NBER WORKING PAPER SERIES

SHOCK RESTRICTED STRUCTURAL VECTOR-AUTOREGRESSIONS

Sydney C. Ludvigson
Sai Ma
Serena Ng

Working Paper 23225
http://www.nber.org/papers/w23225

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2017

Ludvigson acknowledges financial support from the C.V. Starr Center for Applied Economics at NYU. Ng acknowledges support from the National Science Foundation under grant SES-0962431 and SES-1558623. We are grateful to Lutz Kilian, Jóse Luis Montiel Olea, Stephen Terry, Jim Stock, Eric Swanson, and Ian Dew-Becker for many valuable discussions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Shock Restricted Structural Vector-Autoregressions
Sydney C. Ludvigson, Sai Ma, and Serena Ng
NBER Working Paper No. 23225
March 2017
JEL No. C01,C5,C51,E17

ABSTRACT

Identifying assumptions need to be imposed on autoregressive models before they can be used to analyze the dynamic effects of economically interesting shocks. Often, the assumptions are only rich enough to identify a set of solutions. This paper considers two types of restrictions on the structural shocks that can help reduce the number of plausible solutions. The first is imposed on the sign and magnitude of the shocks during unusual episodes in history. The second restricts the correlation between the shocks and components of variables external to the autoregressive model. These non-linear inequality constraints can be used in conjunction with zero and sign restrictions that are already widely used in the literature. The effectiveness of our constraints are illustrated using two applications of the oil market and Monte Carlo experiments calibrated to study the role of uncertainty shocks in economic fluctuations.

Sydney C. Ludvigson
Department of Economics
New York University
19 W. 4th Street, 6th Floor
New York, NY 10002
and NBER
sydney.ludvigson@nyu.edu

Serena Ng
Department of Economics
Columbia University
420 West 118th Street
New York, NY 10027
and NBER
Serena.Ng@columbia.edu

Sai Ma
Department of Economics
New York University
19 W. 4th STreet, 6th Floor
New York, NY 10012
sai.ma@nyu.edu

1 Introduction

A challenge in economic analysis is that the data represented by a vector autoregression (VAR) can be consistent with many causal structures differentiated by distinct economic models and primitive assumptions. Structural vector-autoregressive models (SVAR) provide a simple framework that enables researchers to perform counter-factual experiments without fully characterizing all primitives or micro-foundations that lead to the dynamic system. Concisely stated, a n-variable SVAR analysis consists of finding an $n \times n$ matrix $\mathbf{B}$ that relates the reduced form errors $\boldsymbol{\eta}_t$ to the structural errors $\mathbf{e}_t$:

$$\boldsymbol{\eta}_t = \mathbf{B}\mathbf{e}_t.$$

The data provide $\boldsymbol{\eta}_t$, but the above relationship only provides $n(n+1)/2$ pieces of information about $\mathbf{B}$. Hence $n(n-1)/2$ restrictions are necessary for identification. Sims (1980) originally proposed a triangularized system to identify $\mathbf{B}$, but long-run economic restrictions, statistical restrictions based on heteroskedasticity of the VAR innovations, or additional information in the form of high frequency data, variables external to the SVAR, or narrative descriptions have also been employed to identify $\mathbf{B}$.[1] Point identification exists if these restrictions are enough to yield a unique solution. In some cases, point identification is achievable only by imposing restrictions that are difficult to defend. It may then be desirable to abandon the goal of point identification in favor of less restrictive economic assumptions that are supported by the data.

In this paper we are concerned with applications for which defensible economic restrictions may not permit point identification, but may nonetheless be enough to achieve a substantial constriction of the set of model parameters consistent with the data. Our focus specifically is on two new methods for winnowing the set of plausible solutions, both of which directly restrict the identified structural shocks. The first approach rests on restrictions that require the identified shocks to be consistent with economic reasoning in a small number of extraordinary events, or to accord with our historical understanding of events at particular points in the sample. We refer to these as “event timing constraints,” or simply *event constraints*. The second approach rests on non-linear restrictions pertaining to the comovement between components of external variables and the SVAR shocks. We refer to these as “component correlation constraints,” or *correlation constraints* for short. We show in this paper that the event and component correlation constraints can be used separately, and possibly in conjunction with other identification schemes that already exist in the literature. The effectiveness of these constraints in constricting the set of plausible solutions is also investigated.

The motivation for the event constraints is as follows. Every identification scheme necessarily yields a set of primitive shocks $\mathbf{e}$. While the dynamic causal effects and the forecast error variance attributable to the primitive shocks are often intensively analyzed, a typical SVAR

[1] For a comprehensive review of SVAR models, see Ramey (2016), Kilian and Lutkepohl (2016).

analysis pays much less attention to the estimated shocks themselves, even though the stated goal of the exercise is to identify $\mathbf{e}$. A credible identification scheme should produce estimates of $\mathbf{e}$ with features that accord with our ex-post understanding of historical events, at least during episodes of special interest. For example, a scheme that identifies a large positive output shock in the 1982 recession would be dismissed because the existence of a shock would be hard to defend given the historical account of the events at the time. Such special events turn out to be valuable for identification because, although two feasible structural models $\mathbf{B}$ and $\tilde{\mathbf{B}}$ will generate shocks $\mathbf{e}_t$ and $\tilde{\mathbf{e}}_t$ with equivalent first and second moments, the $\mathbf{e}_t$ and $\tilde{\mathbf{e}}_t$ are not necessarily the same at any given t. In other words, two series with equivalent properties "on average"can still have distinguishable features in certain subsamples.

The idea of using specific events to identify shocks is not new. For example, the narrative approach to summarizing historical information has been widely used to construct shock series. Most often, this is done by introducing a dummy variable that takes on the value of one when a primitive shock to some variable is thought to have occurred based on a reading of historical events. Oil price shocks have been constructed from disruptive political events, tax shocks from specific fiscal policy announcements, and monetary policy shocks from a reading of FOMC meetings. In such narrative analyses, the constructed shock series is typically used as though it were exogenous and accurately measured, and subsequently used in an SVAR with additional restrictions imposed. But as noted in Ramey (2016), the narrative construction does not ensure that the shocks are truly exogenous or are correctly measured. Mertens and Ravn (2014) suggest that failure to account for measurement error in the tax shocks constructed by the narrative approach can explain why some SVAR estimates of tax multipliers can be close to zero. Their approach is to use the narrative tax changes as an external instrument, to generate additional moments that complement those generated from the covariance structure. But exogeneity of the tax shocks is still assumed.

The event constraints used in this paper differ from the narrative approach to constructing events in a number of ways. Our event constraints are data driven and not narrative based. We use features of the shocks during selected episodes to determine whether a possible solution is admissible. This is tantamount to creating dummy variables from the timing of specific events, and then putting restrictions on their correlation with the identified shocks. Note that the same SVAR is used to identify all shocks simultaneously; it is not a two-step procedure that identifies some shocks ahead of others. A challenge with exactly identified SVARs is that there are few tools for model validation. In our framework, a recursive ordering can be recovered if such a structure is consistent with the constraints. Such flexibility comes at the expense of foregoing point identification in favor of set identification.

Our second type of economic restriction, what we call component correlation constraints, relates to the external instrumental variable (IV) approach studied by Stock (2008) and Mertens and Ravn (2013) that exploits external variables $\mathbf{Z}$ to identify one shock in a SVAR. For the external IV idea to work, each external variable must be a valid instrument in the sense of satisfying both a relevance and an exogeneity condition. Our approach differs in several ways. We identify not just one, but all shocks in the system, and we do not require the external variables to be exogenous, or at least not known to be exogenous a priori. To be clear that our external variables are not valid instruments, we refer to our external variables as *synthetic proxies* and label them $\mathbf{S}$. Such synthetic proxies must be presumed to have valuable information about the parameters of interest. 'Valuable' is defined in terms of a lower bound on the unit-free correlation between relevant components of $\mathbf{S}$ and the identified shocks, akin to the instrument relevance condition.

In what follows, we first use a SVAR for the oil market to illustrate how event and correlation constraints can be separately used with zero and sign restrictions to shrink the set of plausible solutions. The constraints thus have more general applicability than what was considered in Ludvigson, Ma, and Ng (2015) (LMN hereafter). In that paper, both types of constraints were used jointly to study the role of uncertainty shocks in economic fluctuations. We then use the LMN application to design Monte Carlo experiments. This is used to assess the sampling errors that could be expected of these constraints.

2 Econometric Framework

Let $\mathbf{X}_t$ denote a $n \times 1$ vector time series. We suppose that $\mathbf{X}_t$ has a reduced-form vector autoregression and an infinite moving average representation given, respectively, by:

$$\begin{aligned} \mathbf{A}(L)\mathbf{X}_t &= \boldsymbol{\eta}_t. && (1)\\ \mathbf{X}_t &= \Psi(L)\boldsymbol{\eta}_t && (2)\\ \boldsymbol{\eta}_t &\sim (0,\boldsymbol{\Omega}), \quad \boldsymbol{\Omega} = \mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t') = \mathbf{PP}' \end{aligned}$$

where $\mathbf{A}(L) = \mathbf{I}_n - \sum_{j=1}^{p} \mathbf{A}_j \mathbf{L}^j$, $\boldsymbol{\Psi}(L) = \mathbf{I}_n + \boldsymbol{\Psi}_1 L + \boldsymbol{\Psi}_2 L^2 + \ldots$ is a polynomial in the lag operator L of infinite order, and $\boldsymbol{\Psi}_s$ is the $(n \times n)$ matrix of coefficients for the sth lag of $\boldsymbol{\Psi}(L)$. The reduced form parameters are collected into $\phi = \big(\text{vec}(\mathbf{A}_1)' \quad \ldots \quad \text{vec}(\mathbf{A}_p)', \quad \text{vech}(\boldsymbol{\Omega})'\big)'$. The reduced form innovations $\boldsymbol{\eta}_t$ are related to the structural-form vector autoregression (SVAR) shocks $\mathbf{e}_t$ by an invertible $n \times n$ matrix $\mathbf{H}$:

$$\begin{aligned} \boldsymbol{\eta}_t &= \mathbf{H\Sigma e}_t \equiv \mathbf{Be}_t && (3)\\ \mathbf{e}_t &\sim (0, \mathbf{I_K}), \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 & \cdot & 0 \\ 0 & \sigma_{22} & 0 & 0 \\ 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \sigma_{KK} \end{pmatrix}, \quad \sigma_{jj} \geq 0 \; \forall j. && (4) \end{aligned}$$

where $\mathbf{B} \equiv \mathbf{H}\boldsymbol{\Sigma}$. The structural shocks $\mathbf{e}_t$ are mean zero with unit variance, and are serially and mutually uncorrelated. The objective is to study the dynamic effects of the structural shocks, or the impulse responses functions (IRF), defined by $\boldsymbol{\Theta}(L) = \boldsymbol{\Psi}(L)\mathbf{B}$,

By definition, $\boldsymbol{\Omega} = \mathbf{P}\mathbf{P}'$ where $\mathbf{P}$ is the unique lower-triangular Cholesky factor with non-negative diagonal elements. Let $\mathbb{O}_n$ be the set of $n \times n$ orthonormal matrices. Now any $\mathbf{B} = \mathbf{P}\mathbf{Q}$ is consistent with the reduced form covariance matrix $\boldsymbol{\Omega} = \mathbf{B}\mathbf{B}'$ provided that $\mathbf{Q} = \begin{pmatrix} q_1 & q_2 & \dots & q_n \end{pmatrix} \in \mathbb{O}_n$. The set of observationally equivalent $\mathbf{B}$s, given $\boldsymbol{\Omega}$, can thus be defined as $\mathcal{B} = \{\mathbf{B} = \mathbf{P}\mathbf{Q} : \mathbf{Q} \in \mathbb{O}_n.\}$ At this point, the only innocuous restriction we can impose is

$$\text{diag}\,(\mathbf{H}) = 1, \quad \text{or} \quad \text{diag}(\mathbf{B}) \geq 0, \tag{5}$$

where the restriction on the sign of $\mathbf{B}$ follows from combining the unit effect normalization on $\mathbf{H}$ with the restriction $\sigma_{jj} \geq 0$. We may then interpret a unit change in structural shock j as a standard deviation increase in variable j, and express the set of observationally equivalent $\mathbf{B}$ corresponding to $\boldsymbol{\Omega}$ as

$$\tilde{\mathcal{B}}(\phi) = \{\mathbf{B} = \mathbf{P}\mathbf{Q} : \mathbf{Q} \in \mathbb{O}_n, \text{diag}(\mathbf{B}) \geq 0.\}.$$

But economic theory, intuition, and our understanding of historical events often allow us to discard some $\mathbf{B}$ in $\tilde{\mathcal{B}(\phi)}$. As in Rubio-Ramirez, Waggoner, and Zha (2010), Moon, Schorfheide, and Granziera (2013) and related work in this literature, let $FZ(\mathbf{Q};\phi)$ denote the collection of zero restrictions imposed on the model. Point identification obtains if there is a unique $\mathbf{Q}$ that satisfies $FZ(\mathbf{Q};\phi) = 0$. If $\mathbf{Q} = \mathbf{I}_n$, then the solution is the recursive structure consistent with the ordering of the variables used to construct $\mathbf{P}$. Let f_{zi} be the number of zero restrictions on q_i and $f_z = \sum_{i=1}^{n} f_{zi}$. For models with variables ordered such that $f_{z1} \geq f_{z2} \ldots \geq f_{zn} \geq 0$, Rubio-Ramirez, Waggoner, and Zha (2010) show that a necessary and sufficient condition for point identification is $f_{zi} = n - i$ for all $i = 1, \ldots, n$. A model is under-identified by zero restrictions if $f_{zi} \leq n - i$ with strict inequality for at least one i.

