Improving Global Vector Autoregressions

Neil R. Ericsson*

January 31, 2011

[Preliminary draft: comments welcome]

Abstract: Global vector autoregressions (GVARs) have several attractive features: a standardized economically appealing choice of variables for each country or region examined, a systematic treatment of long-run properties through cointegration analysis, and flexible dynamic specification through vector error correction modeling. The current paper re-examines the theoretical and empirical underpinnings for GVARs, focusing on exogeneity assumptions, parameter constancy, and data aggregation. This paper proposes refinements in these areas, with the aim of achieving an even more robust approach to GVAR modeling. The GVAR in Dées, di Mauro, Pesaran, and Smith (2007) is used to illustrate these refinements.

Keywords: cointegration, conditional models, data aggregation, error correction, exogeneity, global vector autoregression, GVAR, impulse indicator saturation, marginalization, model design, model selection, parameter constancy, reduction, VAR.

JEL classifications: C52, E41.

*The author is a staff economist in the Division of International Finance, Board of Governors of the Federal Reserve System, Washington, D.C. 20551 U.S.A.; and he may be reached on the Internet at ericsson@frb.gov. The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. The author is grateful to David Hendry, Jesper Lindé, Filippo di Mauro, and Hashem Pesaran for helpful discussions and comments; to Vanessa Smith for invaluable guidance in replicating estimation results from Dées, di Mauro, Pesaran, and Smith (2007); and to Erica Reisman for excellent research assistance. An earlier version of this paper was presented at the workshop “The impact of the crisis on the assessment of the international environment: Lessons for modellers and practitioners”, European Central Bank, Frankfurt am Main, Germany, December 16–17, 2010. All numerical results were obtained using PcGive Version 13.10 and Autometrics Version 1.5e in OxMetrics 6.10: see Doornik and Hendry (2009) and Doornik (2009).
1 Introduction

The recent financial crisis and ensuing Great Recession have highlighted the importance and pervasiveness of international linkages in the world economy—and the importance of capturing those linkages in empirical macroeconomic models that are used for economic analysis, forecasting, and policy analysis. Pesaran, Schuermann, and Weiner (2004) propose and implement global vector autoregressions (or GVARs) as an ingenious approach to capturing international linkages between country- or region-specific error correction models. Dées, di Mauro, Pesaran, and Smith (2007) (hereafter DdPS) extend that work to a larger number of countries and regions; and Pesaran, Schuermann, and Smith (2009) assess the forecasting properties of the GVAR implemented in DdPS.

The GVAR methodology has several attractive features:

- a standardized economically appealing choice of variables (both domestic and foreign) for each country or region,
- a systematic treatment of long-run properties through cointegration analysis, and
- flexible dynamic specification through vector error correction modeling.

The current paper re-examines the theoretical and empirical underpinnings for global vector autoregressions (or GVARs), focusing on the following areas:

- treatment of exogeneity assumptions,
- parameter constancy, and
- data aggregation.

This paper proposes extensions and improvements in these areas, with the aim of achieving an even more robust approach to GVAR modeling.

This paper is organized as follows. Section 2 summarizes the current approach taken to modeling GVARs, as developed in Pesaran, Schuermann, and Weiner (2004) and DdPS inter alia. Section 3 discusses the theory of reduction and exogeneity, thereby providing the background for the analysis in subsequent sections. Sections 4 and 5 respectively examine the implicit data aggregation assumptions in GVARs and the underlying conditioning assumptions. Section 6 summarizes the computer-automated model selection algorithm in Autometrics. That algorithm is central to testing several aspects of GVARs: parameter constancy (through the use of impulse indicator saturation), data aggregation (by being able to consider more potential variables than number of observations), and weak exogeneity. Section 7 empirically implements those classes of tests for DdPS’s GVAR, focusing on the subsystem specifications for the United States, the euro area, the United Kingdom, and China. Section 8 discusses various implications and extensions, and Section 9 concludes.
2 Global Vector Autoregressions

This section briefly describes the current approach taken to modeling GVARs, as developed in Pesaran, Schuermann, and Weiner (2004) and DdPS *inter alia*. For further research on GVARs, see Pesaran and Smith (2006), Dées, Holly, Pesaran, and Smith (2007), Pesaran, Smith, and Smith (2007), Hieberta and Vansteenkiste (2009), Pesaran, Schuermann, and Smith (2009), Castrén, Dées, and Zaher (2010), Chudik and Pesaran (2010), and the comments and rejoinders to Pesaran, Schuermann, and Weiner (2004) and Pesaran, Schuermann, and Smith (2009). Juselius (1992) provides a conceptual precursor to GVARs in her sector-by-sector analysis of the Danish economy to obtain multiple long-run feedbacks entering an equation for domestic inflation.

The GVAR is an ingenious structure for capturing international linkages between country- or region-specific error correction models. The current approach to modeling a GVAR starts with a standardized economically appealing choice of variables (both domestic and foreign) for each country or region, provides a systematic treatment of long-run properties through cointegration analysis, and adopts flexible dynamic specification through vector error correction modeling. These features are very appealing, and they balance naturally the roles of data and economic theory in empirical modeling. The GVAR explicitly aims to capture international economic linkages, especially linkages between the macroeconomic and financial sides of economies. Weak exogeneity plays an important role through allowing conditional subsystem analysis on a country-by-country basis. Data aggregation—empirically implemented but based on economic theory—achieves a high degree of parsimony in the estimated models.

Algebraically, the structure of a GVAR is as follows, drawing on expositions and notation in Pesaran, Schuermann, and Weiner (2004) and DdPS.

The $N + 1$ countries being modeled are indexed by $i$ ($i = 0, 1, 2, \ldots, N$), where country 0 is the reference country (here, the United States), and “country” may in fact refer to a region, such as the euro area. A $k_i \times 1$ vector of country-specific variables $x_{it}$ depends on a $k_i^* \times 1$ vector of foreign variables $x_{it}^*$, where the vector $x_{it}^*$ is specific to country $i$. In particular, $x_{it}^*$ is an aggregation across the variables of the individual foreign countries, where “foreign” is from the perspective of country $i$. Both $x_{it}$ and $x_{it}^*$ can thus be expressed as combinations of the entire set of domestic and foreign variables, which is $(x_{it}^0 : x_{it}^1 : \ldots : x_{it}^N)'.$ Throughout, $t$ and $i$ are the time and country indexes, there are $T$ observations ($t = 1, \ldots, T$) and $N + 1$ countries ($i = 0, 1, 2, \ldots, N$), and an asterisk * denotes “foreign”.

The variables $(x_{it}^0 : x_{it}^*)'$, which are associated with country $i$, are assumed to be driven by a finite-order vector autoregression (VAR). That VAR is interpretable as a joint distribution for $(x_{it}^0 : x_{it}^*)'$, conditional on the lags of those variables. It is convenient to factorize that joint distribution into a conditional distribution for the domestic variables $x_{it}$ (i.e., conditional on $x_{it}^*$) and a marginal distribution for
the foreign aggregates $x^*_it$. Thus, the specific relationship between $x_{it}$ and $x^*_it$ is a subsystem autoregressive distributed lag, equivalently known as an augmented VAR model, denoted VARX*(p1, q1).

\[ \Phi_i(L, p_i)x_{it} = a_{i0} + a_{i1}t + \Upsilon_i(L, q_i)d_t + \Lambda_i(L, q_i)x^*_it + u_{it}, \]  

where $a_{i0}$ and $a_{i1}$ are the intercept and trend coefficients; $\Phi_i(L, p_i)$, $\Upsilon_i(L, q_i)$, and $\Lambda_i(L, q_i)$ are suitably conformable matrix polynomials in the lag operator $L$, of orders $p_i$, $q_i$, and $q_i$ respectively; $d_t$ is a vector of global variables such as oil prices; and $u_{it}$ is a $k_i \times 1$ vector of idiosyncratic shocks.

