the shape of impulse-response functions (IRFs) should remain the same across these sub-periods. This testable assumption provides us with new information about the structure of an SVAR model.

Inspired by Rigobon (2003), Lanne and Lütkepohl (2008) notice that heteroskedasticity can be utilized in many general cases for SVAR specification tests. While in Rigobon (2003) the observed series are simply split at a pre-determined breakpoint that is often not so easy to select, Lanne et al. (2010) show they can also be split according to a Markov Switching VAR (MSVAR) model. In this way, we can obtain a unique pattern of impulse-response functions, and do not need any further identification method like sign restrictions. Lütkepohl and Netšunajev (2014) built on their work to establish a series of standard test procedures.

Although Lanne and Lütkepohl (2008), Lanne et al. (2010), and Lütkepohl and Netšunajev (2014) have established SVAR model specification tests, drawing inference from SVAR models often remains problematic. As an example, consider an SVAR system with two variables A and B. For this system, we can draw four impulse-response plots showing the response of A and B to structural shocks on A and B. But we still do not know which plot represents the response of A to shocks on A, and which represents its response to shocks on B - a problem we will describe in more detail in the following

section. Only in a few cases can this problem be solved by prior economic knowledge, e.g., while demand shocks are transient, supply shocks have long-term effects.

This paper solves this shock-labelling issue by making heteroskedastic models part of a general-purpose SVAR identification strategy. In the context of the two-variable system, if we introduce a new variable that interacts (simultaneously) with one existing variable whilst not interacting (simultaneously) with the other, shocks can be labelled at once. Hereinafter, we call this newly introduced variable a 'catalyst.' Essentially, the catalyst works because its relationship with the two variables we care about is asymmetric, and therefore informs us which variable is which. While an instrumental variable has to be uncorrelated with the error term, about which we have little information to justify the choice of the instrumental variable, an catalyst in our model only needs to be uncorrelated with one of the variable we are interested in.[3]

In this paper, we use crude oil inventory levels as a catalyst. Obviously, these interact instantly with oil prices. But since index traders are passive traders who keep proportional positions in a blanket of commodities, it is unlikely that they would adjust their positions according to real time inventory data, if they ever had them.[4]

III. Methodology

We start by closely tracing the path laid out by Lanne et al. (2010) and Lanne and Lütkepohl (2008). We argue that their approach enables us to 'locally' identify the impact of shocks. Subsequently we move towards a more general identification procedure that allows us to fully identify the impact of the same shocks.

[3]Our method can be easily generalized to the cases with more variables: for a system with K variables, $K-1$ additional catalysts enable us to tell the shocks apart.

[4]Meanwhile, except for the Commodity Futures Trading Commission's weekly report, there is no other source for the inventory holders to learn index traders' positions, let alone react to them. As we shall show later, using inventory levels as the catalyst is also statistically justifiable, since they pass the model specification tests proposed by Lanne et al. (2010) and Lütkepohl and Netšunajev (2014).

A. Local identification through heteroskedasticity

Consider a typical stable k-dimensional VAR(p) model:

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + v_t, , \tag{1}$$

where $y_t = (y_{1t}, y_{2t}, \ldots, y_{kt})'$ are the observations of k endogenous variables at time t, c is the intercept, $\{A_i\}_{i=1}^{p}$ are the coefficient matrices, and v_t is a vector white noise error term that has zero mean and a covariance matrix Σ. v_t represents a mixture of different shocks and can thus be decomposed as $v_t \equiv B\epsilon_t$, where ϵ_t has a mean of zero and a unit covariance matrix I_k. Now ϵ_t represents the orthogonalized shocks that are mutually independent. Since the matrix B contains information about how the independent shocks are reconstructed in v_t, Σ can also be decomposed as the inner product of B, $\Sigma = BB'$. Hereafter we call B the 'structural matrix.'

Rewriting the error term in the reduced-form VAR model with orthogonal errors, we obtain the SVAR(p) model:

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + B\epsilon_t. \tag{2}$$

Compared with the reduced-form VAR model, the structural matrix B is now the extra unknown parameter for an SVAR model. Given the constraint $\Sigma = BB'$, the estimation of B is equivalent to finding a proper decomposition of the reduced-form covariance matrix Σ. However, there are infinitely many ways to decompose Σ.[5] So we need additional information to estimate the true B that generates the data.

For this additional information, we need to understand the economic process we are investigating. For example, if we are certain that some variables are immune to orthogonal instantaneous shocks, we can implement zero restrictions. An effective set of zero restrictions must have at least $k(k-1)$ restrictions, which correspond to at least

[5]Obviously, if B is a decomposition such that $\Sigma = BB'$, then BQ is also a decomposition, where Q is an orthogonal matrix, such that $\Sigma = BQQ'B'$.

$k(k-1)$ elements in B to be set to zero a priori. Sometimes, the hierarchy among shocks is fairly obvious. For example, at times we may expect a certain shock to have an impact on the entire system instantaneously, whereas another has an impact on the entire system but for one variable instantaneously, and so forth, until the final shock that can only have an impact on a single variable instantaneously. This then corresponds to a triangular matrix B that satisfies the conditions for the zero restrictions. However, in practice ranking all the shocks in a strict order is often not possible. For example, we simply do not know the order of shocks between CIT positions and oil futures prices, which is indeed the very reason we are interested in examining their relationship.[6]

In this paper we try to identify an SVAR model with much weaker assumptions based on heteroskedasticity, which proves to contain important additional information (Rigobon, 2003). Our method allows a pair of variables in an SVAR system to interact freely and does not require any sign restrictions, thus enabling us to observe the intricate short-term dynamics among variables in the system that are previously considered unobservable, including the case in this paper, i.e., the instantaneous relationship between CIT flows and crude oil prices.

Instead of the previously mentioned rather strong assumptions, for our approach we have to make a relatively mild assumption: that the data are generated by a heteroskedastic process. Hence, although the variables we are interested in influence one another in the same way during the entire sample period, the strength of the shocks that reflect their dynamics varies.

We start with the simplest situation, in which volatility suddenly jumps from one level to the other at a single breakpoint time T, which we assume - for now - is known a priori. We then get a heteroskedastic reduced-form VAR model with a jump in volatility

[6]Another popular strand of the literature tries to identify a VAR model with sign restrictions. This requires weaker assumptions. However, in our case we still do not have information to support sign restrictions, which also introduce other problems (e.g., the loss of the uniqueness of the solution; see Fry and Pagan, 2011).

similar to Lanne and Lütkepohl (2008):

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + u_t, , \tag{3}$$

where $\mathsf{E}[u_t] = 0$, $\mathsf{E}[u_t u_t'] = \Sigma_1 \cdot \mathbf{1}_{\{t<T\}} + \Sigma_2 \cdot \mathbf{1}_{\{t\geq T\}}$, and the remainder stays the same as in equation (1). Since the relationship between the variables is assumed to be stable, the coefficients of the model are constant. The identification of orthogonal shocks is still equivalent to finding the proper decomposition of the reduced-form residual covariance matrices. But the breakpoint T divides the model into two states, $t < T$ and $t \geq T$, with different volatilities. We call them State 1 and State 2, respectively.

Lanne and Lütkepohl (2008) shows that the assumption that the relationship among the variables we are interested in remains stable across different volatility states leads to the local identification of the model in equation (1). We call it a local identification because the possible decompositions of the reduced-form residual covariance matrices are still not unique.

In line with the standard SVAR model, the IRFs have the same shapes across different states. Therefore, if B is a proper decomposition of Σ_1 for State 1, then the decomposition of Σ_2 must be in the form of $B\Lambda^{1/2}$, where Λ is a nonnegative-definite diagonal matrix, in order to keep the shapes of IRFs unchanged across states. Indeed, for a pair of symmetric matrices Σ_1 and Σ_2, if matrices B and $\tilde{B}$ both satisfy $\Sigma_1 = BB' = \tilde{B}\tilde{B}'$ and $\Sigma_2 = B\Lambda B' = \tilde{B}\Lambda\tilde{B}'$, where Λ is a nonnegative-definite diagonal matrix whose diagonal elements are pairwise unequal, then there must exist one and only one orthogonal matrix Q so that $\tilde{B} = BQ$ and $Q = TR$, where T is a column-switching elementary matrix and R is a diagonal matrix whose diagonal elements are either 1 or -1.[7] Interestingly, this decomposition can also be generalized to more than two states, so in fact we can have as many breakpoints as we want. If there are S states, and the covariance matrix of the i-th state Σ_i can be decomposed as $\Sigma_i = B\Lambda_i B'$, then the

[7]Its proof can be found in many textbooks. Lanne and Lütkepohl (2008) also provide a proof in their appendix.

sufficient condition of the local uniqueness of B up to sign and permuation of columns is the existence of $j \in \{1, \ldots, S\}$ such that the diagonal elements of Λ_j are not pairwise equal (Lanne et al., 2010).