A model is partially identified when there is more than one $\mathbf{Q}$ that satisfy $FZ(\mathbf{Q};\phi) = 0$. In such cases, researchers have used various additional restrictions to dismiss solutions in $\tilde{\mathcal{B}}(\phi)$ leading to a smaller set $\overline{\mathcal{B}}(\phi)$ that satisfies the additional identifying restrictions. A notable example is sign restrictions, or more generally inequality restrictions, of the form $FS(\mathbf{Q};\phi) \geq 0$ (e.g., Uhlig (2005)). Existing theoretical and empirical work tends to place these constraints on the impulse response functions and/or $\mathbf{B}$. In terms of $\mathbf{X}_t = \boldsymbol{\Psi}(L)\mathbf{B}\mathbf{e}_t$, the restrictions have focused on $\boldsymbol{\Psi}(L)\mathbf{B}$. Our restrictions involve the identified shocks $\mathbf{e}_t$ either on their own in the form of event constraints $FE(\mathbf{Q};\phi,\mathbf{X},\tau^*,\bar{k})$, or combined with external variables in the form of component correlation constraints $FC(\mathbf{Q};\phi,\mathbf{X},\mathbf{Z},\bar{c})$. These constraints are explained in the next two subsections.

2.1 Event Constraints

Event constraints put bounds $\bar{k}$ on the sign and magnitude of $\mathbf{e}_t = \mathbf{B}^{-1}\boldsymbol{\eta}_t$ during selected episodes collected into τ^*. These shocks are useful for identification because, from $\mathbf{e}_t = \mathbf{B}^{-1}\boldsymbol{\eta}_t = \mathbf{Q}'\mathbf{P}^{-1}\boldsymbol{\eta}_t$, we see that for $\tilde{\mathbf{Q}} \neq \mathbf{Q}$,

$$\tilde{\mathbf{e}}_t = \tilde{\mathbf{Q}}'\mathbf{P}^{-1}\boldsymbol{\eta}_t = \tilde{\mathbf{Q}}\mathbf{e}_t \neq \mathbf{e}_t$$

at any given t. This implies that constraints involving the shocks at specific time periods in the sample could be used to constrict the number of solutions in $\tilde{\mathcal{B}}(\phi)$. To illustrate the point, consider the $n = 2$ case:

$$\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}.$$

so that solving for e_{1t} gives

$$e_{1t} = |\mathbf{B}|^{-1}(B_{22}\eta_{1t} - B_{12}\eta_{2t}),$$

where $|\mathbf{B}| = B_{11}B_{22} - B_{12}B_{21}$ is the determinant of $\mathbf{B}$. The values of η_{1t} and η_{2t} are given since we have data for t in the span $[\tau_1, \tau_2]$. Hence, a restriction on the behavior of e_{1t_1} at specific time t_1 is a non-linear restriction on $\mathbf{B}$, or equivalently, on $\mathbf{Q}$. With non-Gaussian reduced-form errors, one can also see that the third and higher order moments of $\mathbf{e}_{1t}$ are not invariant to $\mathbf{B}$, hence $\mathbf{Q}$. This is in spite of the fact that the first and second moments of $\mathbf{e}_t$ are invariant to $\mathbf{Q}$. There is thus information in $\mathbf{e}_t$ that can be used to identify $\mathbf{B}$,

Imposing restrictions on $\mathbf{e}_t$ at specific time periods τ^* in SVAR was an idea first considered in LMN. Recent work by Antolin-Diaz and Rubio Ramírez (2016) also suggest to use restrictions on the sign and relative importance of the shocks to help identification. That the properties of the shocks have not been much exploited in identification is somewhat surprising since economic theory and reasoning, narrative interpretations of history, and/or statistical analyses of the data often provide guidance as to what shocks have occurred when, and which ones should be systematically related to variables we can observe. We denote these event constraints with the vector

$$FE(\mathbf{Q}; \phi, \mathbf{X}, \tau^*, \bar{k}) = (FE_1(\tau_1^*, \bar{k}_1), \ldots, FE_n(\tau_n^*, \bar{k}_n))' \geq 0,$$

where $\mathbf{X}$ enters because $\boldsymbol{\eta}_t = \mathbf{A}(L)\mathbf{X}_t$. Each $FE_i(\tau_i, \bar{k}_i)$ represents a vector of constraints of magnitude $\bar{k}_i$ on e_{it} for all $t \in \tau_i^*$. The motivation is that if a $\mathbf{Q}$ implies a shock series that is difficult to defend in certain episodes, it can be removed from $\tilde{\mathbf{B}}(\phi)$. Such constraints could be imposed on extraordinary events such as the major recessions, wars, and natural disasters that have been well-documented. For example, if the first shock (say monetary policy) is presumed to be strongly contractionary in $\tau_i^* = (1979{:}10, 1979{:}11, 1979{:}12)$, then one could formulate

$$FE = \begin{pmatrix} -\sum_{t=1}^{T} 1_{t=,1979:10} \cdot e_{1t} \\ -\sum_{t=1}^{T} 1_{t=,1979:11} \cdot e_{1t} \\ -\sum_{t=1}^{T} 1_{t=,1979:12} \cdot e_{1t} \end{pmatrix} - \begin{pmatrix} \bar{k}_1 \\ \bar{k}_1 \\ \bar{k}_1 \end{pmatrix} \geq 0$$

to dismiss solutions that imply highly expansionary monetary policy shocks in these episodes. The parameter $\bar{k}_1$ could be two (standard deviations) or some other lower bound that reflects how contractionary these shocks are thought to be. The k-th row of FE_1 represents an inequality with $1_{t\in\tau^*_{i,k}}$ as instrument. In essence, $FE(\phi, \mathbf{Q}, \mathbf{X}; \tau^*; \bar{k})$ defines conditions based on the timing, sign, and magnitude of the events to help identification. Of course, if $\bar{k}_i$ is too big, or if the timing of the event τ^*_i is inaccurate, the solutions will be meaningless even if they exist. On the other hand, if shocks are systematically found at particular episodes when no restrictions are explicitly imposed, we can be more confident of their occurrence.

2.2 Component Correlation Constraints

Component correlation constraints put bounds $\bar{c}$ on the correlation between the identified shocks and certain components of variables external to the VAR, which we denote by $\mathbf{S}$. Similar to the event constraints, constraints involving the correlation between $\mathbf{e}_t = \mathbf{B}^{-1}\boldsymbol{\eta}_t$ and observables $\mathbf{S}$, can be used to constrict the number of solutions in $\tilde{\mathcal{B}}(\phi)$. Thus let

$$FC(\mathbf{Q}; \phi, \mathbf{X}, \mathbf{S}, \bar{c}) \geq 0,$$

be the collection of component correlation constraints that put possibly nonlinear restrictions across the parameters. An example helps understand the motivation of these constraints.

Consider a two variable model

$$\mathbf{A}(L)\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \qquad \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \sim \mathcal{N}(0, I_2),$$

where I_2 is a 2×2 identity matrix. The covariance structure $\boldsymbol{\Omega} = \mathbf{B}\mathbf{B}'$ provides three unique pieces of information, so one more restriction would be needed for point identification. If an instrumental variable Z_t exists such that (i) $\mathbb{E}[Z_t e_{1t}] = 0$ (exogeneity) and (ii) $\mathbb{E}[Z_t e_{2t}] \neq 0$ (relevance), then a unique solution for $\mathbf{B}$ can be obtained.

Suppose that, instead of a valid instrument Z_t, we have an external variable S_t that is not assured to be exogenous. Being endogenous means that S_t is contemporaneously correlated with at least one structural shock. This suggests we could represent S_t by

$$S_t = \gamma e_{1t} + \Gamma e_{2t} + \sigma_S e_{St} \tag{6}$$

where e_{St} is a S-specific shock uncorrelated with e_{1t}, e_{2t} by assumption. To be able to use S_t to identify $\mathbf{B}$, we need to say more about γ and Γ. If we are willing to assume that the variable S_t has a component that is correlated with e_{2t} but uncorrelated with e_{1t}, which in the above example will be true whenever $\Gamma \neq 0$, we may want to discard solutions in $\tilde{\mathcal{B}}(\phi)$ for which the absolute correlation between the component $S_t - \gamma e_{1t}$ and e_{2t} is too small.[2] The quantity

$$c(\mathbf{Q}) = \frac{\Gamma}{\sqrt{\Gamma^2 + \sigma_S^2}},$$

measures the correlation between the component $S_t - \gamma e_{1t}$ and e_2. Note that a small Γ does not necessarily imply a small c as the value of σ_S^2 must be taken into account. Requiring that $c(\mathbf{Q}) > \bar{c}$ is the same as $\frac{\Gamma^2}{\Gamma^2+\sigma_S^2} > \bar{c}^2$, which is a non-linear constraint on the parameters of the S_t equation, which are themselves functions of the shocks and data on S_t. That c is between zero and one facilitates the parameterization of $\bar{c}$.

A feature of the component $S_t - \gamma e_{1t}$ is that it is uncorrelated with e_{1t}, but correlated with e_{2t}, as is required of an instrument. We refer to $S_t - \gamma e_{1t}$ as a *synthetic proxy variable* to distinguish it from an external instrument. Below we will denote such component proxy variables $Z(\mathbf{Q})$ to emphasize that they are a function of parameters to be estimated. In conventional IV estimation, instrument exogeneity is a maintained assumption. Of course, S_t itself is a valid exogenous instrument if $\gamma = 0$. But when validity of the exogeneity assumption is questionable, then S_t is at best *plausibly exogenous* in the terminology of Conley, Hansen, and Rossi (2012). These authors consider estimation of the equation for X_{1t} with endogenous regressor X_{2t} when instrument exogeneity is not known to hold exactly, but the parameter is point identified when the exogeneity assumption is valid. They put bounds on the effect due to the invalid instrument, or what we refer to as S_t, on X_{1t}. In contrast, we analyze X_{1t} and X_{2t} jointly, and we put more structure on S_t so that a lower bound can be placed between its subcomponent that is relevant as an instrument and the shocks. Like Conley, Hansen, and Rossi (2012), such a bound will not, in general, be enough to achieve point identification. But it could dismiss solutions that do not achieve this bound, akin to dismissing weak instruments.

The correlation restrictions can also be imposed from a systems perspective. Consider instead of a SVAR in $\mathbf{X}_t$ an augmented SVAR in $(\mathbf{X}, \mathbf{S})$. Suppressing lags for convenience:

$$\begin{aligned} X_{1t} &= B_{11}e_{1t} + B_{12}e_{2t} + B_{1S}e_{St} \\ X_{2t} &= B_{21}e_{1t} + B_{22}e_{2t} + B_{2S}e_{St} \\ S_t &= \gamma e_{1t} + \Gamma e_{2t} + B_{SS}e_{St}. \end{aligned}$$

The objective is to identify both e_{1t} and e_{2t}, not merely one of the two shocks. Exogeneity of S_t imposes that S_t does not affect $\mathbf{X}_t$. But $B_{1S} = B_{2S} = 0$ along with the covariance structure of $\mathbf{X}_t$ are not enough to identify the parameters $B_{11}, B_{12}, B_{21}B_{22}$ of the two variable SVAR. One way to proceed is to impose further restrictions in the form zero, sign, or event constraints,

[2]Note that, if S_t has exactly the structural form given in (6), it is not actually necessary to construct this component to impose the correlation constraint. From (6), it is evident that in this case a bound for the minimum absolute correlation between e_{2t} and $S_t - \gamma e_{1t}$ is isomorphic to a bound on the correlation between e_{2t} and S_t, since $e_{1t}, e_{2,t}$, and e_{st} are uncorrelated by construction.

without appealing to the excluded variable $\mathbf{S}$. But it is also possible use information in $\mathbf{S}$ more fully. To do so, we need to know the values of γ, Γ and B_{SS}. The purpose of the component correlation constraints is to restrict the plausible values of these parameters. Collectively, the winnowing constraints turn an under-identified system into a partially identified one.

In conventional analyses, the instruments are observed and can be readily used to formulate sample orthogonality conditions. If the synthetic variables $\mathbf{Z}(\mathbf{Q})$ were to be thought of as instruments, at least conceptually, they would need to be constructed prior to setting up the estimating equations for $\mathbf{B}$. But the construction of $\mathbf{Z}(\mathbf{Q})$ itself necessitates estimates of $\mathbf{B}$. Iteration would thus be necessary to ensure that the $\mathbf{B}$ used to construct $\mathbf{Z}(\boldsymbol{\beta})$ is internally consistent with the solution that emerges, which comes at a computation cost. In practice, we find that initializing $\mathbf{Q}$ randomly already yields $\mathbf{e}_t(\mathbf{Q})$ and $\mathbf{Z}_t(\mathbf{Q})$ for all the necessary computations. The solutions are almost identical to ones obtained by iteration, even though the synthetic variables are jointly estimated.

In the systems approach, $\mathbf{S}$ are variables of the augmented SVAR. This contrasts with the previous approach that uses an SVAR for $\mathbf{X}$ with $\mathbf{S}$ specified as external variables. There are advantages and disadvantages of each to consider. In the systems approach, $\mathbf{S}$ can only be explained by lags of $\mathbf{S}$ and $\mathbf{X}$. Other predictors of $\mathbf{S}$ are all subsumed in the error e_S, which could be restrictive. And as is well documented, misspecification of one equation in a system can have ramifications for the entire system. But the systems approach allows lags of $\mathbf{S}$ to affect $\mathbf{X}$, and hence allows for feedback that treating $\mathbf{S}$ as external variables would not permit. The SVAR for $\mathbf{X}$ thus imposes over-identification restrictions on the larger system that can in principle be evaluated.