Equation (1) is a conditional equation—of $x_{it}$ conditional on its own lags and on current and lagged values of $d_t$ and $x^*_it$. The corresponding marginal equation for $x^*_it$ can be written as:

\[ \Lambda^*_i(L, q_i)x^*_it = a^*_{i0} + a^*_{i1}t + \Upsilon^*_i(L, q_i)d_t + \Phi^*_i(L, p_i)x_{it} + u^*_{it}, \]  

where $a^*_{i0}$ and $a^*_{i1}$ are the intercept and trend coefficients; $\Lambda^*_i(L, q_i)$, $\Upsilon^*_i(L, q_i)$, and $\Phi^*_i(L, p_i)$ are suitably conformable matrix polynomials in $L$, of orders $q_i$, $q_i$, and $p_i$ respectively; and $u^*_{it}$ is a $k_i^* \times 1$ vector of idiosyncratic shocks. Because equation (2) is the marginal equation that pairs with the conditional equation (1), by construction the current-dated domestic variables ($x_{it}$) do not enter equation (2). That is, the zeroth-order matrix coefficient in the polynomial $\Phi^*_i(L, p_i)$ is itself zero ($\Phi^*_{i0} = 0$, say). Additionally, $u_{it}$ and $u^*_{it}$ are uncorrelated with each other because equations (1) and (2) are a conditional-marginal factorization.

The variables $x_{it}$, $d_t$, and $x^*_it$ may be integrated. If they are integrated, they may also be cointegrated, so it can be useful to rewrite equations (1) and (2) in their error correction form:

\[ \Theta_i(L, p_i - 1)\Delta x_{it} = a_{i0} + \Psi_i(L, q_i - 1)\Delta d_t + \Gamma_i(L, q_i - 1)\Delta x^*_it 

+ \alpha_i\beta_i^t(x^*_it - x^*_it_{-1} : \Delta d_{it} - 1 : \Delta x^*_it_{-1} : t) + u_{it}, \]  

and

\[ \Gamma^*_i(L, q_i - 1)\Delta x^*_it = a^*_{i0} + \Psi^*_i(L, q_i - 1)\Delta d_t + \Theta^*_i(L, p_i - 1)\Delta x^*_it 

+ \alpha^*\beta^t_i(x^*_it - x^*_it_{-1} : \Delta d_{it} - 1 : \Delta x^*_it_{-1} : t) + u^*_{it}. \]  

The cointegrating vectors are $\beta_i$; the feedback coefficients are $(\alpha_i^t : \alpha^*_i)^t$; and the polynomials $\Theta_i(L, p_i - 1)$, $\Psi_i(L, q_i - 1)$, $\Gamma_i(L, q_i - 1)$, $\Gamma^*_i(L, q_i - 1)$, $\Psi^*_i(L, q_i - 1)$, and $\Theta^*_i(L, p_i - 1)$ are the transformations of $\Phi_i(L, p_i)$, $\Upsilon_i(L, q_i)$, $\Lambda_i(L, q_i)$, $\Lambda^*_i(L, q_i)$, $\Upsilon^*_i(L, q_i)$, and $\Phi^*_i(L, p_i)$ that arise from transforming the VAR represented by equations (1) and (2) into the vector error correction model in equations (3) and (4).
In this framework, if $\alpha_i^* = 0$ in equation (4), then the foreign variables $x_{it}^*$ are weakly exogenous for the cointegrating vectors $\beta_i^*$ in equation (3). DdPS assume weak exogeneity of $x_{it}^*$. That weak exogeneity allows them to focus on the conditional error correction model (3) and ignore the marginal error correction model (4). DdPS (Table IV) also test for weak exogeneity by testing $\alpha_i^* = 0$, and DdPS find that that restriction is satisfied for almost all equations across a central subset of countries.

The GVAR itself is constructed from the conditional error correction model (3). Specifically, the conditional error correction models for all of the countries are stacked as a single vector error correction model, which is the GVAR. That GVAR has many potential uses, such as private-firm policy regarding risk, banking supervision and regulation, central bank policy, and forecasting; cf. Pesaran, Schuermann, and Weiner (2004), DdPS, and Pesaran, Schuermann, and Smith (2009). In some of these situations, strong exogeneity, super exogeneity, or both may be required for valid analysis; see Ericsson, Hendry, and Mizon (1998) and Section 3 below.

To provide a sense of some of the empirical aspects involved, consider the GVAR in DdPS. The set of domestic variables $x_{it}$ is as follows (with a few exceptions, as noted in DdPS).

Table 1: Variables in DdPS’s GVAR.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_{it}$</td>
<td>Real GDP</td>
</tr>
<tr>
<td>$\Delta p_{it}$</td>
<td>CPI inflation</td>
</tr>
<tr>
<td>$q_{it}$</td>
<td>Real equity prices</td>
</tr>
<tr>
<td>$e p_{it}$</td>
<td>Real exchange rate</td>
</tr>
<tr>
<td>$r_{it}$</td>
<td>Short-term interest rate</td>
</tr>
<tr>
<td>$l r_{it}$</td>
<td>Long-term interest rate</td>
</tr>
</tbody>
</table>

The country-specific aggregated foreign variables ($x_{it}^*$) are then constructed from the full set of domestic variables across all countries, using trade weights. The VARX$^*(p_i, q_i)$ for each country—as given in equation (1)—is initially a VARX$^*(2, 2)$. In some instances, however, shorter lags are used, based on standard information criteria. Cointegration in the VARX$^*$ is then tested, following the procedure in Harbo, Johansen, Nielsen, and Rahbek (1998) and using critical values from MacKinnon, Haug, and Michelis (1999). The number of cointegrating vectors may differ from country to country. In the conditional error correction model (3), the country’s cointegrating vectors are written in their reduced form, i.e., with $\beta_i^*$ beginning with an identity matrix. The data are quarterly, taken from the IMF’s International Financial Statistics (except for Singapore’s, which are from Datastream); and estimation

---

1 As discussed below, a cut is also required for weak exogeneity.
is typically over 1979Q4–2003Q4 ($T = 97$). This GVAR from DdPS provides the empirical illustration in Section 7.

In order to assess the current approach to GVARs and offer some directions for improvement, the next two sections digress to consider the theory of reduction and exogeneity.

3 The Theory of Reduction

This section examines the relationship between the data generation process and an empirical model, linking the two through the theory of reduction. This relationship is central to understanding the roles of data aggregation and exogeneity in a GVAR. Section 3.1 defines what the data generation process is, Section 3.2 shows how an empirical model is obtained as a reduction or simplification of the data generation process, and Section 3.3 specifies the implied derived empirical model.


Before proceeding to the theory of reduction, a few comments are in order. First, data are generated from an unknown high-dimensional probability distribution (the data generation process, or DGP), which is indexed by a set of parameters. Some functions of those parameters are of interest to an investigator. A central aim of modeling is to determine the numerical values of those parameters, which can be used for testing theories, forecasting, conducting economic policy, and learning about the economy.