So far, we have closely followed Lanne and Lütkepohl (2008) and Lanne et al. (2010), who mainly develop their approach for local identification as benchmarks for tests of established identification strategies. In order to fully identify the model, we still need to solve four outstanding problems. First, the matrix R allows us to arbitrarily switch the signs of each column in B. Second, the local uniqueness of B only exists for non-trivial cases where the diagonal elements of Λ are all unequal in at least one state. Third, the matrix T allows us to arbitrarily swap different columns of matrix B. Fourth, the model specification outlined so far (Rigobon, 2003) requires a choice of breakpoint *a priori*, the determination of which is often problematic itself. We solve each of these four problems in the next subsection.

B. Towards full identification

As it turns out, providing a solution to the first two problems discussed above is rather trivial, as we can directly follow earlier work. The second set of problems, however, requires somewhat more work.

The first problem, of arbitrarily switching signs, is indeed trivial. Technically, any locally identified orthogonal shock can be either positive or negative, since the signs of the columns of B are arbitrary. But we can simply assume that the instantaneous response of a variable to a shock originating from itself is always positive, following Lanne and Lütkepohl (2008) and Lanne et al. (2010). For example, CIT positions always increase instantaneously when a CIT inflow shock takes place. As a result, we require the diagonal parameters of B to be positive, thus effectively determining the signs of the columns of B.

For the second problem, the semi-identification, we can follow in the footsteps of Lanne et al. (2010) and Lütkepohl and Netšunajev (2014). They have proposed a series of Wald tests for testing the equality among the diagonal elements of Λ.

The third problem, of arbitrarily swapping columns of B, is however less trivial. Since we have not imposed any further restrictions on B, we do not know to which shock any particular IRF belongs, even if we have obtained all IRFs. As a result, we have at most quasi-identification: sometimes orthogonal shocks can be labelled based on prior knowledge, but more often this is simply not possible.

We provide a solution to this problem by making a minimal concession. Note that for a 3×3 sized matrix B, if two elements in B that are neither in the same row nor the same column are known, the whole matrix is almost surely identified. Based on this, we introduce a new 'catalyst' variable that interacts with one of the existing variables but does not interact with the other variable(s). Assume the system in equation (3) is composed of $y_t = (y_{1t}, y_{2t}, y_{3t})'$ and y_{3t} only interacts simultaneously with y_{1t} and not with y_{2t}. Then we know the structural matrix B must have the form

$$B = \begin{pmatrix} * & * & * \\ * & * & 0 \\ * & 0 & * \end{pmatrix},$$

where the asterisks represent free parameters to be estimated. We can see that the columns of B are now identified by the two zeros.

For a three-variable system, a set of traditional triangular restrictions requires three elements in B to be zero, while our method requires two. One less restriction is far more than a marginal improvement. First, even for a three-variable case, losing one restriction can make the SVAR model applicable to far more economic systems—while triangular restrictions correspond to a strictly recursive causal chain, which is uncommon in reality, the difficulty of building a system that meets our requirements is comparable to finding an instrumental variable, which is related to one variable and unrelated to the other. Second, when the dimensionality of the system increases, we only require $k - 1$ restrictions, in contrast with the previous $k(k-1)/2$ restrictions, hence our advantage over the existing approach grows at a quadratic rate.

Our approach also compares favorably to a traditional instrumental variables ap-

proach. An instrumental variable is required to be uncorrelated with the error term, i.e., all potential omitted variables. Since we can never be completely sure what variables are omitted, finding a good instrumental variable is far from trivial. The requirements for our 'catalyst', however, are much weaker, as the latter is only required to be uncorrelated with one of the two variables in our analysis.

As is the case with an instrumental variable, the fact that we choose a catalyst based on our prior knowledge puts us at the risk of having a misspecified model. However, a second advantage of the identification strategy we follow here is that it allows us to test for possible misspecification of the structural matrix in an intuitive and straightforward manner. Lanne et al. (2010) and Lütkepohl and Netšunajev (2014) have already shown that any restriction on the structural matrix becomes over-binding if the true model is in reality unrestricted. This means we can build a simple likelihood-ratio test: if the likelihood of the restricted model is significantly lower than the unrestricted model, we know that our restrictions are very likely to be misspecified.

The fourth issue is the choice of the breakpoint. Lanne et al. (2010) show that to circumvent this issue, we can introduce a Markovian Regime-Switching VAR (MSVAR) model with two or more different volatility states. We only introduce the simplest case with two states here, since the approach easily generalizes to more states. Let a Markov chain $\{s_t\}_{t=0,\pm1,\pm2,\dots}, s_t \in \{0,1\}$ denote the underlying state of the system at time t, with a transition probability matrix

$$Q = \begin{pmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{pmatrix},$$

where $q_{ij} \equiv \Pr[s_t = j | s_{t-1} = i], i, j \in 0, 1$. Then the model can be specified as

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + u_t, \tag{4}$$

where $\mathsf{E}[u_t] = 0$, and $\mathsf{E}[u_t u_t' | s_t] = \Sigma_{s_t}$. At each time t, the system can be in either state with certain probabilities and the probability distribution can be estimated from the

data. There is no need to identify the states ex ante any more. The identification of this model is equivalent to finding the proper decomposition of Σ_0 and Σ_1 and thus can be achieved in the same way as for equation (3).

These properties make the MSVAR model particularly suited to represent financial market variables, since financial markets are often described with ARCH-type models and the MSVAR model with different volatility states shares similar features. The estimation of the MSVAR parameters and the identification of the structure can be achieved simultaneously through maximum likelihood estimation (Lanne et al., 2010).[8] A brief Monte Carlo analysis of the estimators can be found in the Appendix. Since we are mainly interested in the case where the true data generating process has a higher-frequency structure than the observations, we also construct an example to examine the performance of our model in this situation.

IV. Baseline Model

To demonstrate our approach, we now turn to the relationship between CIT positions and commodity futures excess returns. We start by introducing our data, before we turn to our model setup and results.

A. Data

The empirical analysis in this paper will begin with the baseline model containing the three most important variables: West Texas Intermediate (WTI) futures excess returns, CIT position changes, and crude oil inventory shifts. For these variables, we can obtain weekly observations for the period starting January 17th, 2006 and ending December 31st, 2012. We use the weekly changes of US crude oil inventory as the catalyst variable. First, inventory is a critical factor in commodity pricing. Inventory adjustments can directly affect the balance of demand and supply and thereby the spot price. Moreover, inventory is closely correlated with the convenience yield (Gorton

[8]The computer program used in this paper builds on the MATLAB code published together with Lütkepohl and Netšunajev (2014) and is available upon request.

et al., 2008), which is an important pricing factor of futures. Second, as index traders are believed to be passive traders tracking commodity indices, they are not likely adjust their positions according to inventory information promptly. The other way round, the only data source that an ordinary inventory holder can get access to is the weekly report published by the Commodity Futures Trading Commission (CFTC), which makes it reasonable to assume that it takes at least one week for the inventory holders to react towards CIT flows, if they do. Since a catalyst is expected to interact with only one of the other two variables we are concerned, inventory shifts constitute an ideal catalyst.

US crude oil inventory data and WTI futures price data can be easily found at the US Energy Information Administration (EIA) website and Datastream. Futures excess returns are calculated in the same way as Singleton (2014), and we use the log excess returns of the front contracts in the baseline model. Correspondingly, the changes of inventories are also represented by the weekly log difference.

The remaining variable, CIT positions, is not readily available. The 'Commodity Index Trader Supplement' (CITS) report and the 'Index Investment Data' (IID) report released by the CFTC are the sole reliable data sources about the commitment of CITs, but neither of them can be used as is: although the former is announced weekly, it only contains information about agricultural commodity markets; the latter does contain information for the crude oil market, but it is released monthly.[9] For these reasons CIT positions in the WTI market can be only estimated based on the available data. The standard practice in the existing literature (Singleton, 2014, Hamilton and Wu, 2015) is based on a mapping algorithm proposed by Masters (2009). Masters (2009) considers two major commodity indices, S&P GSCI and DJ-UBSCI and finds that CIT positions in any non-agricultural commodity can be deduced from the information released in the CITS reports and the weights of the two commodity indices, after noticing each of the two indices contains a characteristic agricultural contract that is not shared by the other. Hamilton and Wu (2015) further notice that the reliance of Masters (2009) on characteristic contracts is not necessary, and the position data in every agricultural

[9]Before June, 2010, the IID report was even released on a seasonal base.

contract made public by the CITS reports can be utilized for the estimation of CITs' engagement in a non-agricultural contract. Hereafter we refer to the estimation through the algorithm of Hamilton and Wu (2015) as the Masters-Hamilton-Wu (MHW) model.