2.3 Overview

The event and correlation constraints can be used individually, jointly, and possibly in conjunction with zero restrictions, short- or long-run restrictions, and/or sign restrictions. Collecting all the restrictions into $F(\mathbf{Q}; \phi, \bar{k}, \bar{c}, \tau^*)$ or simply $F(\bar{k}, \bar{c}, \tau^*)$, the *winnowed* solution set is

$$\begin{aligned}
\bar{\mathcal{B}}(\phi, F) \quad = \quad & \{\mathbf{B} = \mathbf{PQ} : \mathbf{Q} \in \mathbb{O}_n, \ \text{diag}(\mathbf{B}) > 0, \\
& FZ(\mathbf{Q}; \phi) = 0, \\
& FS(\mathbf{Q}; \phi) \geq 0, \\
& FC(\mathbf{Q}; \phi, \mathbf{X}, \mathbf{S}, \bar{c}) \geq 0, \\
& FE(\mathbf{Q}; \phi, \mathbf{X}, \tau^*, \bar{k}) \geq 0\}.
\end{aligned}$$

A particular solution can be in both $\tilde{\mathcal{B}}(\phi)$ and $\overline{\mathcal{B}}(\phi, F)$ only if all restrictions defined by $F(\cdot)$ are satisfied. In general, $\bar{\mathcal{B}}(\phi, F)$ is a set. Setting the thresholds $\bar{c}$ or $\bar{k}$ at values that are too restrictive will lead $\bar{\mathcal{B}}(\phi, F)$ to be empty. A unique solution in $\bar{\mathcal{B}}(\phi, F)$ can be obtained by making additional assumptions. For example, Fry and Pagan (1911) consider the solution that yields an impulse response closest to the median. This implementation would be difficult when there are multiple impulse response functions of interest. In our setup, one can consider the solution that yields the smallest (or largest) $|e_{t_1}|$ amongst those that satisfy the event constraints. It is also possible to use the fact that associated with each solution in $\bar{\mathcal{B}}(\phi, F)$ is a vector of correlation coefficients $c(\mathbf{Q}) = (c_1(\mathbf{Q}), c_2(\mathbf{Q}), \ldots, c_{f_c(\mathbf{Q})})'$. If there are K solutions indexed by $k = 1, ...K$, we will have $c^1(\mathbf{Q}), \ldots, c^K(\mathbf{Q})$ vectors of correlation coefficients. Then there is only one vector that obtains

$$C^{k_{\max}}(\mathbf{Q}) \equiv \max_k \sqrt{c^k(\mathbf{Q})' c^k(\mathbf{Q})}. \tag{7}$$

For future reference, we refer to the single solution $\mathbf{B}$ that generates $C^{k_{\max}}(\mathbf{Q})$ as the "max-C" solution. Metrics other than the quadratic norm can be specified by the researcher. What is important is that the maximum is unique. It should, however, be noted that selecting one particular solution for further analysis does not mean that the solution is more likely to explain the data, as discussed in Kilian and Murphy (2012).

In the illustrations to follow, $FZ(\mathbf{Q}; \phi)$ is a vector of $n(n+1)/2$ restrictions implied by the covariance structure of the model, but these are not enough to identify n^2 parameters. The first application in Section 3 does not involve correlation constraints of type $FC(\mathbf{Q}; \phi, \mathbf{X}, \mathbf{S}, \bar{c})$, while the second application includes all types of winnowing constraints. The example considered in Section 4 does not involve constraints of type $FS(\mathbf{Q}; \phi)$. In all cases, the possible solutions in $\tilde{\mathcal{B}}(\phi)$ are obtained by initializing $\mathbf{B}$ to be the lower Cholesky factorization of $\mathbf{\Omega}$ for an arbitrary ordering of the variables, and then rotating it by $K = 1.5$ million random orthogonal matrices $\mathbf{Q}$. More precisely, each rotation begins by drawing an $n \times n$ matrix $\mathbf{G}$ of NID(0,1) random variables. Then $\mathbf{Q}$ is taken to be the orthonormal matrix in the $\mathbf{QR}$ decomposition of $\mathbf{G} = \mathbf{QR}$ and $\mathbf{QQ}' = \mathbf{I}_n$.

3 An Empirical Analysis of Oil Market

As an example of the identifying potential of the previously proposed shock-restrictions, this section re-examines shocks to the oil market using event and correlation constraints. The SVAR system builds on the oil market system studied in Kilian and Murphy (2012) (KM hereafter). They consider an SVAR with three variables estimated from 1973:01-2008:09:

$$\mathbf{X}_t = \begin{pmatrix} X_{1t} \\ X_{2t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \Delta prod_t \\ rea_t \\ rpo_t \end{pmatrix}, \tag{8}$$

where $\Delta prod_t$ is the percentage change in global crude oil production, rea_t is the global demand of industrial commodities variable constructed in Kilian (2009), and rpo_t is the real oil price. KM refer to rea_t simply as an "aggregate demand" shock.

The three structural shocks of interest are collected into the vector $\mathbf{e}_t$,

$$\mathbf{e}_t = \begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix} = \begin{pmatrix} e_{\Delta prod,t} \\ e_{rea,t} \\ e_{rpo,t} \end{pmatrix}.$$

The first is a shock to the production of crude oil (oil supply shock), the second is a shock to aggregate commodity demand, and the third is a oil market specific demand shock. Kilian (2009) imposes sign restrictions, which we denote $FS_K(\phi, \mathbf{Q})$, on the impact responses as follows:

	Shock		
X	Oil supply disruption	Aggregate demand	Oil demand
$\Delta prod_t$	$-$	$+$	$+$
rea_t	$-$	$+$	$-$
rpo_t	$+$	$+$	$+$

The table shows the presumed impact response of each variable named in the row to an impulse in each shock named in the column. Note that an oil supply disruption is by construction a negative supply shock. KM impose further restrictions, denoted $FS_{KM}(\phi, \mathbf{Q})$, that limit the elasticity of supply to demand shocks, which amounts to imposing an upper bound on the ratios of impact impulse responses $\frac{B_{13}}{B_{33}}$ and $\frac{B_{12}}{B_{32}}$. The upper bound for both is set to be 0.0258. We refer to the Kilian sign restrictions in combination with the KM elasticity bounds as the KKM restrictions. The next two subsections re-examine the SVAR using some new identifying constraints.

3.1 Application 1: Event Constraints on Oil Production

Much has been written about the correlation between oil price on the one hand, and geopolitical as well economic events on the other. See, for example, Hamilton (2013) and Baumeister and Kilian (2016) for recent reviews. However, their causal relations and more precisely the relative importance of the sources of fluctuations in the oil market is still a matter of debate. The latent shocks that we recover necessarily depend on the identifying assumptions used. Instead of a recursive structure or an elasticity bound, we now consider an application that combines sign restrictions with event constraints to constrict the set of possible solutions.

Figure 1 plots the change in the real oil price over time. The three largest spikes in the unconditional oil price change occur in (i) 1974:01, following the start of the OPEC embargo in October 1973 (which lasted until March of 1974), (ii) 1986:02, following the collapse of OPEC, and (iii) 1990:08, the month of the Kuwait invasion by the U.S. If a reading of events in these months suggests that the large oil price changes are partly attributable to oil supply shocks, we would expect a spike of the appropriate sign in the structural e_1 shock during these episodes. Such an assumption leads to the following event constraints on oil supply shocks:

Oil event constraints $FE(\phi, \mathbf{Q}, \mathbf{X}, \tau^*, \bar{k})$

i $e_{1t_1}(\mathbf{Q}) \leq \bar{k}_1$ where t_1 is the period 1974:01 during the OPEC Embargo.

ii $e_{1t_2}(\mathbf{Q}) \geq \bar{k}_2$ where t_2 is the period 1986:02 following the the collapse of OPEC.

iii $e_{1t_3}(\mathbf{Q}) \leq \bar{k}_1$ where t_3 is the period 1990:08 of the Kuwait invasion.

Note that there is nothing in the above that explicitly precludes the presence of the demand shocks e_2 and e_3 from playing an important role in these episodes, or that restricts the relative importance of the three shocks in any particular episode. The event constraints merely require that supply shocks played some role in the price spikes, where the magnitude and sign of that role are determined by the bounds $\bar{k}_1$ and $\bar{k}_2$. The constraints dismiss parameter estimates that suggest overly favorable supply conditions during the post-OPEC embargo event, or the month of the Kuwait invasion, and too unfavorable conditions during the OPEC collapse. To implement these restrictions, we need to set bounds $\bar{k} \equiv \left(\bar{k}_1, \bar{k}_2\right)'$ for the oil supply shocks. In the baseline case, we set $\bar{k}_1 = -3$ and $\bar{k}_2 = 2.5$. This is guided by the fact that the unconditional change in the real price of oil was at least 4 standard deviations higher than its mean in the periods corresponding to events (i) and (iii), and 4 standard deviation lower than its mean in the period corresponding to event (ii). As a sensitivity check, we also consider bounds that are less restrictive by setting $\left(\bar{k}_1, \bar{k}_2\right) = (-2.5, 2)$. In total, the $F(\bar{k}, \tau^*) = \{FS_K,\ FE(\cdot)\}$ constraints for this application combine the Kilian sign restrictions with the event constraints.

We extend the data used in KM to the sample 1973:01 to 2016:06. Our baseline case uses six lags in the VAR. After losing observations to lags and differencing, the estimation sample is 1973:08-2016:06. The first price spike occurs in 1974:01, shortly after the onset of the OPEC embargo that occurred in 1973:10. We therefore estimate the SVAR with lags $p = 6$ so that the large price change episode is in the sample. The KM analysis starts in 1973:01 and uses $p = 24$ lags to capture the long swings in the oil market. We also consider a 24 lag version, which shortens the sample to 1975:02-2016:06. Estimation over this shorter sample eliminates the first event entirely, thereby removing an identifying restriction.

To assess how the constraints affect the identified impulse response functions, we first impose sign restrictions alone as in Kilian (2009). Among 1.5 million rotations, $4,878$ solutions satisfy the sign restrictions. We then impose the KKM constraints that combine the sign restrictions with the elasticity bounds of KM. Only 34 solutions satisfy the KKM constraints, implying that the KM constraint alone severely shrinks $\tilde{\mathcal{B}}(\phi)$. Not surprisingly, there is no solution that survives the sign restrictions, the elasticity bounds, and our event constraints. However, there are $2,143$ solutions that satisfy our $F(\bar{k}, \tau^*)$ constraints with $\bar{k}_1 = -3$ and $\bar{k}_2 = 2.5$, and $3,147$ solutions when $\bar{k}_1 = -2.5$ and $\bar{k}_2 = 2$.

Figure 2 shows the responses of the real price of oil to an oil supply shock, an aggregate commodity demand shock, and an oil market specific demand shock. All figures show responses to one standard deviation shocks in the direction that raise the price of oil. The left panel displays the results under the KKM constraints, and the right panel shows the results under the $F(\bar{k}, \tau^*)$ constraints. Note first that the response of the oil price to the two demand shocks is rather robust to the set of restrictions imposed. Hence we focus on the response to the oil supply shock. The left panel shows that the KKM constraints yield identified responses of the oil price to an oil supply shock that constitute a small subset of those formed from the Kilian sign restrictions alone. The identified responses in this small subset are numerically small (top panel), indicating that the solutions generating the large responses of oil price to a supply shock have been eliminated by the KM constraint.

The right panel of Figure 2 shows the responses produced by our $F(\bar{k}, \tau^*)$ constraints which combine the Kilian sign restriction with the event constraints. This is shown for the base case with $\left(\bar{k}_1, \bar{k}_2\right) = (-3, 2.5)$, and also for the less restrictive constraint of $\left(\bar{k}_1, \bar{k}_2\right) = (-2.5, 2)$. As seen from the right panel of Figure 2, the responses identified by the tighter bounds are a subset of those identified by the weaker bound, which are in turn a subset of the responses constructed from the sign restrictions alone. In contrast to the KKM constraints, the $F(\bar{k}, \tau^*)$ restrictions preserve solutions that produce larger oil price responses to an oil supply shock. These results are robust to using 24 lags and losing the first event, as reported in the Appendix. The results are also robust to using the shorter KM sample, and are available on request.

Interestingly, among all solutions that survive our winnowing constraints, the smallest elasticity $\frac{B_{13}}{B_{33}}$ is 0.1792, which is bigger than the KM bound of 0.0258. The lowest elasticity $\frac{B_{12}}{B_{32}}$ is 0.00004. Amongst the solutions that survive the KKM constraints, the smallest $e_{1t_1}(\mathbf{Q})$ is 0.2777 and the largest $e_{1t_2}(\mathbf{Q})$ is 0.6950. But both of these solutions are rejected by our event constraints, even when we use the weaker bound of $\left(\bar{k}_1, \bar{k}_2\right) = (-2.5, 2)$. Taken together, the estimates imply that, if we restrict the values for the elasticity of oil production to oil *demand* shocks, as in KM, the survived solutions imply that oil *supply* shocks play little or no role in oil price fluctuations. On the other hand, solutions that satisfy the event constraints imply a greater elasticity of oil supply to oil demand shocks than that permitted by KM.

Since a stated objective of an SVAR analysis is to identify the structural shocks, the properties of the shocks are of interest. The e_{1t} series identified by the $F(\bar{k}, \tau^*)$ constraints exhibits non-Gaussian features. Averaged across solutions, the coefficient of skewness and kurtosis are -0.7736 and 6.8513, respectively. By contrast, the e_{1t} series implied by the KKM constraints have an average skewness of -1.4603 and a kurtosis of 11.0722. The stronger departure from Gaussianity arises because the KKM constraints accept solutions that imply larger shocks.[3]

[3]Specifically, the KKM constraints identified large negative e_1s in 1975:10, 1977:01, 1978:01, 1980:10, 1983:02, 1986:09, 1990:08, and large positive e_1s in 1975:09, 1977:02 and 1987:07.

To have a clearer picture of what shocks are identified by $F(\bar{k}, \tau^*)$, the left panel of Figure 3 plots the timing of "large shocks" over the sample; the solid vertical lines are three event dates given above. Large shocks are defined to be those in excess of two standard deviations above or below the mean. The figure reports the average magnitude of the large shocks so defined across all solutions in the winnowed set $\bar{\mathcal{B}}(\phi, F)$. The figure compares the large shock episodes under KKM with those under $F(\bar{k}, \tau^*)$. By design, the e_1 shocks generated by $F(\bar{k}, \tau^*)$ displayed in red should be less than -3 standard deviations in 1974:01 and 1990:08, which is indeed the case. While no additional large oil supply shocks are found, there are 19 occasions when e_{1t} exceeds two standard deviations. By way of comparison, the KKM constraints also identified two negative supply shocks around the events corresponding to our t_1 and t_3 dates, but the negative supply shock in the first event occurs slightly earlier, in 1973:11 rather than 1974:01. On the other hand, the KKM restrictions find no big positive supply shocks in the middle event; on the contrary, all big supply shocks surrounding these dates are negative rather than positive. Interestingly, the KKM constraints only identify a few big demand shocks of either type prior to 1990. On the other hand, the $F(\bar{k}, \tau^*)$ constraints did not find the big negative supply shock associated with the Iranian Revolution around 1978 that the KKM constraints recovered.