Second, the DGP itself involves far too many parameters to estimate on available data, so reductions of the DGP are essential to produce an operational model. A key feature of every reduction is whether or not it involves a loss of information about the parameters of interest. Such a loss may be total (as when the parameters of interest no longer enter the model) or partial (as when some parameters can be gleaned but others not, or when some cease to be constant), or it may just lower the statistical efficiency with which the parameters may be estimated. Logically, reductions lead from the original DGP—which involves the universe of variables—to the distribution of a small subset thereof—which is the local DGP.

Third, an empirical model of that subset of variables then approximates their local DGP. Estimation of the parameters of that model comes last, logically speaking.

Fourth, an empirical model may include variables that do not actually enter the local DGP, in which case a more parsimonious representation can be selected from
the sample evidence. In that context, Section 6 discusses data-based simplification using the general-to-specific approach. The discussion below shows how the general-to-specific approach is the analogue in modeling of reduction in theory.

3.1 The Data Generation Process

This subsection formally (and briefly) discusses the statistical basis for the data generation process.

Let \( \{ u_t \} \) denote a stochastic process for a vector \( u_t \) of random variables at time \( t \) that is defined on the probability space \( (\Omega, \mathcal{F}, P) \), where \( \Omega \) is the sample space, \( \mathcal{F} \) is the event space (sigma field), and \( P \) is the probability measure. Let \( \psi \) denote a vector of parameters, which are entities that do not depend on \( \mathcal{F} \).

Consider the full sample \( U^1_T = (u_1 \ldots u_T) \), which is for \( t = 1, \ldots, T \) where the notation \( U^j \) means \( (u_i \ldots u_j) \) for \( j \geq i \); and denote the initial (pre-sample) conditions by \( U_0 = (\ldots u_{-2} u_{-1} u_0) \). The distribution function \( D_U (\cdot) \) of \( U^1_T \), conditional on \( U_0 \), is denoted by \( D_U (U^1_T | U_0, \cdot) \), which is often called the Haavelmo distribution; see Spanos (1989). To make explicit the observed phenomenon of parameter nonconstancy, \( D_U (\cdot) \) is represented parametrically by the \( kT \)-dimensional vector of parameters \( \psi_T = (\psi'_1 \ldots \psi'_T)' \), where each time period has an associated parameter \( \psi_t = (\psi'_{1,t} \ldots \psi'_{k,t})' \). Thus, elements of \( \psi \) need not be the same at each time \( t \), and some of the \( \{ \psi_{i,t} \} \) may reflect transient effects or regime shifts. The parameter space is \( \Psi \subseteq \mathbb{R}^{kT} \), so \( \psi_T \in \Psi \).

The data generating process of \( \{ u_t \} \) is therefore written as:

\[
D_U (U^1_T | U_0, \psi_T) \quad \text{with} \quad \psi_T \in \Psi \subseteq \mathbb{R}^{kT}.
\]

From equation (5), the complete sample \( \{ u_t, t = 1, \ldots, T \} \) is generated from \( D_U (\cdot) \) by a population parameter value, which is denoted \( \psi_p \).

3.2 The Reduction Sequence

This subsection considers the sequence of reductions that obtains the empirical model from the DGP. Because \( U^1_T \) is unmanageably large, operational models are defined by a sequence of data reductions, which can be viewed in ten stages:

- parameters of interest,
- data transformations and aggregation,
- sequential factorization,
- data partition,
- marginalization,
- mapping to stationarity,
- conditional factorization,
constancy,  
lag truncation, and  
functional form.

It is assumed that empirical modeling aims to determine the values of a set of parameters of interest that are relevant to an investigator’s objectives, such as testing theories or undertaking policy analysis. The key concern of any given reduction is its effect on the parameters of interest. To derive that effect, this subsection briefly considers each of these ten stages in turn, while noting that some of these stages do not involve a reduction per se.

**Parameters of interest.** Let the parameters of interest be denoted by \( \mathbf{\mu} \in \mathcal{M} \). Both economic theory and empirical properties may suggest that certain parameters are parameters of interest. Parameters that are identifiable and invariant to an empirically relevant class of interventions are likely to be of interest. Other parameters may be of interest, depending on the purpose of the exercise.

Also, if \( \mathbf{\mu} \) is not a function of \( \mathbf{\psi}_T^1 \), then the modeling exercise will be vacuous, so assume that \( \mathbf{\mu} = g (\mathbf{\psi}_T^1) \). After each reduction, it is essential to check that \( \mathbf{\mu} \) can still be retrieved from the parameters characterizing the lower-dimensional data density.

**Data transformations and aggregation.** Consider a one-to-one mapping of \( \mathbf{U}_T^1 \) to a new dataset \( \mathbf{W}_T^1 \): \( \mathbf{U}_T^1 \rightarrow \mathbf{W}_T^1 \). The variables in \( \mathbf{W}_T^1 \) may include aggregates of the original variables, their growth rates, etc. The transformation from \( \mathbf{U} \) to \( \mathbf{W} \) affects the parameter space, so \( \mathbf{\Psi} \) is transformed into \( \mathbf{\Phi} \) (say). Because densities are equivariant under one-to-one transformations, the DGP of \( \mathbf{W}_T^1 \) is characterized by the joint density of \( \mathbf{U}_T^1 \):

\[
D_{\mathbf{W}} \left( \mathbf{W}_T^1 \mid \mathbf{W}_0, \mathbf{\phi}_T^1 \right) = D_{\mathbf{U}} \left( \mathbf{U}_T^1 \mid \mathbf{U}_0, \mathbf{\psi}_T^1 \right),
\]

where \( \mathbf{\phi}_T^1 \) is the set of transformed parameters, with \( \mathbf{\phi}_T^1 \in \mathbf{\Phi} \).

For \( D_{\mathbf{W}} (\cdot) \) in equation (6), the key issue is how the transformation from \( \mathbf{\psi}_T^1 \) to \( \mathbf{\phi}_T^1 \) alters the properties of the parameters. Some parameters in \( \mathbf{\phi}_T^1 \) may be more constant than ones in \( \mathbf{\psi}_T^1 \); others may be less constant; and a smaller (or larger) number of the parameters in \( \mathbf{\phi}_T^1 \) may be needed to characterize the parameters of interest \( \mathbf{\mu} \).

**Sequential factorization.** Using the basic result that a joint probability equals the product of the conditional and marginal probabilities, and noting that time is irreversible, then sequentially factorize the density of \( \mathbf{W}_T^1 \) into its (martingale-difference) components:

\[
D_{\mathbf{W}} \left( \mathbf{W}_T^1 \mid \mathbf{W}_0, \mathbf{\phi}_T^1 \right) = \prod_{t=1}^{T} D_{\mathbf{w}} \left( \mathbf{w}_t \mid \mathbf{W}_{t-1}, \mathbf{\delta}_t \right),
\]

where \( \mathbf{W}_{t-1} = (\mathbf{W}_0 : \mathbf{W}_{t-1}^1) \), \( \mathbf{w}_t \) is the \( t \)-th column in \( \mathbf{W}_T^1 = (\mathbf{w}_1 \ldots \mathbf{w}_T) \), and \( \mathbf{\delta}_t \) is the parameterization resulting from the sequential factorization. The right-hand
side of equation (7) implicitly defines an innovation process \( \eta_t \), which equals \( w_t - \mathcal{E}(w_t | W_{t-1}^1) \).