The MHW model is not without critics, though. Irwin and Sanders (2012) argue that there exists a considerable discrepancy between the MHW estimated values and the official monthly values in the IID reports. We confirm their findings for the WTI market, as is shown by the spread between the dashed line and the double dashed line in Figure 1. Nevertheless, despite the spread between the level values, we find the trends of the two lines are to a large extent consistent, which implies that apart from the spread, the MHW estimation is close to the true values. We therefore try to improve the MHW estimation by combining the information in both the CITS and the IID reports. We employ the interpolation algorithm devised by Chow and Lin (1971) and rectify the MHW estimation with the IID reported values while keeping the general trend. The result is shown by the solid line in Figure 1. We then use the log growth of the rectified estimation of CIT positions in the baseline model.

[Insert Figure 1 here]

For simplicity, we abbreviate the WTI front futures contract excess returns, inventory growth, and CIT position growth as ER1, INVENTORY, and CITPOS, respectively.

B. Model Setup

To further specify our MSVAR model, we need to fix two parameters: the lag order and the number of states. Taking into account that the market variables we are studying are expected to respond to new information very quickly, we use one lag in the baseline model.[10]

[Insert Table 2 here]

Next, we have to determine the number of states. As can be seen in Table 2, both AIC

[10]Indeed, increasing the lag order results in substantial increases in AIC values.

and Schwartz Criterion (SC) values point to three states. The existence of three states becomes more intuitive once we have a look at the estimated probability distribution for different states, as shown in Figures 2a–2c. With three different volatility states (Figure 2a), we obtain a clear division between the three sample periods. Judging from the residual covariances reported in Table 1, the volatilities are the lowest in State 1 and the highest in State 3. This pattern is consistent with our impression about the recent crude oil market, as the most volatile State 3 roughly corresponds to the spectacular bust after mid-2008, and the market also saw large volatilities in the first half year of 2008 and during the rally after 2010, which is included in the distribution of State 2. For comparison, consider the results with two states, shown in Figure 2c. The state probability distribution lacks an obvious pattern, and neither of the two states remains stable for a long time.[11]

[Insert Figure 2 here]

[Insert Table 1 here]

Since the identification of the baseline model relies on the assumption that Inventory and CITPos do not respond to the shocks to each other instantly, we test whether the model is misspecified. The unrestricted model has 33 free parameters and its log likelihood is 3336.09. When we restrict the instantaneous responses between Inventory and CITPos to zero, the number of free parameters drops to 31 and the log likelihood drops to 3335.38. With a p-value of 0.7011, the likelihood ratio test cannot reject the validity of our restrictions in the baseline model.

[Insert Table 3 here]

[Insert Table 4 here]

Before we start analyzing the baseline model, we also have to test whether the diagonal

[11] In fact, the state distribution shown in Figure 2c appears only to be generated by splitting State 2 in the 3-state case into two states.

elements in the estimated Λ's are different. The estimated values of the Λ's of both the unrestricted model and the model with restrictions on the instantaneous responses between INVENTORY and CITPOS are reported in Table 3, where λ_{ij} denotes the j-th diagonal element in the matrix Λ_i of State i. We see that the estimations of the λ's do not change much after we impose the restrictions. For each state, no pair of λ's in the restricted model appears to be equal. The Wald tests in Table 4 provide the details: the p-value of the equality test for the first diagonal elements of Λ_2 at State 2 slightly exceeds 0.05, while the tests for the other two pairs at this state are highly significant. However, the statistics for the inequality of the λ's at State 3 cannot provide us with strong enough evidence against the null hypothesis, as λ_{31} is not significantly different from λ_{33}, mainly driven by the large standard deviation of the estimations. Nonetheless, since a Λ with all different diagonal elements is only required in one among all the states for the model to be identified, we can proceed with our restricted baseline model based on the inequalities of the λ's for State 2.

C. A shock to CIT inflows and other impulse responses

Once we obtain the proper structural matrix through the decomposition of the covariance matrices at different states, we can identify three orthogonal shocks from the non-orthogonal reduced-form residuals. The impulse-response functions for our restricted baseline model are shown in Figure 3.

We observe the instantaneous responses and the responses in the following nine weeks. Each column in Figure 3 represents a specific orthogonal shock as identified by the estimation of the structural matrix, namely the shock of WTI futures excess returns, inventory shifts, and CIT inflows. Each row represents the instantaneous and continued responses of a certain variable, ER1, INVENTORY, and CITPOS ordered from top to bottom. The solid line denotes the true IRF and the dashed lines mark off the 68% confidence intervals. Comparing the magnitude of each of the responses is not straightforward, as these very across states.

We first consider the three IRFs on the diagonal of Figure 3. Recall that we assume

that all orthogonal shocks are positive shocks, such that all the variables respond positively to the shocks to themselves at the beginning. Therefore the three IRFs on the diagonal all start with positive values. The impact of an excess return shock is transient and only significant within the same week, which is typical for a financial asset. In contrast, the impact of both the inventory and the CIT inflow shock remains significant for several weeks: interestingly, a positive shock in the inflow of index investors can persist for weeks. Since CITs are mainly institutional investors, this is probably a result of the fact that the funding of institutional investors is mostly affected by slow-moving factors such as the business cycle. The relatively slow decay of the inventory shock is less surprising, and likely related to its adjustment costs (Jin, 2013).

The assumption that inventory and CIT do not react instantaneously to each other is the key to identification, and means we can impose zero restrictions on the respective positions in the structural matrix. It enables us to impose the necessary zero restrictions on two positions in the structural matrix that stand for the instantaneous interaction between Inventory and CITPos. This assumption is reflected in the responses of Inventory to CIT inflow shocks and CITPos to inventory shocks, which are both zero in the starting week. Although by no means the consequence of the (initial) zero restriction, the response of Inventory to CIT inflow shocks remains insignificant and close to zero for the whole observed period, indicating that inventory holders ignore the CITs in the futures market almost completely. This appears to reflect the insensitivity of participants in the physical market to financial investors' behavior. Finally, the IRF of CITPos to inventory shocks is not significant either, which is in line with the view of CITs as passive investors.

Our SVAR model can also reveal some interesting aspects about the relationship between inventory and crude oil futures excess returns. While it is well known that the inventory level of commodities is an important determinant of convenience yields (Pindyck, 1994), the precise mechanism is typically rather opaque. Gorton et al. (2008) predict a negative relationship between futures excess returns and commodity inventories. In line with their theory, the subplot in the first row, the second column in

Figure 3 shows a negative response of ER1 to positive inventory shocks, which is only significant in the first week. In what is perhaps the consequence of the release of inventories to stabilize prices, responses of INVENTORY to excess return shocks turn out to be negative.

[Insert Figure 3 here]

Now that we have discussed how each of our variables of interest behaves when subject to a shock, it is time to focus on our main point: what happens to crude oil futures excess returns after a shock to CIT inflows? In answering this question, we follow in the footsteps of, e.g., Singleton (2014), Hamilton and Wu (2015) and Irwin and Sanders (2012). Although both focus on long-term CIT flows, they find contradicting evidence as to how these flows affect crude oil: Irwin and Sanders (2012) and Hamilton and Wu (2015) find no effect, whereas Singleton (2014) does.

Our results are in line with Singleton (2014), and rather strong: as Figure 3 shows, the response of ER1 - excess returns - to CIT inflow shocks is positive and persists for many weeks following the initial shock. Apparently, the information contained in CIT flows is absorbed by the market very slowly, even compared with the findings of Henderson et al. (2014). Summing up, our analysis so far shows that there is indeed an instantaneous response of ER1 to CIT inflow shocks, and capturing it does not take away from a longer-run impact.

D. Additional Tests

Recall that we impose two restrictions on the structural matrix in order to fully identify the baseline model. Although we cannot reject these restrictions statistically, one may still wonder how these restrictions affect the results. We can now compare the IRFs from the restricted baseline model with those based on the unrestricted model. The unrestricted IRFs are shown in Figure 4.

Note that the order of the λ's for the restricted and the unrestricted models is the same. Hence we can label the three columns of Figure 4 in the identical order as for the

restricted case. Comparing Figure 4 with Figure 3, we find most of the IRFs remain similar. Of course for the response of ER1 to inventory shocks, we have effectively introduced a new channel in our fully identified model, via the CIT positions. Indeed, this is where the IRFs from both models are markedly different. An inventory shock now also has a positive instantaneous impact on CIT positions, reflecting investors' reaction to physical traders' activities. Although the instantaneous impact of a CIT inflow shock on INVENTORY is only marginally significantly positive at the 68% level, the overall picture from both models is therefore different.