We have imposed restrictions on the supply shock e_{1t}, and it is of interest to examine what they imply for the demand shocks e_{2t} and e_{3t}. The left panel of Figure 3 shows that our event constraints by no means preclude other big shocks from occurring at the same time. Indeed, the $F(\bar{k}, \tau^*)$ constraints generate several large aggregate commodity demand shocks of both signs after 2010, several negative oil-specific demand shocks before 1980, and negative demand shocks of both types concurrently in 1986. While demand shocks appear to be important, large supply shocks outside of constrained event dates are numerous under both identification schemes. Regardless of which identification scheme is used, large negative supply shocks coexist with large positive demand shocks in both 1973/74 and 1990.

We have thus far illustrated our approach using restrictions on e_{1t} at three particular episodes. While supply shocks during these three episodes seem reasonable, some might see demand shocks as being relatively more important. In a separate exercise, we place the event constraints on the demand shocks e_{2t} or e_{3t} instead of e_{1t}, thereby forcing at least one them to be large in the earlier event dates characterized by large price changes. This approach is agnostic as to which demand shock moves. That is, the restrictions require that *at least one* of e_{2t} or e_{3t} be large and positive ($\geq$ 3 standard deviations above the mean) in 1974:01 and 1990:08, but large and negative (≤ -2 standard deviations) in 1986:02. Sixty solutions survived these constraints. But as seen from the right hand panel of Figure 3, we still find supply shocks during the oil embargo in 1973:11 and again in the Kuwait invasion month 1990:08 to be large and negative. This is similar to when the supply shock e_{1t} was explicitly constrained, though the negative supply shock in the first episode is now two months earlier than the price spike itself. However, e_{1t} does not spike up at all in any month of 1986. Together with the consistent finding of large negative demand shocks in 1986, the interpretation of the 1986 OPEC collapse episode is less clear. We return to this point below.

The above analysis focused on the identified sets of impulse responses. It is straight forward to likewise construct sets of variance decompositions. We briefly report a few of these statistics here. For the long-horizon (infinite-step-ahead) forecast error variance, the KKM restrictions imply that effectively none of the forecast error variance in the real price of oil is attributable to supply shocks: the fraction explained by supply shocks covers the range [0.02, 0.04]. That is, of all the solutions that survive the KKM restrictions, the lowest fraction of variance explained is 2% and the highest is 4%. By contrast, the aggregate commodity demand shock explains between 32% and 91% of the oil price variance across the solutions, while the oil demand shock explains between 7% and 65%. The corresponding numbers using the $F(\bar{k}, \tau^*)$ constraints are [0.19, 0.68] for the supply shock, [0.19, 0.80] for the aggregate demand shock, and [0.00, 0.32] for the oil demand shock. Thus, supply shocks play a much larger relative role in oil price fluctuations under these identifying restrictions. Under either type of identifying restrictions, the relative importance of the demand shocks ranges widely across the solutions in the winnowed set. These ranges can be tightened with additional constraints on the forecast errors attributable to the identified shocks. Such constraints can also be interpreted as a type of correlation constraint.

The purpose of this application is to show that event constraints can be combined with sign restrictions to shrink the set of plausible solutions. Our intent is not to argue in favor of one event constraint over another, but rather to show that shocks identified by different $\mathbf{Q}$s have distinguishable features at specific points in the sample even though they may have equivalent first and second moments computed from the full sample. Such distinguishable features can substantially narrow the set of solutions deemed credible. The results of this application bear this out.

3.2 Application 2: Shortfall in Oil Production as Synthetic Proxy

Instrumental variables are typically chosen because economic reasoning or theory suggests that they ought to be correlated with the endogenous variable of interest. In the oil example above, some have argued that oil production shortfalls are at least partly caused by political events such as wars or embargoes (e.g., Hamilton (2013)). Hence if the production shortfalls are uncorrelated with the two demand shocks e_{2t} and e_{3t}, one might consider a measure of such shortfalls as an instrument for oil supply shocks, if data for such a variable is available. Kilian (2008) constructed such a series, which Olea, Stock, and Watson (2015) use as an external instrument in an SVAR designed to identify oil supply shocks.[4]

To use the shortfall series as an external instrument, the variable must be truly exogenous and the data measured without error. Relaxing the exogeneity assumption would yield inconsistent estimates. However, as discussed above, even if the shortfall series is not exogenous with respect to the two demand shocks, it may still be relevant for oil supply shocks and we can restrict the set of possible solutions by placing a lower bound on the presumed relevance. We therefore consider a second oil shock application that illustrates the use of the correlation constraints discussed above, in which the oil supply shortfall variable is our S_t. The data for this application are based on the replication files of Olea, Stock, and Watson (2015) and span the period 1973:02 to 2004:09, or 1973:09-2004:09 after losing observations to lags and differencing. Hence the sample is slightly different from the previous application.

Consider the following equation for the oil production shortfall S_t:

$$S_t = \gamma_0 + \gamma_1 e_{2t} + \gamma_2 e_{3t} + \gamma_s(L) S_{t-1} + Z_t. \tag{9}$$

This specification captures the main feature that S_t is correlated with the shocks of interest, hence endogenous. Equation (9) forms an orthogonal decomposition of S_t into a component that is spanned by the two oil demand shocks e_{3t} and e_{2t} and a composite residual component orthogonal to these. The residual is the constructed component Z_t, which plays the role of $S_t - \gamma e_{1t}$ in the two-variable example above. In general, Z_t will be serially correlated as it may include lagged values of $\mathbf{X}_t$. The constructed component is again denoted $Z(\mathbf{Q})$ to emphasize that it is a function of parameters $\mathbf{B} = \mathbf{PQ}$ in the SVAR for $\mathbf{X}_t$. The component correlation constraint for this application is defined as follows:

Oil Correlation Constraint $FC(\phi, \mathbf{Q}, \mathbf{X}, \mathbf{S}, \bar{c})$**:**

- $|c(\mathbf{Q})| > \bar{c}$, where $c(\mathbf{Q}) = \text{corr}(Z_t(\mathbf{Q}), e_{1t}(\mathbf{Q}))$ is the sample correlation between $Z_t(\mathbf{Q})$ defined as in (9) and the supply shock $e_{1t}(\mathbf{Q})$.

As c is a correlation coefficient, $\bar{c}$ must be between zero and one in absolute value. Olea, Stock, and Watson (2015) are concerned that the production shortfall variable could be a weak instrument. We do not treat S_t as a valid instrument, but nonetheless assume that it is relevant for supply shocks while taking the weak IV consideration into account. In the baseline

[4] This variable is an estimate of production short-falls caused by wars or civil disturbances in OPEC countries. As explained in Kilian (2008), the production short-fall variable is generated by first computing a counterfactual production level for the country in question that would have occurred if the war had not. The counterfactual production level is generated by extrapolating its pre-war production level based on the average growth rate of production in other countries that are subject to the same global macroeconomic conditions and economic incentives, but are not involved in the war. Kilian chooses the countries that belong in the benchmark group on a case-by-case basis drawing on historical accounts and industry sources.

implementation, we set a relatively unrestrictive bound of $\bar{c} = 0.03$. We assess the sensitivity to this bound.

The results are displayed in Figure 4. The benchmark is the identified set of impulse responses based on Kilian's sign restrictions alone, $FS_K(\phi, \mathbf{Q})$. These are shown on the left. This is compared to the set of responses after imposing the correlation constraint $FC(\phi, \mathbf{Q})$, combining them with sign restrictions, and also adding the oil event constraints on e_{1t} considered in the previous application to both FS_K and FC to arrive at $F(\bar{k}, \bar{c}, \tau^*) = \{FS_K,\ FC(\cdot),\ FE(\cdot)\}$. These results are shown on the right panel. Because the sample size is different from the first application and correlation constraints are now imposed, the $\bar{k}$ bounds used earlier renders the winnowed set empty. Hence we adjust the bounds on the event constraints to be less restrictive at $\bar{k}_1 = -2$ and $\bar{k}_2 = 2$, compared with $\bar{k}_1 = -3$ and $\bar{k}_2 = 2.5$ used earlier. Among 1.5 million rotations, $3,838$ solutions satisfy the sign restrictions. While 330 solutions satisfy the sign and correlation constraints, only eight solutions satisfy sign, event, and correlation constraints. All eight solutions yield a non-zero estimate of γ_1 and γ_2, casting doubt on the exogeneity of production shortfall variable. Further investigation reveals that the main reason that only eight solutions survive all three constraints is that few solutions deliver a positive oil supply shock in 1986:02.

The plots on the left show that sign restrictions alone leave a wide range of possible responses to each structural shock, with no clear response pattern associated with any one structural shock, as KM have previously emphasized. The responses based on sign and correlation constraints are narrower, while those based on $F(\bar{k}, \bar{c}, \tau^*)$ that combine the three sets of restrictions are significantly narrower for all three shocks. That the identified impulse responses are narrower than using sign restrictions alone arises in part because external information in the form of correlation constraints contributes to identification, especially for the responses to the aggregate demand shock e_{2t}. The event constraints are relatively more important for the supply shock. Taken together, this analysis suggests that oil supply shocks and aggregate commodity demand shocks have large and persistent effects on the price of oil, while oil specific demand shocks have much smaller effects.

To learn more about the effectiveness of the correlation constraints, Figure 5 depicts the large shocks based on sign and correlation constraints, but without imposing event constraints. A direct comparison with Figure 3 must be made with care because the sample for the second application ends in 2004:9. Keeping in mind this difference, we see that imposing FC yields large negative oil supply shocks in 1974:01 and 1990:08, the previously imposed event constraint dates and both episodes of dramatic price spikes upward. This shows that the identified set obtained by imposing the component correlation constraints using the external proxy variable S_t implies that large negative supply shocks contributed non-trivially to the price spikes during the oil embargo and Kuwait invasion, even if we don't impose event constraints in those episodes.

However, a large (positive) supply shock is found in the OPEC collapse of 1986:02 only if we additionally impose an event constraint that seems incongruent with all but eight solutions. Meanwhile there were large negative shocks to aggregate demand e_2 in 1986:02, and oil specific demand in 1986:09. Taking the results from the two applications together, the sharp drop in oil price in 1986 is likely due not to one shock, but to a combination of shocks. This is unlike the oil price hikes during the oil embargo and the Kuwait invasion when evidence for important negative supply shocks is more compelling.

4 Monte Carlo Simulation: Uncertainty Shocks

In point-identified models, sampling uncertainty can be evaluated using frequentist confidence intervals or Bayesian credible regions, and they coincide asymptotically. Inference for set-identified SVARs is, however, more challenging because no consistent point estimate is available. As pointed out in Moon and Schorfheide (2012), the credible regions of Bayesian identified impulses responses will be distinctly different from the frequentist confidence sets, with the implication that Bayesian error bands cannot be interpreted as approximate frequentist error bands. Our analysis is frequentist, and while the two applications presented above illustrate how the dynamic responses vary across estimated models, where each model is evaluated at a solution in $\bar{\mathcal{B}}(\phi, F)$, we still need a way to assess the robustness of our procedure, especially since it is new to the literature.

Unfortunately, few methods are available to evaluate the sampling uncertainty of partially identified SVARs from a frequentist perspective, and these tend to be specific to the imposition of particular identifying restrictions. Moon, Schorfheide, and Granziera (2013) suggest a projections based method within a moment-inequality setup, but it is designed to study SVARs that only impose restrictions on one set of impulse response functions. Furthermore, the method is computationally intense, requiring a simulation of critical value for each rotation matrix. Gafarov, Meier, and Olea (2015) suggest to collect parameters of the reduced form model in a $1-\alpha$ Wald ellipsoid but the approach is conservative. For the method to get an exact coverage of $1-\alpha$, the radius of the Wald-ellipsoid needs to be carefully calibrated. As discussed in Kilian and Lutkepohl (2016), even with these adjustments, existing frequentist confidence sets for set-identified models still tend to be too wide to be informative. It is fair to say that there exists no generally agreed upon method for conducting inference in set-identified SVARs. While we do not have a fully satisfactory solution to offer, our restrictions can further tighten the identified set, and by implication the confidence sets. We now explore this possibility in simulations, taking two approaches.

Let R be the number of replications in the Monte Carlo experiment. The first approach is based on the properties of a set of solutions in repeated samples. In each replication, K

rotation matrices are entertained, but only $K_r \leq K$ rotations of $\mathbf{Q}$ will generate solutions that are admitted into the winnowed set for that replication, $\bar{\mathcal{B}}^r(\phi, F)$. Let $\Theta^{r,k}_{i,j,s}$ be the s-period ahead response of the ith variable to a standard deviation change in shock j at the k-th rotation of replication r. Let $\underline{\Theta}^r_{i,j,s} = \min_{k\in[1,K_r]} \Theta^{r,k}_{i,j,s}$ and $\overline{\Theta}^r_{i,j,s} = \max_{k\in[1,K_r]} \Theta^{r,k}_{i,j,s}$. Each $(\underline{\Theta}^r_{i,j,s}, \overline{\Theta}^r_{i,j,s})$ pair represents the extreme (highest and lowest) dynamic responses in replication r. From the quantiles of $\underline{\Theta}^r_{i,j,s}$, we can obtain the $\alpha/2$ critical point $\underline{\Theta}_{i,j,s}(\alpha/2)$. Similarly, from the quantiles of $\overline{\Theta}^r_{i,j,s}$, we have the $1-\alpha/2$ critical point $\overline{\Theta}_{i,j,s}(1-\alpha/2)$. Eliminating the lowest and highest $\alpha/2$ percent of the samples gives a $(1-\alpha)\%$ percentile-based confidence interval defined by

$$CI_{\alpha,F} = \Big[\underline{\Theta}_{i,j,s}(\alpha/2),\ \ \overline{\Theta}_{i,j,s}(1-\alpha/2)\Big].$$

The second approach is based on the properties of a particular solution in repeated samples. We consider the "max-C" solution discussed above. For replication r with K_r solutions in $\bar{\mathcal{B}}^r(\phi, F)$, the "max-C" solution maximizes the collective component correlations as defined in (7). Let $\breve{\Theta}^r_{i,j,s}$ be the dynamic response of variable i to shock j at horizon s associated with the "max-C" solution. Note that the same "max-C" solution is used to evaluate the dynamic responses at all (i, j, s). The critical points associated with the quantiles of $\breve{\Theta}^r_{i,j,s}$ define the $(1-\alpha)\%$ confidence interval

$$CI^{\text{max-C}}_{\alpha} = \Big[\breve{\Theta}_{i,j,s}(\alpha/2),\ \ \breve{\Theta}_{i,j,s}(1-\alpha/2)\Big].$$

Since the $CI_{\alpha,F}$ interval is formed from the tails of the distribution of solutions, it is conservative and can be expected to be wider than $CI^{\text{max-C}}_{\alpha}$.