**Data partition.** Now, partition \( W_T^1 \) into two sets, one set to be analyzed (\( X_T^1 \)) and one set to be marginalized (\( V_T^1 \)):

\[
W_T^1 = (X_T^1 : V_T^1),
\]

where \( X_T^1 \) is an \( T \times n \) matrix. Consequently, everything about \( \mu \) must be learnt from \( X_T^1 \) alone, which entails that \( V_T^1 \) must not be essential to inference about \( \mu \).

**Marginalization.** Actual marginalization proceeds as follows. Using the partition in equation (8), and noting that \( W_{t-1} = \{X_{t-1}^1, V_{t-1}^1, W_0\} \), factorize \( D_w(\cdot) \) into the conditional distribution of \( v_t \) given \( x_t \), and the marginal distribution of \( x_t \):

\[
D_w(w_t | W_{t-1}, \delta_t) = D_{v|x}(v_t | x_t, W_{t-1}, \delta_{a,t}) \cdot D_x(x_t | V_{t-1}^1, X_{t-1}^1, W_0, \delta_{b,t}).
\]

If only \( \{x_t\} \) is to be analyzed, with only \( D_x(x_t|\cdot) \) retained, then \( \mu \) must be obtainable from \( \{\delta_{b,t}\} \) alone.

If lagged information about \( v \) is also to be eliminated, then \( D_x(x_t|\cdot) \) must be marginalized with respect to \( V_{t-1}^1 \), requiring the very strong condition that:

\[
D_x(x_t | V_{t-1}^1, X_{t-1}^1, W_0, \delta_{b,t}) = D_x(x_t | X_{t-1}^1, W_0, \delta_{b,t}^*). \tag{10}
\]

There is no loss of information from eliminating the history \( V_{t-1}^1 \) if and only if \( \delta_{b,t} = \delta_{b,t}^* \forall t \), in which case the conditional sequential distribution of \( \{x_t\} \) does not depend on \( V_{t-1}^1 \). That is, \( v \) does not Granger-cause \( x \); see Granger (1969). In modeling, another important condition is that there is no loss of relevant information when \( \mu = g(\{\delta_{b,t}^*\}) \). That is still a strong condition, but less stringent than \( \delta_{b,t} = \delta_{b,t}^* \forall t \). Also, marginalizing \( v \) will entail a loss of information unless \( \delta_t = (\delta_{a,t}, \delta_{b,t}) \in \Delta_a \times \Delta_b \) for parameter spaces \( \Delta_a \) and \( \Delta_b \). Otherwise, the parameters of the conditional and marginal distributions in equation (9) are cross-linked.

The above discussion implies that modeling aggregated data (say) can be viewed as a two-step process. First, the disaggregated series for a given variable (such as expenditure) are transformed by a one-to-one transformation into the corresponding aggregated series and all but one of those disaggregated series. Second, those disaggregated series are marginalized. In a GVAR, aggregation arises explicitly: the country-specific aggregated foreign variables (\( x_{it}^* \)) are aggregated from the full set of domestic variables across all countries, using trade weights.

**Mapping to stationarity.** An economy may generate integrated data, where a variable that is integrated of order \( d \) (denoted \( I(d) \)) must be differenced \( d \) times to eliminate all unit roots. Mapping such data to stationarity is a reduction from \( I(d) \)
to \( l(0) \). This mapping is generally useful for interpreting the resulting models, and it is needed to ensure that conventional inference is valid for all parameters. Still, many inferences will be valid even if this reduction is not enforced; see Sims, Stock, and Watson (1990). While differencing the data can map the data to stationarity, cointegration can also eliminate unit roots between linear combinations of variables. Cointegration is merely noted here, as it is treated extensively in numerous books and expository articles: see Banerjee and Hendry (1992), Ericsson (1992), Banerjee, Dolado, Galbraith, and Hendry (1993), Johansen (1995), Hatanaka (1996), Doornik, Hendry, and Nielsen (1998), Hendry and Juselius (2001), and Juselius (2006) *inter alia*.

**Conditional factorization.** Typically in empirical modeling, some variables are treated as endogenous and others are treated as given or non-modeled. Formally, this partitioning of the variables arises by factorizing the density of \( n \) variables in \( x_t \) into sets of \( n_1 \) and \( n_2 \) variables \( y_t \) and \( z_t \):

\[
x_t' = (y_t' : z_t'),
\]

where \( y_t \) denotes the endogenous variables in \( x_t \), \( z_t \) denotes the non-modeled variables in \( x_t \), and \( n_1 + n_2 = n \). Using the partition in equation (11), the joint distribution of \( x_t \) on the right-hand side of equation (10) can always be factorized as:

\[
D_x (x_t \mid X_{t-1}^1, W_0, \delta_{b,t})
= D_{y|z} (y_t \mid z_t, X_{t-1}^1, W_0, \theta_{a,t}) \cdot D_z (z_t \mid X_{t-1}^1, W_0, \theta_{b,t}),
\]

where \( D_{y|z} (y_t \mid \cdot) \) is the conditional density of \( y_t \) given \( z_t \), \( D_z (z_t \mid \cdot) \) is the marginal density of \( z_t \), and \( \theta_{a,t} \) and \( \theta_{b,t} \) are those densities’ parameters. Modeling only \( y_t \) and treating \( z_t \) as given corresponds to modeling only \( D_{y|z} (y_t \mid \cdot) \) and discarding the marginal distribution \( D_z (z_t \mid \cdot) \) on the right-hand side of equation (12). No loss of information in this reduction corresponds to the condition that \( z_t \) is weakly exogenous for \( \mu \). Specifically, weak exogeneity requires that \( \mu = f(\theta_{a,t}) \) alone and that \( (\theta_{a,t}, \theta_{b,t}) \in \Theta_a \times \Theta_b \) for parameter spaces \( \Theta_a \) and \( \Theta_b \); see Engle, Hendry, and Richard (1983).

Equation (12) factorizes the joint distribution of \( x_t \) into the conditional density of \( y_t \) given \( z_t \) and the marginal density of \( z_t \). Equally, the joint distribution of \( x_t \) can be factorized as the conditional density of \( z_t \) given \( y_t \) and the marginal density of \( y_t \). That “reverse” factorization is:

\[
D_x (x_t \mid X_{t-1}^1, W_0, \delta_{b,t})
= D_{z|y} (z_t \mid y_t, X_{t-1}^1, W_0, \theta_{c,t}) \cdot D_y (y_t \mid X_{t-1}^1, W_0, \theta_{d,t}),
\]

where \( D_{z|y} (z_t \mid \cdot) \) is the conditional density of \( z_t \) given \( y_t \), \( D_y (y_t \mid \cdot) \) is the marginal density of \( y_t \), and \( \theta_{c,t} \) and \( \theta_{d,t} \) are those densities’ parameters. Which factorization
is more useful is in part an empirical question (e.g., which factorization delivers an empirically constant model) and in part economic (e.g., which factorization delivers economically more interpretable results). In Section 5, equations (12) and (13) together will aid in interpreting the structure of GVARs.