[Insert Figure 4 here]

Given that the effects of CIT flows can persist for a long time, a logical next step is to investigate how our model fares once we extend our horizon. To do so, we increase the numbers of lags in the baseline model to 13 and observe the IRFs for 52 weeks. The results are displayed in Figure 5.

[Insert Figure 5 here]

Once we review the IRFs over very long periods, it obviously becomes less easy to find a consistent sign, since we can be less sure to assess the impact of a single shock. Indeed, the response of ER1 to inventory shocks is no longer purely positive (or negative, for that matter) over such a long horizon. Nevertheless, the response of INVENTORY to shocks in the futures market is still largely negative, even we extend the lag number to 13. In addition, the impact of CIT inflow shocks on ER1 is still persistently positive, and the impact of excess return shocks on CIT Position changes also remains positive, though the former fades much more quickly than the latter.

V. Models with Extra Controls

It is possible that the relationships identified in the baseline model vary in different financial environments, especially since our sample contains the 2008 crisis. We therefore estimate some models in which the VIX and the US term spread are included as

extra control variables in our model.

We choose these two variables because they are likely to have a direct influence on both crude oil futures excess returns and CIT positions. The yield curve of the US treasury bond is widely used as a leading indicator of economic growth (Dotsey, 1998), which may affect both oil prices and CITs funding. The VIX is often used as an indicator of investors' fear (e.g. Whaley, 2000), and has proven to helpful in predicting returns of many assets, including crude oil futures (see Chevallier and Benoît, 2013). Finally, CITs are also known to actively adjust their positions according to the risks represented by VIX (Cheng et al., 2014). We use the log difference of the VIX (VIX) and the difference of the 10Y-3M term spread of US treasury bonds (TermSprd).

In the models in this section, we therefore regard VIX and TermSpr as exogenous variables. Let y_t be the endogenous vector $(\text{ER1}'_t, \text{Inventory}'_t, \text{CITPos}'_t)'$, and x_t be the exogneous vector VIX_t, TermSprd_t, or $(\text{VIX}'_t, \text{TermSprd}'_t)'$, then the reduced-form model is expressed as:

$$y_t = c + \sum_{i=1}^{p} A_i y_{t-i} + \sum_{i=1}^{p} \Phi_i x_{t-i} + u_t, \tag{5}$$

where u_t still captures the non-orthogonal residuals with Markov-Switching volatility regimes. This reduced-form model is identified in the same way as the baseline model by restricting the instantaneous mutual responses between inventory and CIT positions. To be consistent with the previous results, we again employ three volatility states and one lag. In order to save space, we only report the IRFs of the model with both the VIX and TermSprd included as controls in Figure 6.[12] Despite the difference in scales, there is no difference in the significance and signs between this figure and Figure 3: the impulse-responses with extra controls change little compared with the baseline model, again proving that our baseline model is rather robust.

[Insert Figure 6 here]

[12]The full results are available upon request.

[**Insert Table 5 here**]

Table 5 shows the coefficients of the exogenous variables.[13] The TermSprd is indeed a good predictor for our endogenous variables. Since a higher term spread signals higher growth, its positive predictive power for the crude oil futures excess returns is to be expected. The VIX negatively predicts CITPos, in line with the arguments in Cheng et al. (2014) that index traders follow the convective flow and chase the less risky assets. Although the coefficients for the VIX in the prediction of ER1 are positive, contradicting findings in the literature (e.g. Saria et al., 2011), they are not very significant.[14]

Summing up, our results are robust to both changes in the lag structure and the inclusion of additional controls.

VI. Conclusion

In this paper we design a novel identification strategy for structural vector autoregression models, which enables us to capture (some of) the short-run dynamics between variables that we observe at a low(er) frequency. We show how our method can shed light on the hotly-debated issue on the relationship between commodity prices and commodity index trader flows with weekly data.

Our SVAR identification strategy relies on an underlying Markov-Switching VAR model with two or more volatility states and the assumption that the impulse-response functions have the same shape across different states. If we are interested in the relationship between two variables, the mutual instantaneous impulse-responses between them can be identified with the inclusion of a single additional 'catalyst' variable, which only interacts instantaneously with one of two variables and not with the other. Likewise, we can identify a system with more variables, with the inclusion of other catalyst

[13] We also show Wald statistics, which generally do not have very significant p-values except the prediction of ER1 with TermSprd. However, due to the highly nonlinear nature of the MSVAR model, Wald tests are in fact not that reliable.

[14] Further evidence in this direction can be gathered from the insignificant Wald statistics for the relationship between the VIX and Inventory.

variables.

Compared with traditional SVAR identification strategies such as triangular restrictions or sign restrictions, the restriction conditions of our strategy are much easier to satisfy, and the impulse-responses resulting from estimating the model are uniquely identified. In fact, the impulse responses from our approach can be regarded as the cumulative responses generated by the real data-generating process at a higher frequency. Hence, our methodology can help solve many identification problems for situations where data are available at a relatively low frequency, for example for financial markets, but also for many macroeconomic analyses.

As an illustration, we demonstrate how our methodology can help us understand the relationship between crude oil futures prices and commodity index trader flows. Using the estimated index trader positions and other weekly data, we find that an index trader inflow shock can cause a positive response of the WTI futures excess returns instantly, and the response remains significant in several following weeks. Commodity index traders themselves also positively respond to shocks of crude oil futures excess returns instantly, though the responses will not last for long and are only significant in the immediate week. These findings are robust even after we add some extra control variables that are likely to affect oil prices and index traders into our model. These findings nicely complement the work done with higher frequency data by Singleton (2014).

REFERENCES

Anzuini, A., P. Pagano, and M. Pisani (2014). Macroeconomic effects of precautionary demand for oil. *Journal of Applied Econometrics 30*(6), 968–986.

Beetsma, R., M. Giuliodori, and F. Klaassen (2009). Temporal aggregation and SVAR identification, with an application to fiscal policy. *Economics Letters 105*(3), 253–255.

Cheng, I.-H., A. A. Kirilenko, and W. Xiong (2014). Convective risk flows in commodity

futures markets. *Review of Finance 19*(5), 1733–1781.

Chevallier, J. and S. Benoît (2013). A fear index to predict oil futures returns. Working Paper 62.2013, FEEM.

Chow, G. C. and A. Lin (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics 53*(4), 372–375.

Dotsey, M. (1998). The predictive content of the interest rate term spread for future economic growth. *Economic Quarterly - Federal Reserve Bank of Richmond 84*(3), 31–51.

Droumaguet, M. (2012). *Markov-Switching Vector Autoregressive Models: Monte Carlo Experiment, Impulse Response Analysis, and Granger-Causal Analysis*. Ph. D. thesis, European University Institute.

Faust, J. (2015). The robustness of identified var conclusions about money. *Carnegie Rochester Conference Series on Public Policy 49*, 207–244.

Fry, R. and A. Pagan (2011). Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature 49*(4), 938–960.

Gorton, G., F. Hayashi, and K. Rouwenhorst (2008). The fundamentals of commodity futures returns. Yale School of Management Working Papers.

Hamilton, J. D. and J. C. Wu (2015). Effects of index-fund investing on commodity futures prices. *International Economic Review 56*(1), 187–205.

Henderson, B. J., N. D. Pearson, and L. Wang (2014). New evidence on the financialization of commodity markets. *Review of Financial Studies 28*(5), 1285–1311.

Irwin, S. H. and D. R. Sanders (2012). Testing the masters hypothesis in commodity futures markets. *Energy Economics 34*(1), 256–269.

Jin, X. (2013). Short-run price-inventory dynamics in crude oil market. Southern Methodist University Working Paper.

Juvenal, L. and I. Petrella (2014). Speculation in the oil market. *Journal of Applied Econometrics 30*(4), 621–649.

Kilian, L. (2013). Structural vector autoregressions. In *Handbook of Research Methods and Applications in Empirical Macroeconomics*, pp. 515–554. Edward Elgar, UK.

Kilian, L. and D. P. Murphy (2013). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics 29*(3), 454–478.

Lanne, M. and H. Lütkepohl (2008). Identifying monetary policy shocks via changes in volatility. *Journal of Money, Credit and Banking 40*(6), 1131–1149.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2010). Structural vector autoregressions with markov switching. *Journal of Economic Dynamics and Control 34*(2), 121–131.

Lütkepohl, H. and A. Netšunajev (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural vars. *Journal of Applied Econometrics 29*(3), 479–496.

Masters, M. W. (2009). Testimony before the commodity futures trading commission. Technical report, Commodity Futures Trading Commission.

Mertens, K. and M. O. Ravn (2013). The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review 103*(4), 1212–1247.

Pindyck, R. S. (1994). Inventories and the short-run dynamics of commodity prices. *The RAND Journal of Economics 25*(1), 141–159.

Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics 85*(4), 777–792.