Using winnowing constraints to reduce the number of plausible solutions is an idea that first appeared in LMN, but more is to be learned about its sampling error and implementation. In particular, whether to use S_t as an external variable or in an augmented system? To this end, we design Monte Carlo experiments built around the SVAR considered in LMN. In brief, the goal of LMN is to understand whether uncertainty is a cause or consequence (or both) of real economic fluctuations, taking into account that there are two distinct types of uncertainty: financial and macro uncertainty. Let $\mathbf{X}_t = (U_{Mt}, Y_t, U_{Ft})'$, where U_{Mt} denotes macro uncertainty, Y_t denotes a measure of real activity, and U_{Ft} denotes financial uncertainty. Although LMN use several different measures of Y_t, the simulations here are based on their base-case with Y_t equal to the log of industrial production. The goal is therefore to identify three structural shocks, denoted $\mathbf{e}_t = (e_{Mt}, e_{Yt}, e_{Ft})'$ from a reduced form VAR with errors $\boldsymbol{\eta}_t = (\eta_{Mt}, \eta_{Yt}, \eta_{Ft})'$, where $\boldsymbol{\eta}_t = \mathbf{B}\mathbf{e}_t$. Unlike the oil market application, no sign restrictions are imposed. Instead, there are now two external variables S_t available for analysis. LMN argue that both theory and empirics suggest that stock market returns are reasonable candidates for S_t, where the maintained assumption is that aggregate stock market returns should be correlated with uncertainty shocks. Specifically, one latent component Z_{2t} of stock returns is presumed to be correlated with both types of uncertainty shocks, and a separate latent component Z_{1t} is presumed correlated with just the financial uncertainty shocks. These assumptions suggest the following representation for a two-dimensional $S_t = (S_{1t}, S_{2t})'$:

$$S_{1t} = \gamma_{10} + \gamma_{1Y} e_{Yt} + \gamma_{1S}(L) S_{1t-1} + Z_{1t} \tag{10a}$$

$$S_{2t} = \gamma_{20} + \gamma_{2M} e_{Mt} + \gamma_{2Y} e_{Yt} + \gamma_{2S}(L) S_{2t-1} + Z_{2t}. \tag{10b}$$

The component Z_{1t} in equation (10a) is orthogonal to e_{Yt}, while the component Z_{2t} in equation (10b) is orthogonal to e_{Yt} and e_{Mt}. As before, the constructed components are denoted $\mathbf{Z}(\mathbf{Q}) \equiv (Z_1(\mathbf{Q}), Z_2(\mathbf{Q}))$ to emphasize that they are functions of parameters $\mathbf{Q}$ to be estimated. If $\mathbf{Z}_t$ were observed, then the relevance conditions $E[Z_{1t}e_{Mt}] \neq 0$, $E[Z_{1t}e_{Ft}] \neq 0$, $E[Z_{2t}e_{Ft}] \neq 0$, the exogeneity conditions $E[Z_{1t}e_{Yt}] = 0$, $E[Z_{2t}e_{Yt}] = 0$, $E[Z_{2t}e_{Mt}] = 0$, along with the covariance structure of $\mathbf{X}_t$, are sufficient to permit point identification of $\mathbf{B}$. But $\mathbf{Z}_t$ is not observed, and the exogeneity conditions hold by construction irrespective of $\mathbf{Q}$. Hence when we observe $\mathbf{S}_t$ instead of $\mathbf{Z}_t$, we no longer have point identification. Two sets of constraints are used to shrink the number of possible solutions.

1 **Uncertainty event constraints** $FE(\mathbf{Q}; \phi, \mathbf{X}, \tau^*, \bar{k})$: For $\mathbf{e}_t(Q) = \mathbf{B}^{-1}\boldsymbol{\eta}_t$,

   i $e_{Ft_1}(\mathbf{Q}) > \bar{k}_1$, $t_1 = 1987{:}10$;

   ii $e_{Ft_2}(\mathbf{Q}) > \bar{k}_2$, $t_2 \in \tau_2^* = [2007{:}12, 2009{:}06]$;

   iii $e_{Yt_2}(\mathbf{Q}) < \bar{k}_3$, $t_2 \in \tau_2^* = [2007{:}12, 2009{:}06]$.

2 **Uncertainty correlation constraints** $FC(\phi, \mathbf{Q}, \mathbf{X}, \mathbf{S}, \bar{c})$

   Let $c_{kj}(\mathbf{Q}) = \text{corr}(Z_{kt}(\mathbf{Q}), e_{jt}(\mathbf{Q}))$ be the sample correlation between the k-th column of $\mathbf{Z}(\mathbf{Q})$ and the j-th column of $\mathbf{e}(\mathbf{Q})$. Then

   i $|c_{1M}(\mathbf{Q})| > \bar{c}$, $|c_{1F}(\mathbf{Q})| > \bar{c}$, and $|c_{2F}(\mathbf{Q})| > \bar{c}$.

   ii For $c(\mathbf{Q}) = \left(c_{1M}(\mathbf{Q}), c_{1F}(\mathbf{Q}), c_{2F}(\mathbf{Q})\right)'$, $\sqrt{c(\mathbf{Q})'c(\mathbf{Q})} > \bar{C}$.

The event constraints ensure that the identified shocks at a small number of extraordinary events are not counterintuitive. In this application, the events are the 1987 stock market crash (Black Monday) and the τ_2^* dates are set in accordance with NBER dating of the Great Recession, which coincides with the timing of the financial crisis. In particular, the constraints imply that the financial uncertainty shocks identified in October 1987 and during the 2007-2009 financial crisis must be large and positive, and the identified output shocks during the Great Recession must not take on unusually large positive values. The correlation constraints are imposed so that one stock return component $Z_{1t}(\mathbf{Q})$ is sufficiently highly correlated (in absolute terms) with each type of uncertainty shock, e_{Mt} and e_{Ft}, and another component $Z_{2t}(\mathbf{Q})$ is sufficiently highly correlated with e_{Ft}. In addition, one set of restrictions, (i), is applied to the individual correlations and a second, (ii), to a measure of the collective correlation between the stock return components and the uncertainty shocks. The bounds are set to $\bar{c} = 0.03$, $\bar{C} = 0.24$, $\overline{k}_1 = 4.0$, $\overline{k}_2 = 4.0$, and $\overline{k}_3 = 2$ as in LMN.

4.1 A Three Variable SVAR with External Variable S_t

In this subsection, we use a Monte Carlo type experiment to assess the robustness of our winnowing restrictions when S_t is synthetic proxy variable external to the three variable SVAR. In the simulation exercise, the true data generating process (DGP) is calibrated to one particular solution, the "max-C" solution that has the highest value for $\sqrt{c(\mathbf{Q})'c(\mathbf{Q})}$.

To generate samples from this solution in a way that ensures the events that appear in historical data also occur in our simulated sample, we draw randomly with replacement from the sample estimates of the shocks $\mathbf{e}_t^{\text{max-C}}(\mathbf{Q})$ for the "max-C" solution, with the exception that we fix the values for these shocks in each replication in the periods t_1 and t_2 to be the observed ones, where t_1 is the period 1987:10 of the stock market crash and $t_2 \in [2007{:}12, 2009{:}06]$. Each draw of the shocks $\mathbf{e}_t^{\text{max-C}}(\mathbf{Q})$ is combined with the max-C estimates of the parameters in $\mathbf{A}(L)$ and $\mathbf{B}$ to generate $R = 5,000$ samples of $\mathbf{X}_t$ using the SVAR $\mathbf{A}(L)\mathbf{X}_t = \mathbf{B}\mathbf{e}_t^{\text{max-C}}$. Each sample of of $\mathbf{X}_t$ is length T, as in the data used in the estimation.

Samples of data for the idiosyncratic shocks e_{S_1t} and e_{S_2t} are generated from

$$e_{S_1t} = S_{1t} - d_{1S}(L)S_{1t-1} - d_1'\mathbf{e}_t^{\text{max-C}} \quad (11)$$

$$e_{S_2t} = S_{2t} - d_{2S}(L)S_{2t-1} - d_2'\mathbf{e}_t^{\text{max-C}} \quad (12)$$

where all d parameters in (11) and (12) are calibrated to target $c_1 = -0.035$, $c_2 = -0.185$, and $c_3 = -0.163$. These are the correlations between Z_{1t} and e_{Mt}, Z_{1t} and e_{Ft}, and Z_{2t} and e_{Ft}, respectively, corresponding to the max-C solution estimated in LMN. We generate 5,000 samples of S_1 and S_2 by drawing with replacement from these e_{S1} and e_{S2} vectors, in the same manner described above for $\mathbf{e}_t$, and recursively iterating on (11) and (12) using the first observations on S_1 and S_2 in our historical sample as initial values.[5]

For each of these R=5,000 replications, we construct a set of winnowed solutions $\bar{\mathcal{B}}(\phi|F)$. In each replication r, K =1,000 possible solutions for $\mathbf{B}$ are generated by initializing $\mathbf{B}$ to be the lower Cholesky factorization of $\mathbf{\Omega}$ for an arbitrary ordering of the variables. These are then rotated by 1,000 random orthogonal matrices $\mathbf{Q}$. The number of random rotations for

[5] In the data used by LMN, S_1 is the return on the S&P 500 stock index in excess of the one-month Treasury bill rate, and S_2 is the excess return portfolio-weighted-average of the CRSP value-weighted return and a small-stock index.

this simulation exercise is considerably smaller than what is in general possible in historical applications, in order to make feasible the large number of replications.

There are three variables in the model and therefore nine impulse responses to consider. These are shown in Figure 6. The widest areas (white shading) show CI_α with $\alpha = 10$ when no winnowing constraints at all are imposed, i.e., when the event and correlation constraints are *not* imposed. The dashed line shows the "true" IRFs, which are the ones estimated from the historical data for the max-C solution. While the true IRF is in the CI_α interval, this interval is too wide to be informative. This is because, absent additional constraints, the model is under-identified. The remaining two shaded areas are the intervals of IRFs when the winnowing constraints are imposed. Shaded in medium grey are the $CI_{\alpha,F}$ intervals, again with $\alpha = 10$. Shaded in darkest grey are the $CI_\alpha^{\text{max-C}}$ intervals. Both confidence intervals are noticeably narrower compared to the unconstrained confidence intervals, indicating that the winnowing constraints contribute to the identification. In all nine cases, the "true" IRF is inside the corresponding $CI_{\alpha,F}$. Several economic insights can be gleaned from these intervals that are well determined by the data. The effect of a positive financial uncertainty shock on output is decisively negative. The short-run effect of a positive output shock on macro uncertainty is negative, but the effect on financial uncertainty is positive. A positive macro uncertainty shock increases output with little effect on financial uncertainty.

4.2 A Four Variable System

In this subsection, we carry out a simulation for a four variable SVAR that includes S_t, and which we refer to as the *full system approach* to distinguish it from the three variable SVAR in the previous subsection. The systems approach only restricts e_{St} to be contemporaneously uncorrelated with $\mathbf{X}_t$ which is less restrictive than Granger non-causality. In other words, lags of S_t can be predict X_t in the systems framework. This means in particular that the the dynamic effects of shocks to $\mathbf{X}_t$ on $\mathbf{S}_t$ are allowed to feedback to $\mathbf{X}_t$, a channel that is missing if $\mathbf{S}_t$ is treated as an external proxy.

To simplify the analysis, we let S_t be a scalar. The reduced form errors are collected into the 4×1 vector $\boldsymbol{\eta}_t = (\eta'_{Xt}, \eta_{St})'$. The structural shocks are $\begin{pmatrix}\mathbf{e}'_{Xt} & e_{St}\end{pmatrix}'$ with $\boldsymbol{\eta}_t = \mathbf{B}\mathbf{e}_t$. Under the assumption that $E[\mathbf{X}_t e_{St}] = 0$, the impact sub-vector $\mathbf{B}_{XS} = (B_{MS}, B_{YS}, B_{FS})'$ is therefore zero. These three zero restrictions imply

$$\begin{pmatrix} \eta_{Mt} \\ \eta_{Yt} \\ \eta_{Ft} \\ \eta_{St} \end{pmatrix} = \begin{pmatrix} B_{MM} & B_{MY} & B_{MF} & 0 \\ B_{YM} & B_{YY} & B_{YF} & 0 \\ B_{FM} & B_{FY} & B_{FF} & 0 \\ B_{SM} & B_{SY} & B_{SF} & B_{SS} \end{pmatrix} \begin{pmatrix} e_{Mt} \\ e_{Yt} \\ e_{Ft} \\ e_{St} \end{pmatrix}. \tag{13}$$

The synthetic variables $\mathbf{Z}_t$ are now implicitly defined as

$$Z_{1t} = \eta_{St} - B_{SY}e_{Yt} = B_{SM}e_{Mt} + B_{SF}e_{Ft} + B_{SS}e_{St}$$
$$Z_{2t} = Z_{1t} - B_{SM}e_{Mt} = B_{SF}e_{Ft} + B_{SS}e_{St}.$$

The winnowing constraints used to shrink the set of admissible solutions are the same as above, except that the residual η_{St} is used to construct $\mathbf{Z}_t$ in place of S_t, and the structural parameter matrix $\mathbf{B}$ is now 4×4, as shown in (13).