**Constancy.** Complete parameter constancy in the conditional density $D_{y|z}(y_t|\cdot)$ means that $\theta_{a,t} = \theta_a \forall t$, where $\theta_a \in \Theta_a$. In such a situation, if weak exogeneity holds, $\mu$ itself is constant because $\mu$ is a function of only $\theta_a$. While appearing simple enough, constancy is actually a subtle concept; see Hendry (1996) and Ericsson, Hendry, and Prestwich (1998).

**Lag truncation.** Lag truncation limits the extent of the history $X_{t-1}$ in the conditional density $D_{y|z}(y_t|\cdot)$ in equation (12). For instance, truncation at $s$ lags implies:

$$D_{y|z}(y_t | z_t, X_{t-1}, W_0, \theta_a) = D_{y|z}(y_t | z_t, X_{t-s}, W_0, \zeta).$$

In equation (14), no loss of relevant information requires that $\mu = f(\zeta)$.

**Functional form.** Functional form could be treated as a set of data transformations, but it merits some discussion on its own. Specifically, map $y_t$ into $y_t^*$ (= $h_1(y_t)$) and $z_t$ into $z_t^*$ (= $h_2(z_t)$), and denote the resulting data by $X^\dagger$. Assume that the transformations $y_t^*$ and $z_t^*$ together make $D_{y^*|z^*}(y_t^*|\cdot)$ approximately normal and homoscedastic, denoted $N_{n_1}(\lambda_t, \Upsilon)$. Then there is no loss of information—and no change in the specification—if:

$$D_{y^*|z^*}(y_t^* | z_t^*, X_{t-1}^{t-s}, W_0, \gamma) = D_{y|z}(y_t | z_t, X_{t-1}, W_0, \zeta).$$

A well-known example in which no loss of information occurs is transforming the log-normal density of a variable to the normal density in the logarithm of that variable. The left-hand side density in equation (15) defines the local (conditional) DGP of $y_t^*$. When joint normality holds, as is assumed here, that final conditional model is linear in the transformed space.

### 3.3 The Derived Model

This sequence of reductions delivers the derived model, which has the following specification:

$$A(L) h_1(y_t) = B(L) h_2(z_t) + \varepsilon_t \quad \varepsilon_t \xrightarrow{\text{GAP}} N_{n_1}(0, \Sigma_\varepsilon),$$

where $\varepsilon_t$ is a mean-zero, homoscedastic, mean-innovation process with variance $\Sigma_\varepsilon$, and $A(L)$ and $B(L)$ are constant-parameter polynomial matrices of order $s$ in the lag operator $L$. That is, $A(L)$ and $B(L)$ are matrices whose elements are polynomials. The error $\varepsilon_t$ is a derived process that is defined by:

$$\varepsilon_t = A(L) h_1(y_t) - B(L) h_2(z_t),$$
so $\varepsilon_t$ as given in equation (17) is not autonomous. For the same reason, equation (16) is a derived model, rather than an autonomous model.

Reduction from the DGP to the generic econometric equation in (16) involves all ten stages of reduction discussed above, thereby transforming the parameters $\psi^1_T$ in the DGP to the coefficients of the empirical model. Because the DGP is congruent with itself, equation (16) would be an undominated congruent model if there were no information losses from the corresponding reductions. Thus, valid reductions involve no losses in information; and econometrics has created concepts that correspond to avoiding possible losses. Three aspects are of particular interest for GVARs.

- **Aggregation** entails no loss of information when marginalizing with respect to disaggregates if the retained (aggregated) information provides a set of sufficient statistics for the parameters of interest $\mu$.

- **Conditional factorization** leads to no loss of information if the conditioning variables are weakly exogenous for $\mu$.

- **Parameter constancy** over time and with respect to interventions on the marginal processes (invariance) are essential for sustainable relationships.

The next three sections consider aggregation, conditional factorization, and parameter constancy in the context of a GVAR.

## 4 Aggregation and Marginalization

In a GVAR, one significant marginalization is aggregation across the variables of the individual foreign countries. This marginalization is testable in principle for a GVAR because the variables of all individual countries are available: the variables for a given country are the “domestic” variables for the VARX* for which that particular country is treated as domestic rather than foreign.

Testing this marginalization involves a practical difficulty involving degrees of freedom, as illustrated with DdPS’s VARX* for the United States. In that VARX*, there are 6 endogenous variables ($x_{0t}$) and 3 foreign variables ($x_{0t}^*$), each entering current-dated and at their first two lags. Each foreign aggregate in $x_{0t}^*$ is the trade-weighted average of the individual series for 25 foreign countries, implying 24 restrictions. For a given equation in the VARX*, there are thus 24 restrictions to be tested for each of 3 variables in $x_{0t}^*$, and at each of 3 lags. In total, there are 216 restrictions, which could be tested as 216 “zero” restrictions on regressors added to that equation in the VARX*. However, there are only 97 observations; and an individual equation in the VARX* already has 23 variables. Conventional approaches to testing these marginalizations are infeasible because the implied model has more variables than
observations. In outlining the approach to impulse indicator saturation in Autometrics, Section 6 shows how to deal with this situation. Section 7 implements tests of these aggregation assumptions for the U.S. VARX* in DdPS.

5 Conditioning and Weak Exogeneity

This section considers two roles of weak exogeneity in GVARs: the relationship of the conditional VARX* to the conditional-marginal factorization (Section 5.1), and tests of weak exogeneity (Section 5.2).

5.1 Conditioning

A GVAR—as formulated along the lines discussed in Section 2—involves an internally inconsistent treatment of conditioning. The following example illustrates.

Consider a GVAR with two countries—the United States \( i = 0 \) and the euro area \( i = 1 \)—and where for convenience lags are ignored. The VARX* for the United States expresses \( x_0t \) conditional on \( x^*_0t \), which is simply \( x_{1t} \) in this bi-country example. Likewise, the VARX* for the euro area expresses \( x_{1t} \) conditional on \( x^*_{1t} \), which here is \( x_{0t} \). The GVAR is the system of these two VARX* subsystems, stacked.

This GVAR does not represent a proper density function, as equations (12) and (13) clarify. Let \( y_t \) be the U.S. variable \( x_0t \), and \( z_t \) be the euro area variable \( x_{1t} \). The U.S. VARX* is thus the conditional density \( D_{y|z(x_0t|\cdot)} \) in equation (12), and the euro area VARX* is the conditional density \( D_{z|y(x_{1t}|\cdot)} \) in equation (13). The GVAR combines these two conditional densities, but these densities arise from different conditional-marginal factorizations of the joint density of \( (y_0t : z_0t) \). Hence, the GVAR does not constitute a valid representation of the joint density of \( (y_0t : z_0t) \).

While this illustration is for two countries only, the logic goes through immediately for all GVARs, regardless of the value of \( N \). Specifically, for arbitrary \( N \), the VARX* for country 0 expresses \( x_{0t} \) conditional on \( x^*_0t \). Because \( x^*_0t \) is the aggregate of the variables in all foreign countries \( \{x_{it}, i \neq 0 \} \), that VARX* expresses \( x_{0t} \) conditional on \( (x'_{1t} : x'_{2t} : \ldots : x'_{Nt})' \). The VARX* for country 1 expresses \( x_{1t} \) conditional on \( x^*_{1t} \), and hence conditional on \( (x'_{0t} : x'_{2t} : x'_{3t} : \ldots : x'_{Nt})' \). In general, the VARX* for country \( i \) expresses \( x_{it} \) conditional on \( x^*_{it} \), and hence conditional on \( \{x_{jt}, j \neq i \} \). Each of the \( N + 1 \) VARX* models is thus the initial conditional density in a conditional-marginal factorization of the joint density for \( (x'_{0t} : x'_{1t} : x'_{2t} : x'_{3t} : \ldots : x'_{Nt})' \). However, each of those \( N + 1 \) conditional-marginal factorizations is a different factorization of the joint density. The GVAR combines conditional densities from \( N + 1 \) different factorizations of the joint density and so does not constitute a valid representation of that joint density.