Saria, R., U. Soytasa, and E. Hacihasanoglu (2011). Do global risk perceptions influence world oil prices? *Energy Economics 33*(3), 515–524.

Singleton, K. J. (2014). Investor flows and the 2008 boom/bust in oil prices. *Management Science 60*(2), 300–318.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search* (2nd ed.). Cambridge, Massachusetts: The MIT Press.

Stock, J. H. and M. W. Watson (2008). Lecture 7 - recent developments in structural VAR modeling. Presentd at the NBER Summer Institute Econometrics Lectures: What's New in Economics – Time Series.

Tang, K. and W. Xiong (2012). Index investment and financialization of commodities. *Financial Analysts Journal 68*(6), 54–74.

Whaley, R. E. (2000). The investor fear gauge. *The Journal of Portfolio Management 26*(3), 12–17.

Appendix: Monte Carlo Analyses

Although we focus on our application to the crude oil futures market, in this appendix we present three Monte Carlo analyses to demonstrate the performance of our model. In the first analysis, we show how the estimators of the SVAR model behave. In the second analysis, we assess the performance of the model when the true data generating process is at a higher frequency than the observations. Finally, in the third analysis, we push the model to its limits, by assessing what happens when we increase the frequency of the true data generating process even more.

A1. Analysis 1

Concerning the estimation of the model, we present a brief Monte Carlo analysis here. Consider the model in equation (4). Following Droumaguet (2012), we let

$$\{s_t\}_{t=0,1,2,\ldots} = \{0, 1\},$$

$$p = 1,$$

$$c = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$A_1 = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.9 \end{pmatrix},$$

$$\Sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix},$$

$$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 25 \end{pmatrix},$$

and

$$Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}.$$

Obviously, with these parameters, the correct decomposition of Σ_0 and Σ_1 must be

$$\Sigma_0 = BB' = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{5} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{5} \end{pmatrix} \approx \begin{pmatrix} 1 & 0 \\ 0 & 2.236 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2.236 \end{pmatrix},$$

and

$$\Sigma_1 = B\Lambda B' = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{5} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{5} \end{pmatrix}.$$

The initial values are $y_0 = (0, 0)'$ and $s_0 = 0$. The first 300 simulated observations are dropped to warrant the randomness, and the following 300 observations are kept for the Monte Carlo experiments. We repeat 500 simulations in total, and the results are reported in Table 6.

[Insert Table 6 here]

From the table, we observe that all parameters are estimated with a low bias. Even for the two parameters (b_{12} and λ_{21}) that have a relatively large variance, we find that the bias in the estimated parameters remains limited.

A2. Analysis 2

On top of the estimation of parameters, we are especially interested in the performance of the model when the true data generating process is at a higher frequency than the observations. We present an example here to show how our model performs in this case.

In this analysis, the data generating process generally resembles the one in the previous process, whereas the shocks take place every other time, and the process is only observed in between. Assume

$$y_t = c + A_1 y_{t-1} + u_t,$$

where $u_t = 0$ for $t = 0, 2, 4, \ldots$ and $u_t | s_\tau \sim (0, \Sigma_{s_\tau})$ for $t = 1, 3, 5, \ldots$, and $\tau = (t-1)/2$. Similar with the previous example, $s_\tau, \tau = 0, 1, 2, \ldots$ is a Markov chain whose values

are either 0 or 1, and the transition matrix is the same:

$$Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}.$$

The parameters c, A_1, Σ_0 and Σ_1 take the same values as in the previous analysis. However, here we assume that the the data are censored at $t = 1, 3, 5, \ldots$ and we only observe y_t at $t = 0, 2, 4, \ldots$. The model established on the observed data should therefore be

$$\begin{aligned} y_{2\tau} &= c + A_1 y_{2\tau-1} + u_{2\tau} \\ &= c + A_1(c + A_1 y_{2\tau-2} + u_{2\tau-1}) + u_{2\tau} \\ &= c + A_1 c + A_1^2 y_{2\tau-2} + A_1 u_{2\tau-1}. \end{aligned}$$

for $\tau = 0, 1, 2, \ldots$. An observer who is not informed with the data generating process will thus believe the data are generated by an equivalent process

$$y_{2\tau} = \tilde{c} + \tilde{A}_1 y_{2\tau-2} + \tilde{u}_{2\tau}, \tag{A1}$$

where $\tilde{c} = c + A_1 c$, $\tilde{A}_1 = A_1^2$, and $\tilde{u}_{2\tau} = A_1 u_{2\tau-1} \sim (0, \tilde{\Sigma}_{s_\tau} \equiv A_1 \Sigma_{s_\tau} A_1')$.

We then randomly generate 1,000 observations and then drop every other one, thus obtaining a 500-observation sample. According to equation (A1), we expect to find the parameters in the equivalent model to be:

$$\tilde{c} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\tilde{A}_1 = \begin{pmatrix} 0.36 & 0 \\ 0 & 0.81 \end{pmatrix},$$

$$\tilde{\Sigma}_0 = \begin{pmatrix} 0.36 & 0 \\ 0 & 4.05 \end{pmatrix},$$

$$\tilde{\Sigma}_1 = \begin{pmatrix} 0.36 & 0 \\ 0 & 20.25 \end{pmatrix},$$

and for the observer transition matrix there is also

$$\tilde{Q} = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}.$$

As a result, the decomposition of the equivalent covariance matrices should be

$$\tilde{\Sigma}_0 = \tilde{B}\tilde{B}' = \begin{pmatrix} 0.6 & 0 \\ 0 & \sqrt{4.05} \end{pmatrix} \begin{pmatrix} 0.6 & 0 \\ 0 & \sqrt{4.05} \end{pmatrix} \approx \begin{pmatrix} 0.6 & 0 \\ 0 & 2.012 \end{pmatrix} \begin{pmatrix} 0.6 & 0 \\ 0 & 2.012 \end{pmatrix},$$

and

$$\tilde{\Sigma}_1 = \tilde{B}\tilde{\Lambda}\tilde{B}' = \begin{pmatrix} 0.6 & 0 \\ 0 & \sqrt{4.05} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 0.6 & 0 \\ 0 & \sqrt{4.05} \end{pmatrix}.$$

We run 500 simulations in total again, and the results are reported in Table 7. We see the accuracy of the estimations are comparable to the first analysis, where the SVAR model is consistent with the true data generating process.

[Insert Table 7 here]

A3. Analysis 3

We further consider a case where the data are generated at an even high frequency. Assume $u_{1_1}, \ldots, \epsilon_{1_H}, \ldots, u_{T_1}, \ldots, u_{T_H}$ are a series of shocks, and $y_1, \ldots, y_T$ is the series of data observed at a much lower frequency. For each $\eta = 1, \ldots, H$ in any $t = 1, \ldots, T$, the response of y_t to u_{t_η} is Φu_{t_η}, and the response of y_τ is zero if $\tau \neq t$. Therefore we have

$$y_t = \Phi \sum_{\eta=1}^{H} u_{t_\eta}$$

for $t = 1, \ldots, T$. We then assume that shocks follow two potential distributions, which is determined by an underlying Markov regime-switching process. Assume $s_t, \tau = 0, 1, 2, \ldots$ is a Markov chain whose values are either 0 or 1, and the transition matrix is denoted as Q. If $s_t = 0$, any u_{t_η} is drawn from the multivariate normal distribution $N(0, \Sigma_0)$, and $N(0, \Sigma_1)$ if $s_t = 0$, where $\Sigma_0 = \Psi\Psi'$, and $\Sigma_1 = \Psi\Omega\Psi'$.

Since we can only observe data at a low frequency whilst the shocks actually arrive at a much higher frequency, subjectively, responses to shocks are perceived to arise instantaneously and vanish instantaneously. Therefore the data-generating process is equivalent to an Markov Switching SVAR model with zero mean and covriance matrices of reduced-form residuals switching between $H^2\Phi\Psi\Psi'\Phi'$ and $H^2\Phi\Psi\Omega\Psi'\Phi'$. If we input the observed data into the Markov Switching SVAR model considered in this paper, we would expect the equivalent structural matrix $B = H\Phi\Psi$, and equivalent $\Lambda = \Omega$, while the transition matrix is still Q.