As in simulation for the three variable SVAR above, we draw $\mathbf{e}_X$ from the empirical distribution of shocks corresponding to the max-C solution of system estimation. In order to account for the big shock events, we again fix the $\mathbf{e}_X$ shocks in each replication in the periods t_1 and t_2 to be the observed ones where t_1 is the period 1987:10 of the stock market crash and $t_2 \in$ [2007:12, 2009:06]. The winnowing constraints are the same as those in non-system simulation.

The results for the systems approach are reported in Figure 7. The bounded white areas are the CI_α confidence intervals when no event or correlation constraint is imposed with $\alpha = 10$. As above, though the 'true' IRF (dashed line) is inside the interval, no insight can be gained from the unconstrained estimation because the confidence intervals are too wide. Next, we turn to the two confidence intervals after the constraints are imposed. Several observations are of note. First, the $CI_{10}^{\text{max-C}}$ confidence intervals shaded in dark grey are always inside the $CI_{10,F}$ intervals shaded in medium grey, and both are noticeably narrower than unconstrained solutions shaded in white. Second, the $CI_{10}^{\text{max-C}}$ and $CI_{10,F}$ intervals are both very similar to the ones reported in Figure 6 for the three variable SVAR when S_t was an external variable. While the widths of the $CI_{10}^{\text{max-C}}$ bands are similar in the two figures, the ones produced by systems estimation for $CI_{10,F}$ are tighter, especially for the dynamic response to a macro uncertainty shocks reported in the first row.

A final result of interest is whether this particular SVAR is compatible with a recursive structure. Recursivity is a convenient assumption and is often used in the empirical literature on the effects of uncertainty shocks. The identifying restrictions employed here impose no such structure, and it is straight forward to check whether the estimated values of B are consistent with it. Figure 8 plots the distribution over 5,000 replications of $\hat{B}_{YM}, \hat{B}_{MY}, \hat{B}_{YF}$, and $\hat{B}_{MF}$ based on the "max-C" solution. None of the distributions are centered around zero. The implication is that the recursive structure is inconsistent with any ordering of the variables.

5 Conclusion

Identifying assumptions need to be imposed in order to give impulse responses generated by autoregressions an economically meaningful interpretation. But our information in the form of commonly used zero and sign restrictions are often not rich enough to conclude that the data are consistent with a clear causal pattern among the variables. This paper explores the properties of two new types of identifying restrictions as moment inequalities that can help constrain the number of plausible solutions. The first restricts the sign and magnitude of identified shocks to accord with an historical understanding of events at particular points in the sample, while the second restricts the correlations between the identified shocks and components of external variables. We use applications and simulations to show how these constraints can be used. The issue of how to best conduct frequentist inference in set-identified SVARs remains an important topic for future research.

References

Antolin-Diaz, J., and J. Rubio Ramírez (2016): "Narrative Sign Restrictions for SVARs," Unpublished paper, Emory University.

Baumeister, C., and L. Kilian (2016): "Forty years of oil price fluctuations: Why the price of oil may still surprise us," *The Journal of Economic Perspectives*, 30(1), 139–160.

Conley, T. G., C. B. Hansen, and P. E. Rossi (2012): "Plausibly Exogenous," *The Review of Economics and Statistics*, 94(1), 260–272.

Fry, R., and A. Pagan (1911): "Sign restrictions in structural vector autoregressions: A critical review," *Journal of Economic Literature*, 49(4), 938–960.

Gafarov, B., M. Meier, and J. M. Olea (2015): "Projection Inference for Set-Identified SVARs," Unpublished paper, New York University.

Hamilton, J. D. (2013): "Historical Oil Shocks,," in *Routledge Handbook of Major Events in Economic History*, ed. by R. E. Parker, and R. Whaples, pp. 49–56. Routledge Taylor and Francis Group, New York.

Kilian, L. (2008): "Exogenous oil supply shocks: how big are they and how much do they matter for the US economy?," *The Review of Economics and Statistics*, 90(2), 216–240.

——— (2009): "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review*, 99(3), 1053–1069.

Kilian, L., and H. Lutkepohl (2016): *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Kilian, L., and D. P. Murphy (2012): "Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models," *Journal of the European Economic Association*, 10(5), 1166–1188.

Ludvigson, S. C., S. Ma, and S. Ng (2015): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?," http://www.econ.nyu.edu/user/ludvigsons/ucc.pdf.

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

——— (2014): "A reconciliation of svar and narrative estimates of tax multipliers," *Journal of Monetary Economics*, 68, S1–S19.

Moon, H. R., and F. Schorfheide (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80(2), 755–782.

Moon, H. R., F. Schorfheide, and E. Granziera (2013): "Inference for VARs Identified with Sign Restrictions," Unpublished Paper.

Olea, J. L. M., J. H. Stock, and M. D. Watson (2015): "Inference in Structural VARs with External Instruments," Unpubished manuscript, Harvard University.

Ramey, V. A. (2016): "Macroeconomic shocks and their propagation," NBER Working paper No. 21978.

Rubio-Ramirez, J. F., D. F. Waggoner, and T. Zha (2010): "Structural vector autoregressions: Theory of identification and algorithms for inference," *The Review of Economic Studies*, 77(2), 665–696.

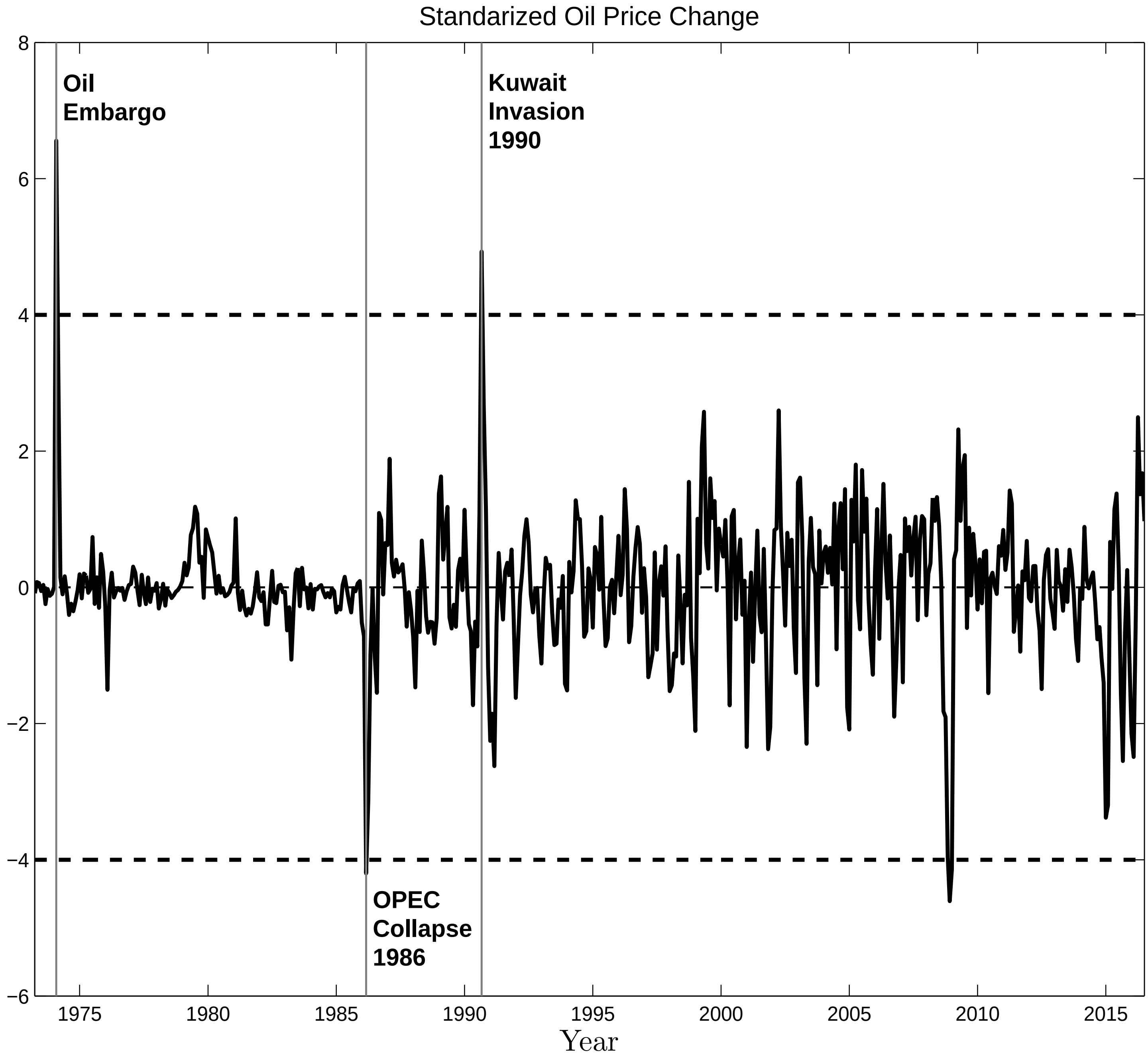
Sims, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, pp. 1–48.

Stock, J. H. (2008): "What's New in Econometrics: Time Series," *NBER Summer Institute.*

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52(2), 381–419.

Figures and Tables

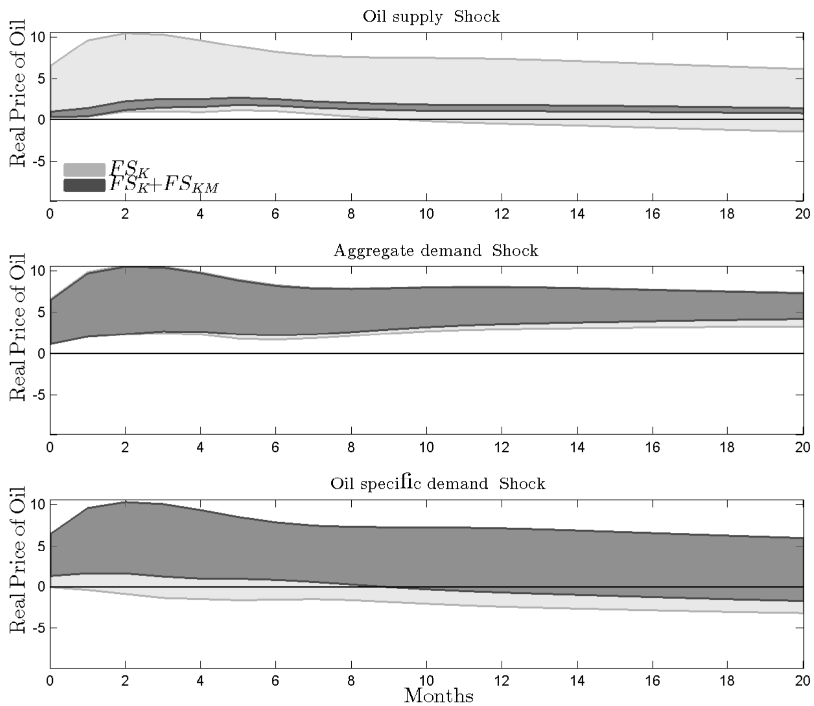
Figure 1: Oil Price Change



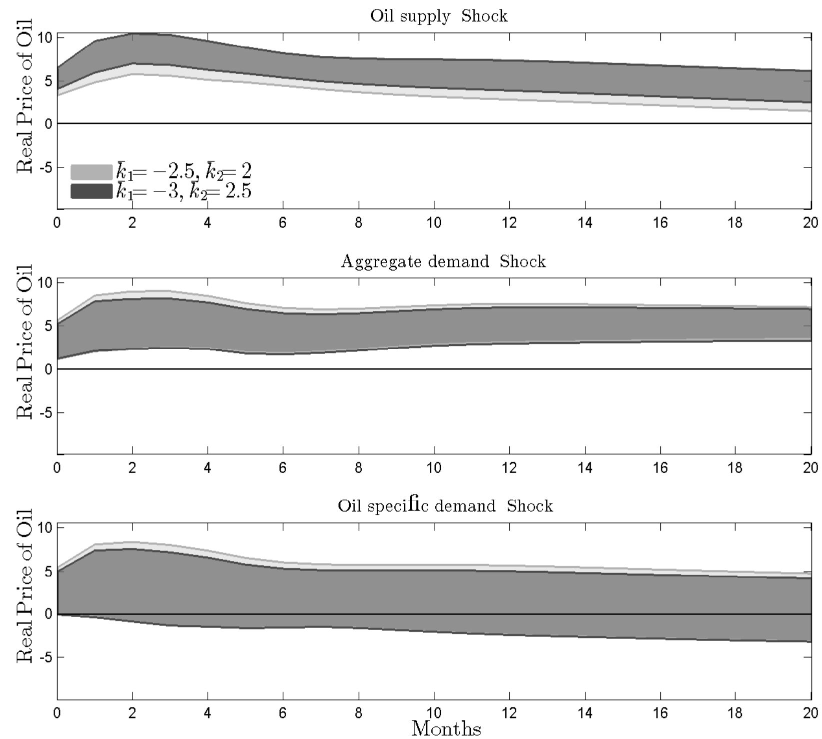
The standardized change in the real oil price change is reported. The sample spans the period 1973:01 to 2016:06.