Two special cases of the bi-country example above help clarify. In the first case,
suppose that the VARX* for the United States is an exact relationship, i.e., \( u_{0t} = 0 \) in equation (1) for the relationship of \( x_{0t} \) conditional on \( x_{1t} \). Then, the VARX* for the euro area is also an exact relationship, with the coefficient on \( x_{0t} \) in the euro area VARX* being the inverse of the coefficient on \( x_{1t} \) in the VARX* for the United States; see Ericsson (1992, Section 2C). No additional information about the economies is gained from looking at the two conditional densities rather than just one of them. That said, additional information would be gained by looking at (e.g.) the marginal density for \( x_{1t} \). That marginal density, along with the conditional density for \( x_{0t} \) (conditional on \( x_{1t} \)), also delivers a complete representation of the joint density.

In the second case, suppose that the joint distribution of \((x_{0t} : x_{1t})'\) is independent normal with a diagonal covariance matrix. Then, the coefficient on \( x_{1t} \) is zero in the VARX* for the United States, and also the coefficient on \( x_{0t} \) is zero in the VARX* for the euro area. As in the first case above, no additional information about the economies is gained from looking at the two conditional densities rather than just one of them, but a marginal density could provide additional information and complete the representation of the joint density.

It is also feasible to analyze more general settings, as in Ericsson (1992, Section 2C). Regardless, the GVAR mixes conditional densities from different factorizations and so does not properly portray the joint density of the variables in the \( N + 1 \) countries. It still remains to decide which conditional-marginal factorization is the most useful, and Section 8 discusses some possible avenues to pursue.

5.2 Tests of Weak Exogeneity

DdPS test whether \( x_{it}^* \) is weakly exogenous for \( \beta_i \) in the VARX* (1) by testing whether the cointegrating vectors in that VARX* enter equation (4), which is the (marginal) vector error correction model for \( x_{it}^* \). That is, DdPS test whether \( \alpha_i^* = 0 \) in equation (4). This test is a direct implementation of the test for weak exogeneity proposed by Johansen (1992). While this test for weak exogeneity is intuitively appealing, the rich structure of the GVAR affords an alternative potentially more powerful approach to testing for weak exogeneity, as detailed below.

This alternative approach can be motivated by observing that each vector of foreign variables \( x_{jt} \) (for \( j \neq i \)) is associated with a set of cointegrating vectors \( \beta_j' (x_{jt-1} : d_{t-1} : x_{jt-1}^* : t)' \) through the VARX* for \( x_{jt} \). Thus, the full set of variables of the individual foreign countries is associated with the set of cointegrating relationships \( \beta_j' (x_{jt-1} : d_{t-1} : x_{jt-1}^* : t)' \) (again, for \( j \neq i \)). In the VARX* for \( x_{it} \), the variable \( x_{it}^* \) is the aggregate of the variables of all the individual foreign countries, so any and all of the cointegrating vectors associated with those foreign countries’ VARX* models may enter the marginal equation (2) for \( x_{it}^* \). This discussion highlights the centrality of parameters of interest—e.g., whether they be \( \beta_i \) or \( \{\beta_j, \ j = 0, \ldots, i, \ldots, N\} \) when discussing weak exogeneity. The implied test of weak exogeneity will differ, depending
upon the choice of parameters of interest.

Two testable implications of weak exogeneity follow from this observation about cointegration in the GVAR. One implication is for the conditional model (3): none of the cointegrating vectors associated with foreign-country VARX* models should enter the conditional model. The second implication is for the marginal model (2): the cointegrating vectors associated with the domestic-country VARX* model should not enter the marginal model, even when including in that marginal model the cointegrating vectors associated with foreign-country VARX* models. Section 7 reports both types of tests, including some with modifications using Autometrics. See also Boswijk (1995) and Ericsson (1995), who discuss similar issues in analyzing subsystems with separate cointegrating vectors.

### 6 Computer-automated Model Selection

This section describes the model selection algorithm in Autometrics (Section 6.1) and impulse indicator saturation (Section 6.2). Both are key to the empirical implementation of the tests of data aggregation, parameter constancy, and weak exogeneity in Section 7.

#### 6.1 The Algorithm in Autometrics

Hoover and Perez (1999) proposed an automated general-to-specific model-selection algorithm that incorporated many of the features of the “Hendry” or LSE methodology. Hendry and Krolzig (2001) developed a second-generation algorithm called PcGets, which extended and improved upon Hoover and Perez’s algorithm; see also Hendry and Krolzig (1999, 2003, 2005) and Krolzig and Hendry (2001). Doornik and Hendry (2009) implement a third-generation algorithm called Autometrics, which is part of PcGive version 13. Autometrics utilizes one-step and multi-step simplifications along multiple paths following a tree search method. Diagnostic tests serve as additional checks on the simplified models, and encompassing tests resolve multiple terminal models. Both analytical and Monte Carlo evidence show that the resulting model selection is relatively non-distortionary for Type I errors. At an intuitive level, Autometrics functions as a series of sieves that aim to retain parsimonious congruent models while discarding both noncongruent models and over-parameterized congruent models. This feature of the algorithm is eminently sensible, noting that the data generation process itself is congruent and is as parsimonious as feasible.

The current subsection summarizes Autometrics as an automated model-selection algorithm, thereby providing the necessary background for interpreting its application in Section 7. For ease of reference, the algorithm is divided into three “stages”, denoted Stage 0, Stage 1, and Stage 2. For full details of Autometrics’s algorithm,

Stage 0: the general model, impulse indicator saturation, and F pre-search tests. Stage 0 involves three parts: the estimation and evaluation of the general model, inclusion of impulse indicator dummies for all observations, and some pre-search tests aimed at simplifying the general model before instigating formal multi-path searches.

First, the general model is estimated, and diagnostic statistics are calculated for it. If any of those diagnostic statistics is unsatisfactory, the modeler must decide what to do next—whether to “go back to the drawing board” and develop another general model, or to continue with the simplification procedure, perhaps ignoring the offending diagnostic statistic or statistics.

Second, and optionally, Autometrics performs block additions and searches of impulse indicator dummies for all observations in a process known as impulse indicator saturation. Doing so generates a robust regression estimator, and it provides a check for parameter constancy. Section 6.2 discusses impulse indicator saturation further.

Third, and also optionally, Autometrics attempts to drop various sets of potentially insignificant variables. Autometrics does so by dropping all variables at a given lag, starting with the longest lag. Autometrics also does so by ordering the variables by the magnitude of their $t$-ratios and either dropping a group of individually insignificant variables or (alternatively) retaining only a group of individually statistically significant variables. In effect, an $F$ pre-search test for a group of variables is a single test for multiple simplification paths, a characteristic that helps control the costs of search. If these tests result in a statistically satisfactory reduction of the general model, then that new model is the starting point for Stage 1. Otherwise, the general model itself is the starting point for Stage 1.