In this experiment, we let $T = 600$, and $H = 100$. The transition matrix is still:

$$Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}.$$

We also let

$$\Sigma_0 = \Psi\Psi' = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix} \cdot \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix} = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix},$$

and

$$\Sigma_1 = \Psi\Omega\Psi' = \Psi \cdot \begin{pmatrix} 25 & 0 \\ 0 & 25 \end{pmatrix} \cdot \Psi' = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}.$$

At last, we let

$$\Phi = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}.$$

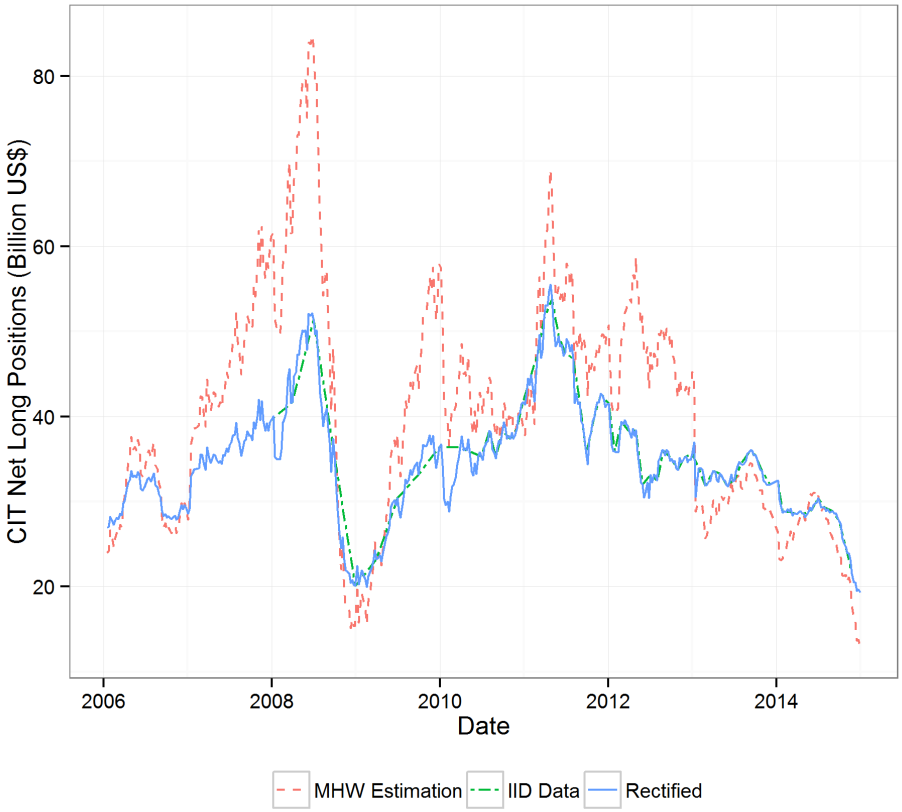
We still run 500 simulations in total, and the results are reported in Table 8. In this table, B^*, Λ^*, and Q^* are respectively the estimations of B, Λ, and Q. Although

the estimations are based on low-frequency observations, the aggregation responses of shocks that arrive at a high frequency are still reflected by the instantaneous responses in the low-frequency SVAR model.

[Insert Table 8 here]

Figures and Tables

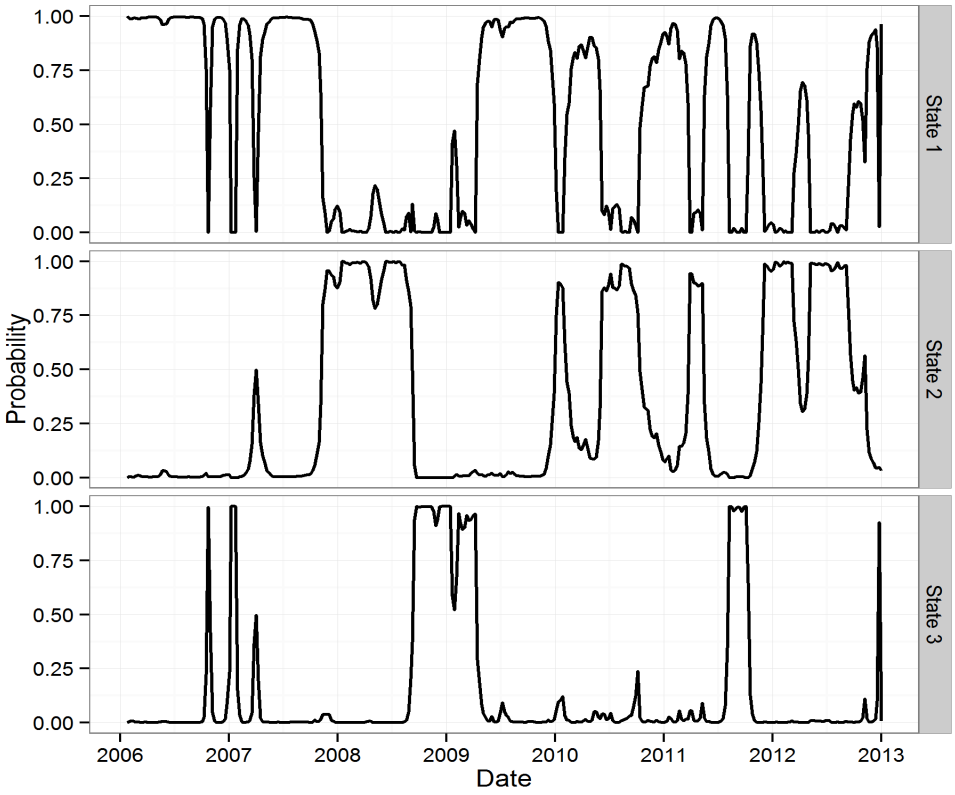
Figure 1. : CIT Net Long Positions Estimations



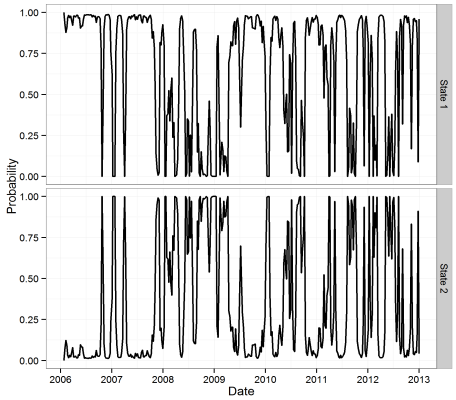
Dashed, double dashed, and solid lines represent estimations of CIT net long positions in billion US$ estimated with the MHW algorithm, reported by the CFTC IID (Index Investment Data) report, and corrected values.

Figure 2. : State Probabilities of Baseline Model with Different Number of States

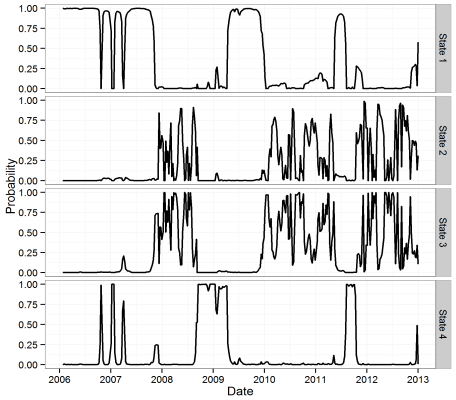
(a) 3 states (preferred specification)