Figure 2: Oil Shock Application 1

$\text{FS}_K + \text{FS}_{KM}$ **Constraints**



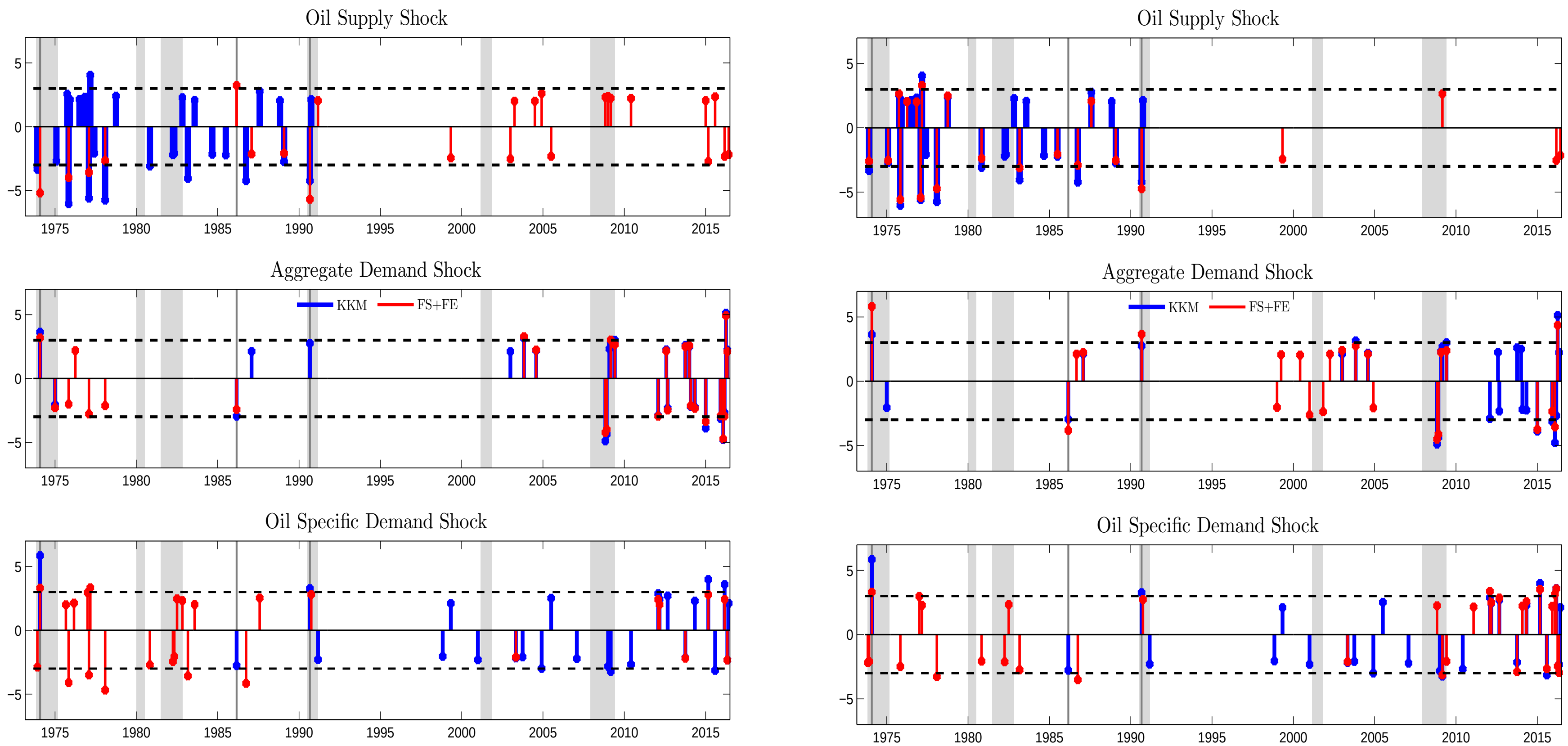
$\text{FS}_K + \text{FE}$ **Constraints**



Shaded areas stack the impulse responses for all admissible solutions based on the type of restriction listed in the subplot heading. The sample spans the period 1973:01-2016:06.

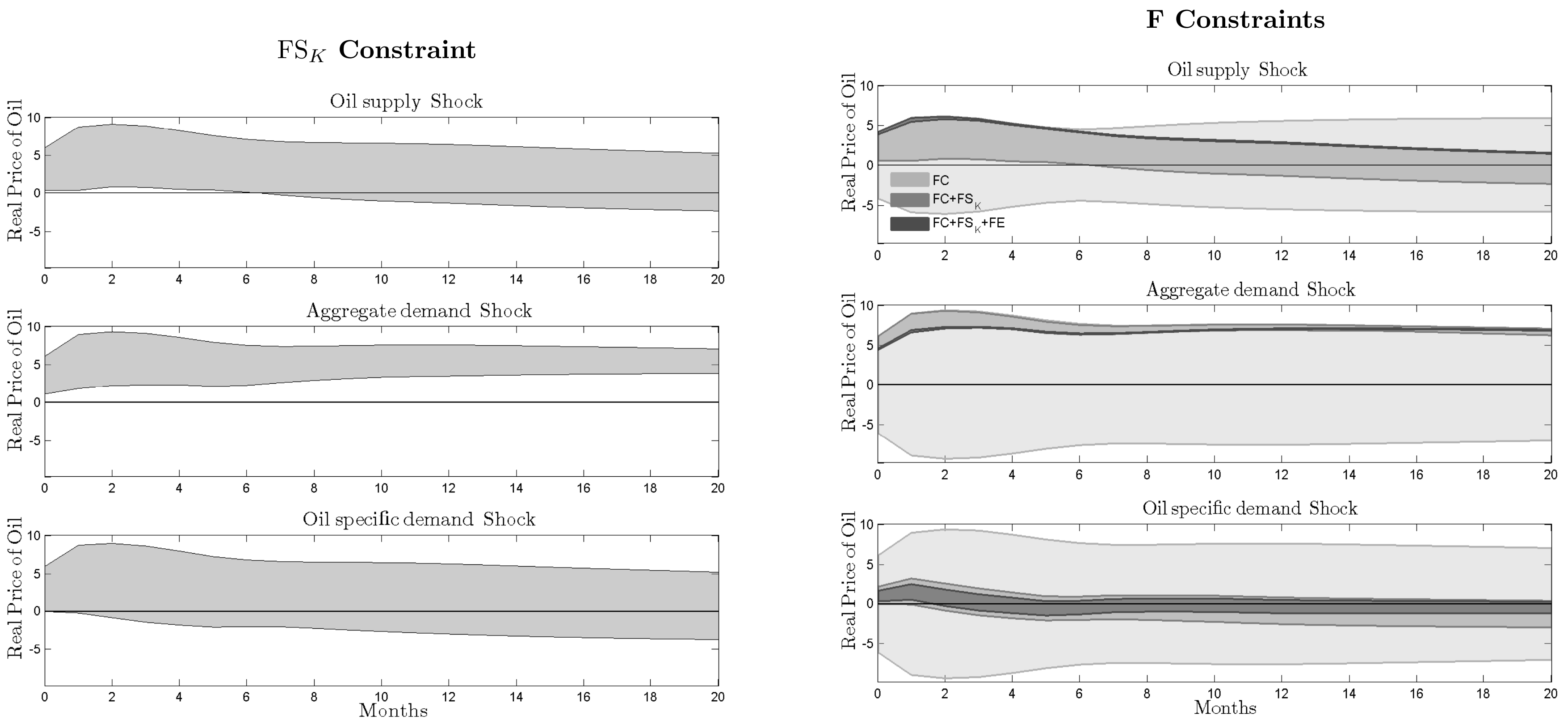
Figure 3: Large Shocks in Oil Application 1

Event Constraints on e_1 | **Event Constraints on e_2 or e_3**



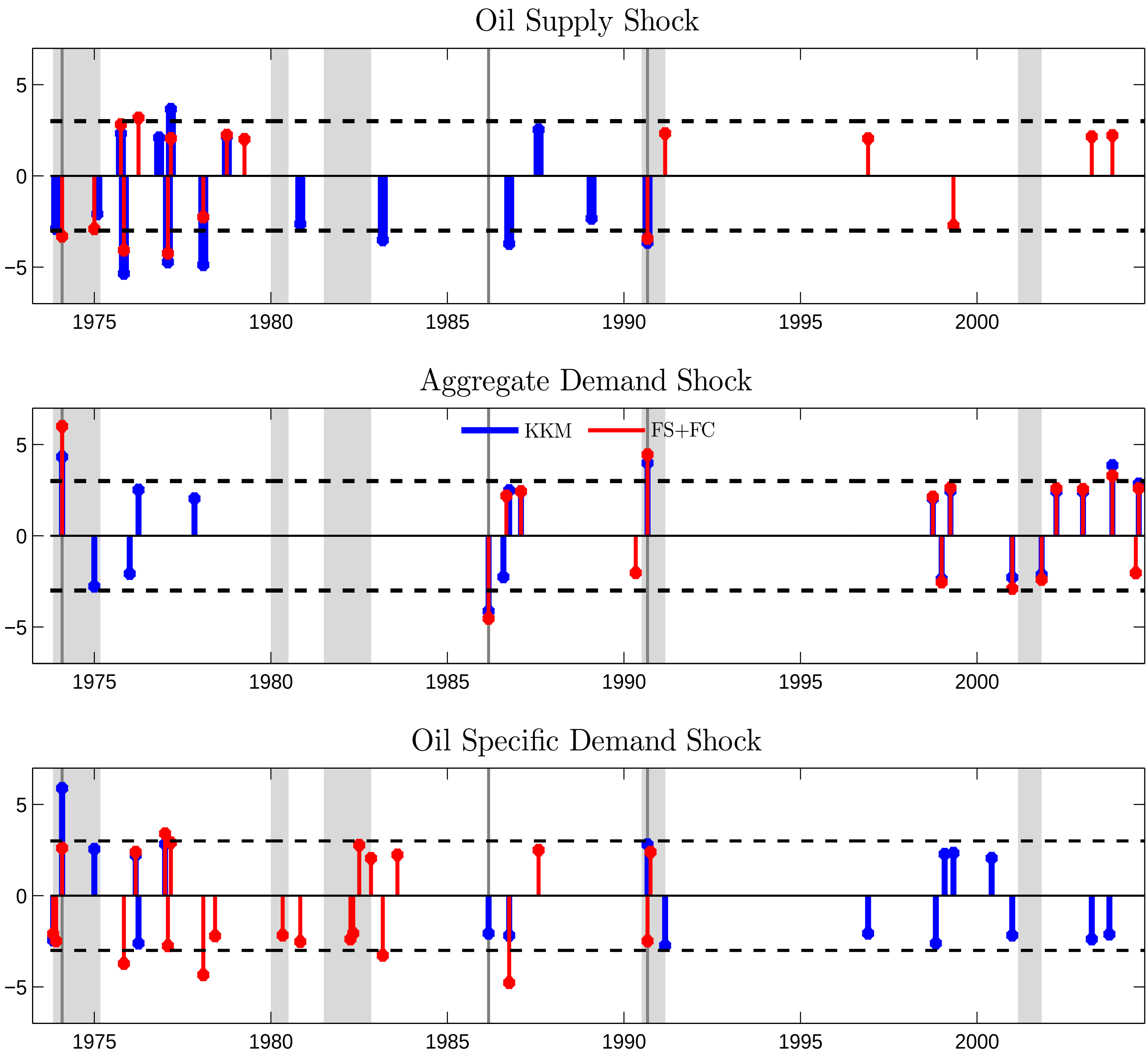
The figure exhibits shocks that are at least 2 standard deviations above/below the unconditional mean for e. The horizontal line corresponds to 3 standard deviations shocks. The solid vertical lines are the event dates 1974:01, 1986:02, and 1990:08 discussed in the text. The light grey shaded areas give the dates of NBER recessions. The sample spans the period 1973:02 to 2004:09.

Figure 4: Oil Shock Application 2



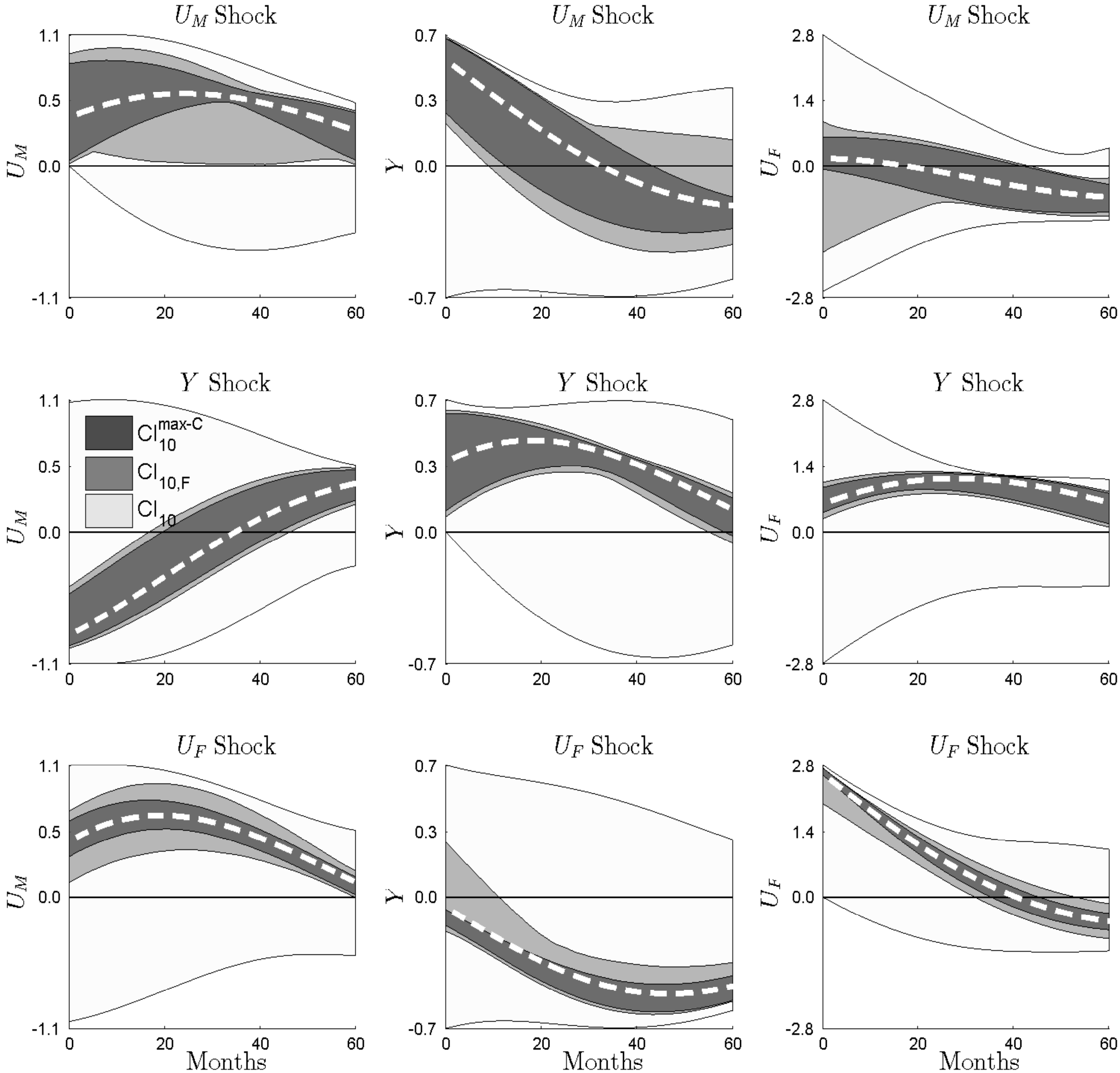
Shaded areas stack the impulse responses for all admissible solutions based on the type of restriction listed in the subplot heading. FC constraint is based on $\bar{c}$=0.03. The sample spans the period 1973:02-2004:09.