Stage 1: a multi-path encompassing search. Stage 1 tries to simplify the model from Stage 0 by searching along multiple paths, ensuring that the diagnostic tests are not rejected. If all variables are individually statistically significant, then the initial model in Stage 1 is the final model. If some variables are statistically insignificant, then Autometrics tries deleting those variables to obtain a simpler model. If a simplification is rejected, Autometrics backtracks along that simplification path to the most recent previous acceptable model and then tries a different simplification path. A terminal model results if the model’s diagnostic statistics are satisfactory and if no remaining regressors can be deleted.

If Autometrics obtains only one terminal model, then that model is the final model. However, because Autometrics pursues multiple simplification paths in Stage 1, Autometrics may obtain multiple terminal models. To resolve such a situation, Autometrics creates a union model from those terminal models and tests each terminal model
against that union model. Autometrics then creates a new union model, which nests all of the surviving terminal models; and that union model is passed on to Stage 2.

**Stage 2: another multi-path encompassing search.** Stage 2 in effect repeats Stage 1 (possibly iteratively) by applying the simplification procedures from Stage 1 to the union model obtained at the end of Stage 1. The resulting model is the final model. If Stage 2 obtains more than one terminal model after applying encompassing tests, then the final model is selected by using the Akaike, Schwarz, and Hannan–Quinn information criteria. See Akaike (1973, 1981), Schwarz (1978), and Hannan and Quinn (1979) for the design of these information criteria, and Atkinson (1981) for the relationships between them.

In addition to optional pre-search simplification and impulse indicator saturation, the algorithm requires several other choices: in particular, target size and fixity of regressors.

**Target size.** Autometrics requires the modeler to choose which tests are calculated and to specify the critical values for those tests. In principle, the modeler can choose the test statistics and their critical values directly, although doing so is tedious because of the number of statistics involved. To simplify matters, Autometrics offers several options for the “target size”, which incorporates pre-designated selections of test statistics and critical values.

The “target size” is meant to equal “the proportion of irrelevant variables that survives the [simplification] process” (Doornik, 2009, p. 100). In the analysis below, Autometrics’s target size is either 5% or 1%—which are values that appear to approximate the liberal and conservative strategies in PcGets—or 0.1%, which is very stringent. The liberal strategy errs on the side of keeping some variables, even although they may not actually matter. The conservative strategy keeps only variables that are clearly significant statistically, erring in the direction of excluding some variables, even although those variables may matter. Which strategy is preferable depends in part on the data themselves, in part on the class of regressors examined (and, in particular, whether indicator saturation is considered), and in part on the objectives of the modeling exercise. Also, the two approaches may generate similar or identical results, as seen below.

**Fixity of regressors.** Autometrics offers the option of designating certain variables as “fixed”. Fixed variables are forced to always be included in regression, whereas free variables (variables that are not fixed) may be deleted by the algorithm.

In short, Autometrics is general-to-specific, multi-path, iterative, and encompassing, with diagnostic tests providing additional assessments of statistical adequacy, and with options for pre-search simplification and robustification. Autometrics can be characterized as having two components:
1. Estimation and diagnostic testing of the general unrestricted model (Stage 0); and
2. Selection of the final model by
   (a) pre-search simplification of the general unrestricted model (Stage 0),
   (b) impulse indicator saturation (Stage 0), and
   (c) multi-path (and possibly iterative) selection of the final model (Stages 1 and 2).

6.2 Impulse Indicator Saturation

Impulse indicator saturation (IIS) uses zero-one impulse indicator dummies to analyze properties of a model. There are $T$ such dummies, so inclusion of all $T$ dummies in a model is infeasible. However, blocks of dummies can be included, and that insight provides the basis for IIS. Two Monte Carlo examples motivate the generic approach in IIS. See Hendry, Johansen, and Santos (2008), Johansen and Nielsen (2009), and Hendry and Santos (2010) for further discussion and recent developments.

Example 1. This example illustrates the behavior of IIS when the model is correctly specified. Suppose that the DGP for variable $y_t$ is:

$$y_t = \mu_0 + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma^2),$$

and that the model estimated is a regression of $y_t$ on a constant. Figure 1a plots Monte Carlo data for a DGP with ($\mu_0 = 20$, $\sigma^2 = 1$, $T = 100$); and Figure 1b plots the estimated model’s residuals.

Now, re-estimate the model, but including impulse indicator dummies for the first half of the sample, as represented by Figure 2a. (That estimation is equivalent to estimating the model over the second half of the sample, ignoring the first half.) Drop all statistically insignificant impulse indicator dummies and retain the statistically significant dummies (Figure 2b). Repeat this process, but start by including impulse indicator dummies for the second half of the sample (Figure 2d) and retain the significant ones (Figure 2e). Now, re-estimate the original model, including all dummies retained in the two block searches (Figure 2g), and select the statistically significant dummies out of that combined set (Figure 2h). Hendry, Johansen, and Santos (2008) have shown that, under the null hypothesis of correct specification, the fraction of impulse indicator dummies retained is roughly $\alpha T$, where $\alpha$ is the target size. In Figure 2h, four dummies are retained, whereas $\alpha T = 5\% \cdot 100 = 5$, a close match.

Example 2. This example illustrates the behavior of IIS when there is an unmodeled break. Suppose that the DGP for variable $y_t$ is:

$$y_t = \mu_0 + \mu_t D_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma^2),$$

17
Figure 1: Actual and fitted values and the corresponding scaled residuals for the model $M_{\text{no break}}$ when the DGP does not have a break.

Figure 2: A characterization of split-sample impulse indicator saturation when the DGP does not have a break.
Figure 3: Actual and fitted values and the corresponding scaled residuals for the model $M_{no\ break}$ when the DGP has a break.

Figure 4: A characterization of split-sample impulse indicator saturation when the DGP has a break.
where $D_t = 0$ ($T = 1, \ldots, 75$), $D_t = 1$ ($T = 76, \ldots, 100$), and the model estimated is a regression of $y$ on a constant, ignoring the break $D_t$. Figure 3a plots Monte Carlo data for a DGP with ($\mu_0 = 20$, $\mu_1 = -10$, $\sigma^2 = 1$, $T = 100$); and Figure 3b plots the estimated model’s residuals. Interestingly, no residuals lie outside the estimated 95% confidence region, even though the break is $10\sigma$.

Figure 4 plots the corresponding graphs for split-sample impulse indicator saturation. The procedure has high power to detect the break, even although the nature of the break was not utilized in the procedure itself.

In practice, IIS in Autometrics utilizes many blocks, and the partitioning of the sample into blocks may vary over iterations of searches. IIS is a statistically valid procedure for integrated, cointegrated data; see Johansen and Nielsen (2009). IIS can also serve as a diagnostic statistic.

Many existing procedures can be interpreted as “special cases” of IIS in that they represent particular algorithmic implementations of IIS. Such special cases include recursive estimation, rolling regression, the Chow (1960) predictive failure statistic (including the 1-step, breakpoint, and forecast versions implemented in OxMetrics), the Andrews (1993) unknown breakpoint test, the Bai and Perron (1998) multiple breakpoint test, intercept correction (in forecasting), and robust estimation. IIS thus provides a general and generic procedure for analyzing a model’s constancy.