(b) 2 states

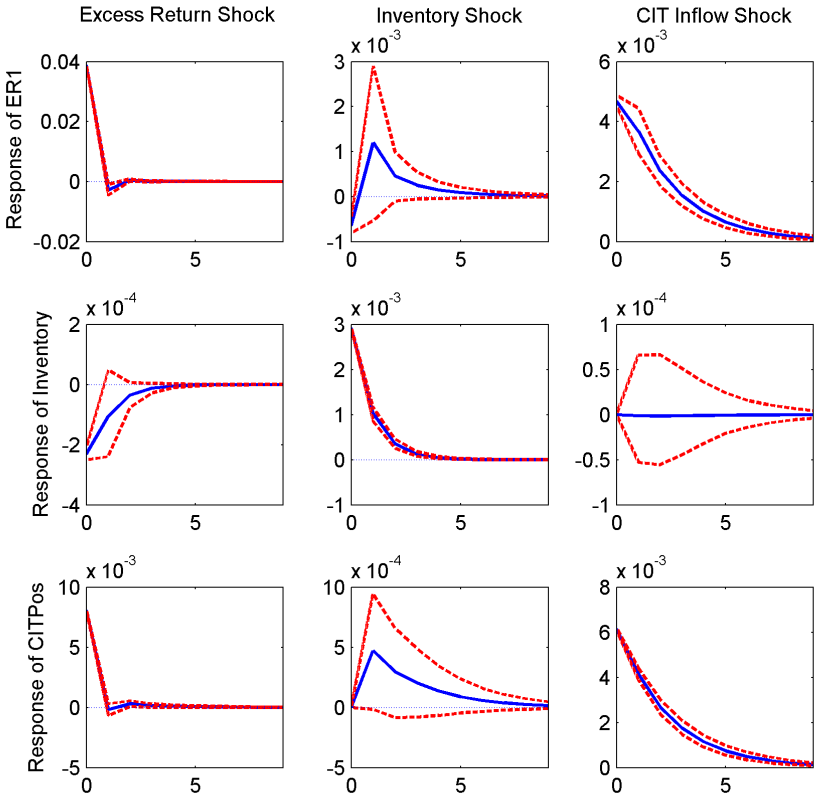


(c) 4 states



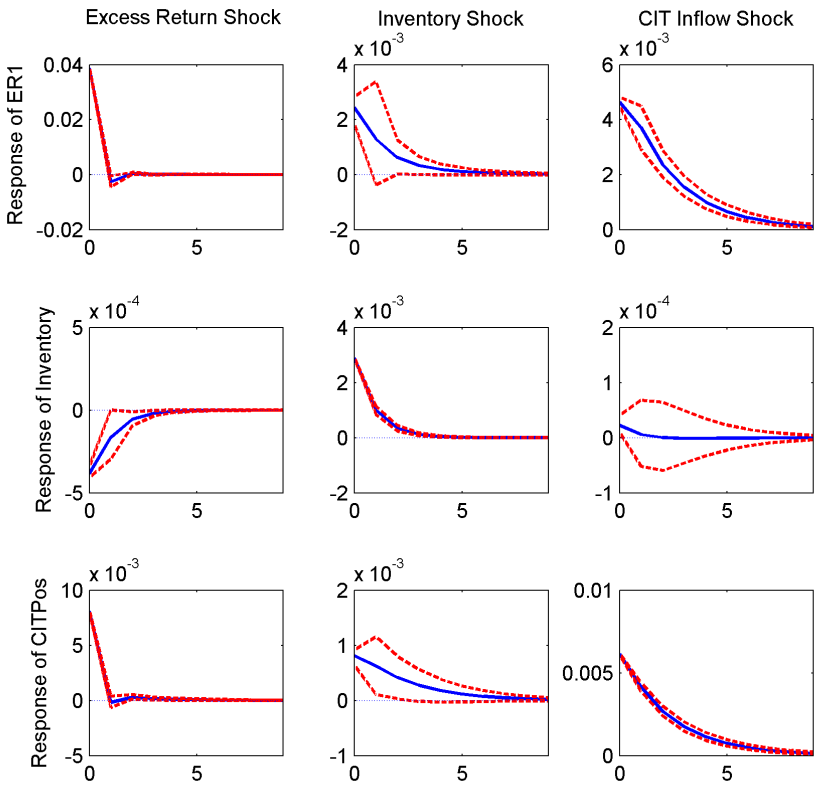
This figure shows the estimated probability of the MSVAR system with three volatility states, which contains WTI front futures contract excess returns, inventory growth, and CIT position growth, to be respectively in States 1, 2, 3 and 4.

Figure 3. : Impulse-Response Functions of the Restricted Baseline Model



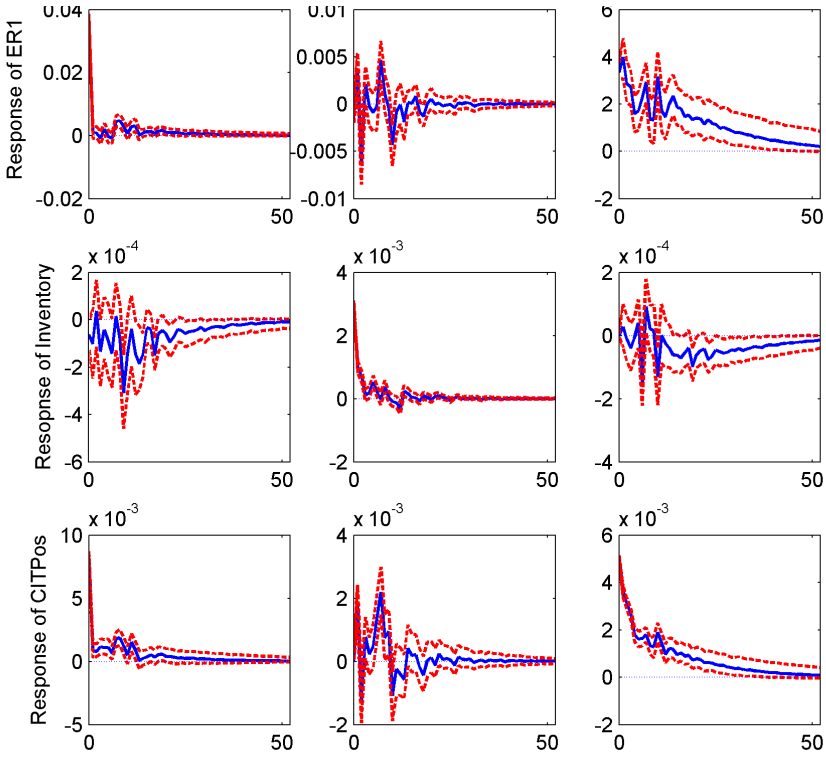
This figure shows the impulse-response functions of the restricted baseline model containing WTI front futures contract excess returns, inventory growth, and CIT position growth with 3 states. Each column represents a specific orthogonal shock as is identified by the estimation of the structural matrix, and each row represents the instantaneous and continued responses of a certain variable within 9 weeks. The solid line denotes the true IRF and the dashed lines mark off the 68% confidence intervals.

Figure 4. : Impulse-Response Functions of the Unrestricted Baseline Model



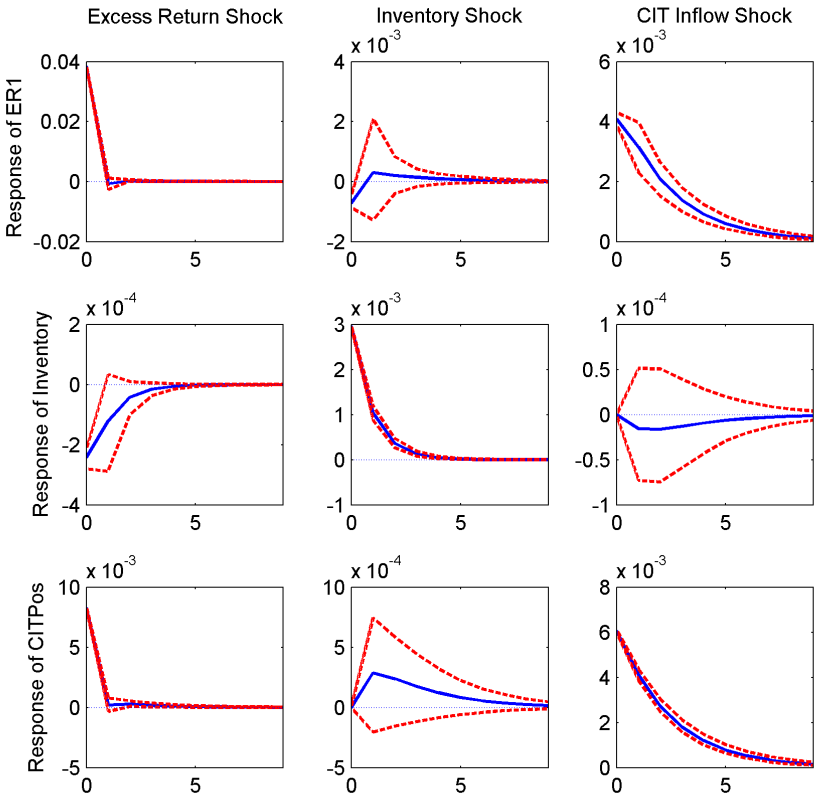
This figure shows the impulse-response functions of the unrestricted baseline model containing WTI front futures contract excess returns, inventory growth, and CIT position growth with three states. Each column represents a specific orthogonal shock as is semi-identified by the estimation of the structural matrix, and each row represents the instantaneous and continued responses of a certain variable within nine weeks. The labels of the shocks is attached through the comparison with the restricted model. The solid line denotes the true IRF and the dashed lines mark off the 68% confidence intervals.

Figure 5. : Impulse-Response Functions of the Restricted Baseline Model with 13 Lags



This figure shows the impulse-response functions of the restricted baseline model containing WTI front futures contract excess returns, inventory growth, and CIT position growth with three states and 13 lags. Each column represents a specific orthogonal shock as is identified by the estimation of the structural matrix, and each row represents the instantaneous and continued responses of a certain variable within 52 weeks. The solid line denotes the true IRF and the dashed lines mark off the 68% confidence intervals.

Figure 6. : Impulse-Response Functions of the Model with VIX and Term Spread Controlled



This figure shows the impulse-response functions of the model containing WTI front futures contract excess returns, inventory growth, and CIT position growth as endogenous variables, and VIX and term spread of US treasury bonds as exogenous variables with three states and one lag. Each column represents a specific orthogonal shock as is identified by the estimation of the structural matrix, and each row represents the instantaneous and continued responses of a certain variable within nine weeks. The solid line denotes the true IRF and the dashed lines mark off the 68% confidence intervals.