Figure 5: Large Shocks in Oil Application 2



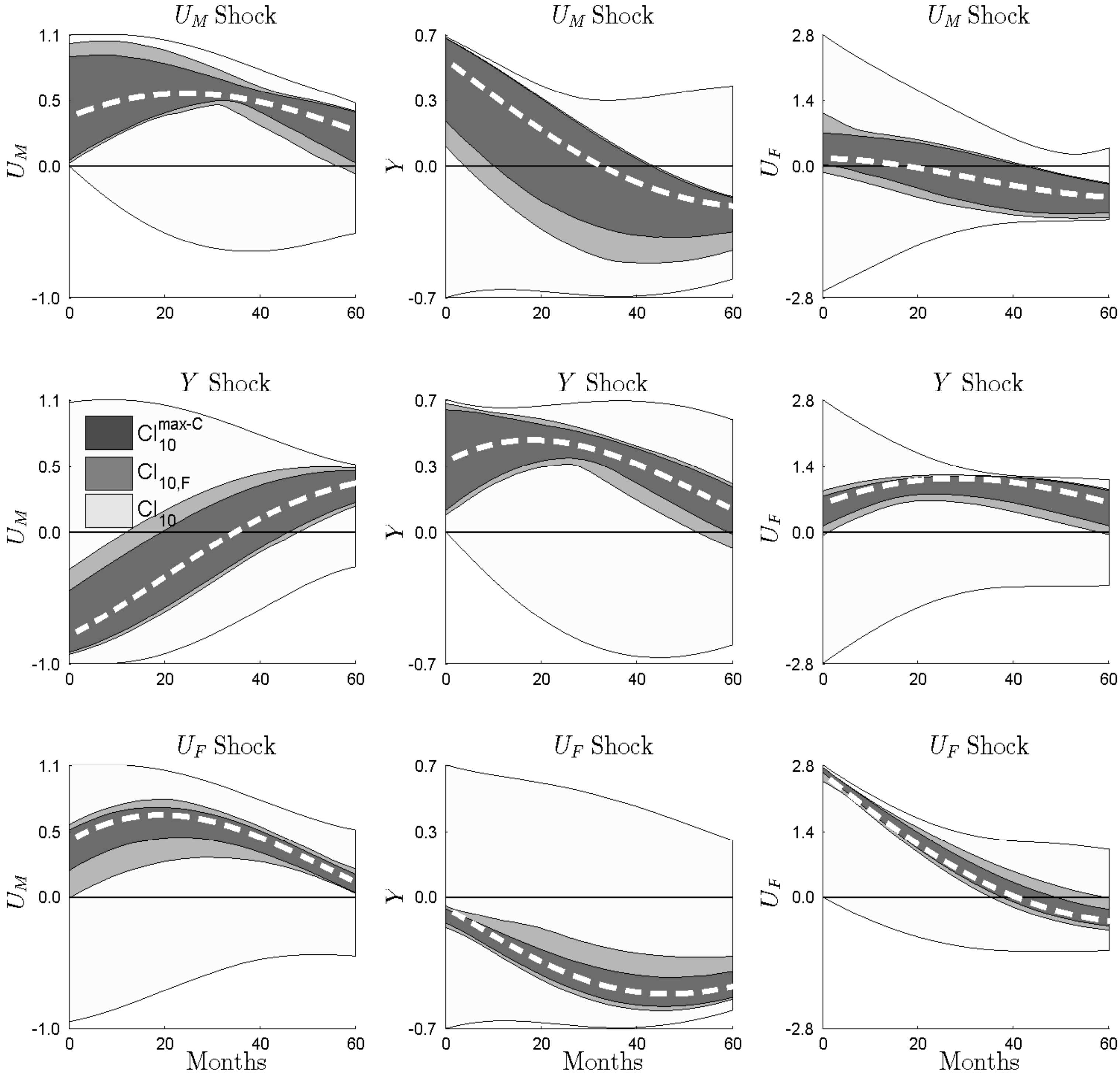
The figure exhibits shocks that are at least 2 standard deviations above/below the unconditional mean for e. The horizontal line corresponds to 3 standard deviations shocks. The solid vertical lines are the event dates 1974:01, 1986:02, and 1990:08 discussed in the text. The light grey shaded areas give the dates of NBER recessions. The sample spans the period 1973:02 to 2004:09.

Figure 6: Monte Carlo Simulation For LMN: External S



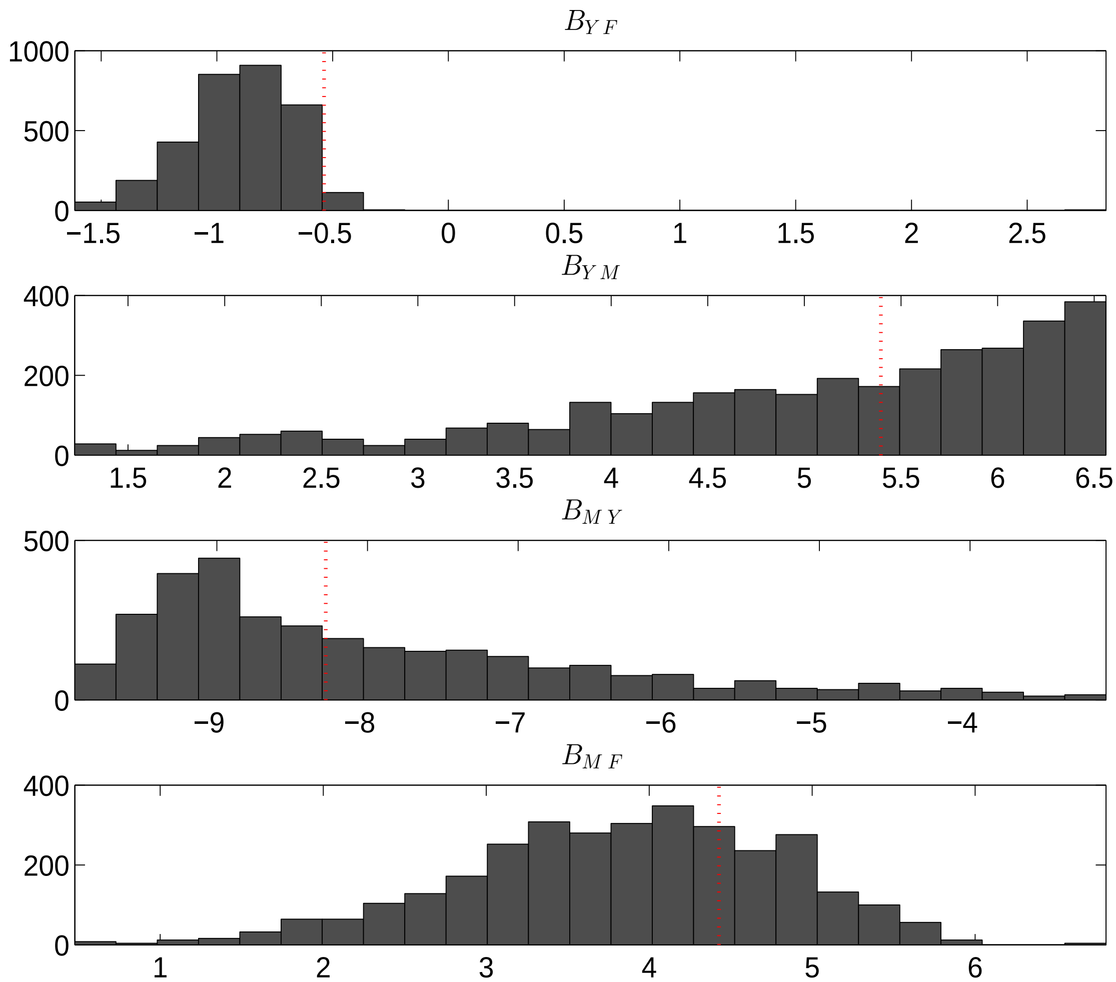
The shaded area reports the 90 percent confidence interval across 5000 replications. Dotted line is the true IRF. $CI_{10}^{\text{max-C}}$ is the CI with max-C solution. $CI_{10,F}$ is the CI with winnowing constraints imposed. CI_{10} is the CI without imposing any constraints. The sample size is $T = 652$ and 1000 random rotations are used for each replication and number of replication is $R = 5000$.

Figure 7: Monte Carlo Simulation For LMN: System Estimation



The shaded area reports the 90 percent confidence interval across 5000 replications. Dotted line is the true IRF. $CI_{10}^{\text{max-C}}$ is the CI with max-C solution. $CI_{10,F}$ is the CI with winnowing constraints imposed. CI_{10} is the CI without imposing any constraints. The sample size is $T = 652$ and 1000 random rotations are used for each replication and number of replication is $R = 5000$.

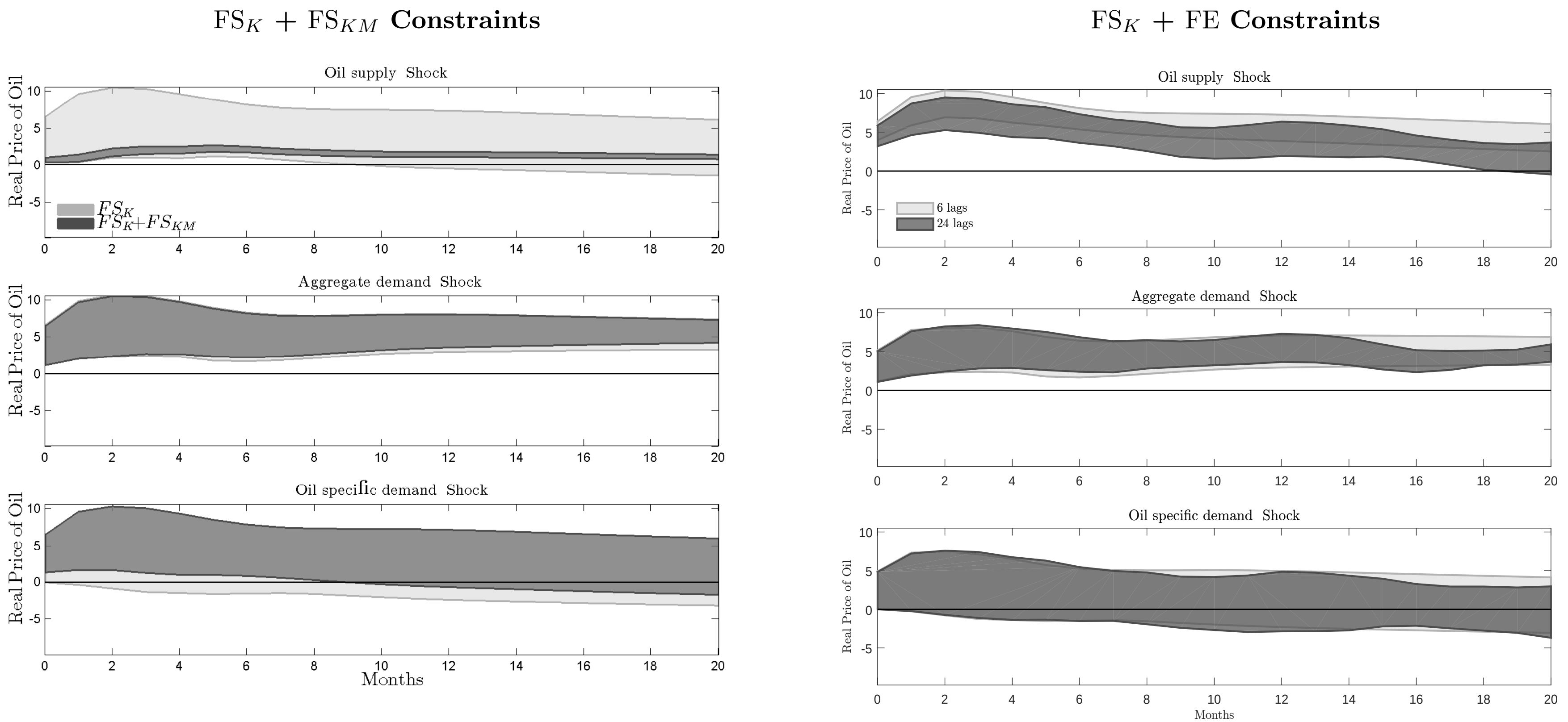
Figure 8: Monte Carlo Simulation For LMN with System Estimation: Density Plot



Histogram of max-C solution over 5000 replications with system estimation. B is reported in the unit of 10^{-3}.

Appendix Figures and Tables

Figure A1: Oil Shock Application 1: Longer Lags



Shaded areas stack the impulse responses for all admissible solutions based on the type of restriction listed in the subplot heading. The sample spans the period 1973:01-2016:06.