Table 1—: Estimated Error Covariances, Baseline Model with 3 States

State No.		Covariances			Eigen Values
		u_{ER1}	$u_{\text{Inventory}}$	u_{CITPos}	
1	u_{ER1}	.001527	-.000011	.000343	.000009
	$u_{\text{Inventory}}$	-.000011	.000009	-.000002	.000025
	u_{CITPos}	.000343	-.000002	.000103	.001606
2	u_{ER1}	.001453	-.000010	.000492	.001641
	$u_{\text{Inventory}}$	-.000010	.000013	-.000002	.000163
	u_{CITPos}	.000492	-.000002	.000351	.000013
3	u_{ER1}	.010723	-.000065	.002560	.011354
	$u_{\text{Inventory}}$	-.000065	.000012	-.000013	.000326
	u_{CITPos}	.002560	-.000013	.000957	.000011

This table shows the estimated covariance matrices and their eigen values of the non-orthogonal residuals in the reduced-form VAR model containing WTI front futures contract excess returns, inventory growth, and CIT position growth with three states. The three variables are respectively abbreviated as ER1, Inventory, and CITPos.

Table 2—: Information Criteria of Different Numbers of States

Number of States	AIC	SC
2	-6577.22	-6483.82
3	-6608.75	-6488.11
4	-6608.31	-6452.65

This table shows the AIC and Schwartz Criterion (SC) values for the SVAR model containing WTI front futures contract excess returns, inventory growth, and CIT position growth with two, three, and four different volatility states. The three variables are respectively abbreviated as ER1, Inventory, and CITPos.

Table 3—: Estimated Λ's in Baseline Model with 3 States

	Unrestricted Model		Restricted Model	
	Estimation	SD	Estimation	SD
λ_{21}	0.840 316	0.200 677	0.851 057	0.275 694
λ_{22}	1.529 337	0.829 916	1.512 752	0.317 398
λ_{23}	7.830 264	3.563 535	7.822 529	1.616 984
λ_{31}	6.974 877	2.915 683	6.930 944	3.758 024
λ_{32}	1.300 846	0.747 950	1.310 487	0.515 381
λ_{33}	13.348 555	6.852 705	13.277 863	6.675 058

This table shows the estimated diagonal elememts of Λ_i of State i of the unrestricted and restricted baseline model, where λ_{ij} denotes the j-th diagonal element in Λ_i.

Table 4—: Tests of the Equality among λ's

Null Hypothesis	Wald Statistic	p-Value
$\lambda_{21} = \lambda_{22}$	1.574	0.057 75
$\lambda_{21} = \lambda_{23}$	4.250	0.000 01
$\lambda_{21} = \lambda_{22}$	3.829	0.000 06
$\lambda_{31} = \lambda_{32}$	1.482	0.069 21
$\lambda_{31} = \lambda_{33}$	0.829	0.203 68
$\lambda_{31} = \lambda_{32}$	1.788	0.036 93

This table shows the Wald statistics and respective p-values of the tests of the pairwise equality among the diagonal elements of Λ_2 and Λ_3 of State 2 and 3 of the restricted baseline model. λ_{ij} denotes the j-th diagonal element in Λ_i.

Table 5—: Coefficients and Wald Statistics of Exogeneous Variables

Model	Endogeneous Var.	TermSprd_{t-1}		VIX_{t-1}	
		Coefficient	Wald Stat.	Coefficient	Wald Stat.
(1)	ER1_t	0.020 154	1.384		
	Inventory_t	0.000 509	0.489		
	CITPos_t	0.004 385	1.006		
(2)	ER1_t			0.016 372	0.853
	Inventory_t			−0.000 950	−0.654
	CITPos_t			−0.001 199	−0.235
(3)	ER1_t	0.019 229	1.375	0.017 133	0.884
	Inventory_t	0.000 601	0.616	−0.001 026	−0.692
	CITPos_t	0.003 978	0.952	−0.000 954	−0.190

This table shows the coefficients of the exogenous variables and respective Wald statistics.

Table 6—: Monte Carlo Analysis 1

Parameter	Component/Element	True Value	Mean	Variance
c	c_1	0.000	0.000 295	0.003 877
	c_2	0.000	−0.000 102	0.054 215
A_1	a_{11}	0.600	0.586 254	0.002 059
	a_{12}	0.000	0.010 764	0.026 284
	a_{21}	0.000	0.000 374	0.000 055
	a_{22}	0.900	0.888 535	0.000 782
B	b_{11}	1.000	0.990 843	0.013 173
	b_{12}	0.000	0.004 290	0.008 727
	b_{21}	0.000	−0.012 335	0.289 684
	b_{22}	2.236	2.082 039	0.148 946
Λ	λ_{11}	1.000	0.996 814	0.149 508
	λ_{22}	5.000	6.190 787	5.000 563
Q	q_{00}	0.800	0.750 209	0.025 488
	q_{01}	0.200	0.249 791	0.025 488
	q_{10}	0.200	0.241 747	0.027 461
	q_{11}	0.800	0.758 253	0.027 461

This table shows the mean and the variance of parameter estimations of the model in equation 4 in the Monte Carlo experiments. c_1 and c_2 are respectively the first and second components of c, q_{00} to q_{11} follow the same notation in equation III.B, and a_{ij}, b_{ij} and λ_{ij} in rest rows denote the element of A_1, B, and Λ in the i-th row and j-th column. Since Λ is required to be diagonal in the estimation, we only report the diagonal elements.

Table 7—: Monte Carlo Analysis 2

Parameter	Component/Element	True Value	Mean	Variance
$\tilde{c}$	$\tilde{c}_1$	0.000	−0.001 702	0.001 317
	$\tilde{c}_2$	0.000	−0.000 330	0.038 614
$\tilde{A}_1$	$\tilde{a}_{11}$	0.360	0.350 682	0.003 062
	$\tilde{a}_{12}$	0.000	0.003 010	0.091 033
	$\tilde{a}_{21}$	0.000	−0.000 131	0.000 004
	$\tilde{a}_{22}$	0.810	0.801 007	0.001 281
$\tilde{B}$	$\tilde{b}_{11}$	0.600	0.598 974	0.004 349
	$\tilde{b}_{12}$	0.000	0.002 910	0.002 957
	$\tilde{b}_{21}$	0.000	−0.028 042	0.228 178
	$\tilde{b}_{22}$	2.012	1.910 030	0.120 709
$\tilde{\Lambda}$	$\tilde{\lambda}_{11}$	1.000	0.968 466	0.133 478
	$\tilde{\lambda}_{22}$	5.000	5.956 050	4.598 460
$\tilde{Q}$	$\tilde{q}_{00}$	0.800	0.758 156	0.025 945
	$\tilde{q}_{01}$	0.200	0.241 843	0.025 945
	$\tilde{q}_{10}$	0.200	0.242 198	0.023 510
	$\tilde{q}_{11}$	0.800	0.757 802	0.023 510

This table shows the mean and the variance of parameter estimations of the censored model in the Monte Carlo experiments. $\tilde{c}_1$ and $\tilde{c}_2$ are respectively the first and second components of $\tilde{c}$, $\tilde{q}_{00}$ to $\tilde{q}_{11}$ follow the same notation in equation III.B, and $\tilde{a}_{ij}$, $\tilde{b}_{ij}$ and $\tilde{\lambda}_{ij}$ in rest rows denote the element of $\tilde{A}_1$, $\tilde{B}$, and $\tilde{\Lambda}$ in the i-th row and j-th column. Since $\tilde{\Lambda}$ is required to be diagonal in the estimation, we only report the diagonal elements.

Table 8—: Monte Carlo Analysis 3

Parameter	Component/Element	True Value	Mean	Variance
B^*	b^*_{11}	1.000	0.997 754	0.002 306
	b^*_{12}	0.000	0.000 422	0.000 179
	b^*_{21}	0.000	0.007 081	0.097 375
	b^*_{22}	4.000	3.960 467	0.069 386
Λ^*	λ^*_{11}	1.000	1.010 285	0.024 300
	λ^*_{22}	25.000	25.807 279	13.772 550
Q^*	q^*_{00}	0.800	0.795 442	0.001 068
	q^*_{01}	0.200	0.204 270	0.001 334
	q^*_{10}	0.200	0.204 558	0.001 068
	q^*_{11}	0.800	0.795 730	0.001 334

This table shows the mean and the variance of parameter estimations of the model in the Monte Carlo Analysis 3. q^*_{ij} represents the transition probaility from State i to State j, and b^*_{ij} and λ_{ij} denote the element of B^*, and Λ^* in the i-th row and j-th column. Since Λ^* is required to be diagonal in the estimation, we only report the diagonal elements.