Algorithmically, IIS also solves the problem of having more regressors than observations by testing and selecting over blocks of variables. That approach permits testing the aggregation assumption discussed in Section 4; and such a test is implemented in the next section, using Autometrics.

7 Evaluation of the GVAR

This section implements the tests proposed above, using the GVAR in DdPS to illustrate and focusing on the VARX* specifications for the United States, the euro area, the United Kingdom, and China.

Data aggregation. Table 2 reports test statistics of data aggregation, at the 0.1% target level in Autometrics, equation by equation and by subsystem, for the United States, the euro area, the United Kingdom, and China.² The data aggregation implicit in the VARX* for the United States is strongly rejected for all equations of that VARX* except for the equation for the oil price. In two equations (those for $\Delta e_{US}$

²Because the subsystem for a given country is unrestricted (either as an unrestricted VARX*, or as a cointegrated VARX* if conditioned on the estimate of $\beta_i$), OLS estimation equation by equation is maximum likelihood estimation of the VARX*. Valid omitted-variables (LM) test statistics can be calculated on either the VARX* as a subsystem, or on the individual equations of the VARX*. These two approaches may imply different alternative hypotheses, even while the null hypothesis is the same. Table 2 reports both the equation-by-equation approach and the subsystem approach.
and $\Delta r_{US}$), the selected disaggregated (country-specific) foreign variables are jointly significant at at least the 0.005% level. Rejection may be occurring because trade weights need not be appropriate for all variables or in all equations. For interest rates in particular, relative capitalization or market activity might be more relevant. Also, the actual volume of trade between countries need not reflect the importance that the potential for trade may have on the actual pricing of goods and services in individual countries; see de Brouwer and Ericsson (1998) for further discussion.

In several equations for each of the euro area, the United Kingdom, and China, aggregation is likewise strongly rejected. In some equations, the included statistically significant country-specific foreign variables are suggestive of the aggregation issues at hand. For instance, in the equation for $\Delta e_{euro}$, those foreign variables are $\Delta e_p$ for Sweden and Switzerland, with an $F$-statistic of 298.74. In the equation for $\Delta^2 p_{UK}$, the foreign variable is $\Delta^2 p$ for Japan.

**Weak exogeneity (test of the conditional model).** Table 3 reports test statistics on the conditional model for the null hypothesis of weak exogeneity, equation by equation and by subsystem, for the United States, the euro area, the United Kingdom, and China. The GVAR in DdPS has 63 cointegrating vectors, two of which are in the VARX* for the United States. Hence, the conditional-model weak exogeneity test statistics for the United States examine whether the remaining 61 cointegrating vectors are statistically significant in the equations of the VARX* for the United States. This test of weak exogeneity is not rejected in the equation for $\Delta p_{oil}$, is strongly rejected in the equation for $\Delta l_r$, and is rejected at around the 3%-5% range in the remaining four equations. For the VARX* as a whole, this test of weak exogeneity is strongly rejected—at at least the 0.005% level. For the euro area and for the United Kingdom, this test of weak exogeneity is rejected for some equations and for the respective subsystems. This test is not rejected for China, either in individual equations or for its subsystem.

**Weak exogeneity (test of the marginal model).** Table 4 reports test statistics on the marginal model for the null hypothesis of weak exogeneity, equation by equation and by subsystem, for the United States, under different assumptions about the treatment of the cointegrating vectors from the foreign countries’ VARX* models. To discuss these assumptions, let $\alpha_{21}$ denote the cointegration feedback coefficients for the U.S. cointegrating vectors in the marginal model for $x^*_{0t}$, and $\alpha_{22}$ denote the cointegration feedback coefficients for all of the foreign-country cointegrating vectors in the marginal model for $x^*_{0t}$. Test statistics are reported for three types of hypotheses:

(i) $\{\alpha_{21} = 0 | \alpha_{22} \equiv 0\}$,

(ii) $\{\alpha_{21} = 0 | \alpha_{22} \text{ unrestricted}\}$, and

(iii) $\{\alpha_{21} = 0 | \alpha_{22} \text{ partially unrestricted}\}$, e.g., with some elements of $\alpha_{22}$ set to zero by Autometrics at a given target level.
Table 4 reports test statistics for all three types of hypotheses (by column), with two calculations for type (iii): one for a target size of 1%, and one for a target size of 0.1%. The effects of variable selection are pronounced in the marginal equations for $y^*$ and $\Delta p^*$, with weak exogeneity being rejected more strongly when foreign-country cointegrating vectors are allowed in the marginal model, and especially more strongly when some selection across those cointegrating vectors occurs. The test statistics for the marginal equation for $ep^*$ do not reject weak exogeneity.

**Parameter constancy (impulse indicator saturation).** Table 5 reports results from impulse indicator saturation at the 0.1% target level, equation by equation and by subsystem, for the United States, the euro area, the United Kingdom, and China. The table also lists the dates of the retained impulse indicator dummies. Parameter constancy is rejected in the VARX* models for all four countries. For the United States, parameter constancy is rejected for all equations except $\Delta p$. IIS of the equation for the short-term interest rate $r$ resulted in the most retained impulse indicator dummies, as might be expected for a variable that could be strongly influenced by shifts in monetary policy. For the United Kingdom, the impulse indicator dummy for 1992(4) in the equation for $\Delta ep$ reflects the substantial devaluation of the pound sterling around the time of Black Wednesday (September 16, 1992), when the British government was forced to withdraw the pound sterling from the ERM.

For China, IIS detects 18 dummies in the equation for $\Delta r$ (the change in the short-term interest rate), with an infinite $F$-statistic. The Chinese short-term interest rate is administered, with stretches of several quarters without change. Hence, $\Delta r$ is a series of zeros, interspersed with impulses that reflect the magnitude of change in the interest rate. It is encouraging that IIS demonstrates the ability to detect scattered impulses such as these, even although the block searches for IIS favor detecting impulses all within a single block.

### 8 Remarks

Several general remarks are germane.

First, which conditional-marginal factorization is the most useful? Economic reasoning may provide some guidance. Tests of super exogeneity may also provide indirect evidence on which way to factorize; see Engle, Hendry, and Richard (1983), Hendry (1988), Engle and Hendry (1993), and Hendry and Santos (2010). Also, in principle, an algorithm such as Autometrics could search the tree implied by the set of feasible factorizations.

---

3DdPS (Table IV) implement type (i), albeit with an additional lag on the domestic variables. Also, note that type (i) and type (ii) can be interpreted as implementing type (iii) for target sizes of 0% and 100% respectively.
<table>
<thead>
<tr>
<th>Equation</th>
<th>United States</th>
<th>Euro area</th>
<th>United Kingdom</th>
<th>China</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta y$</td>
<td>6.4717**</td>
<td>[no disaggregated data detected]</td>
<td>[no disaggregated data detected]</td>
<td>8.2075**</td>
</tr>
<tr>
<td></td>
<td>[0.0006]</td>
<td></td>
<td></td>
<td>[0.0000]</td>
</tr>
<tr>
<td></td>
<td>$F(3, 79)$</td>
<td></td>
<td></td>
<td>$F(6, 79)$</td>
</tr>
<tr>
<td>$\Delta^2 p$</td>
<td>13.204**</td>
<td>13.997**</td>
<td>30.264**</td>
<td>9.6511**</td>
</tr>
<tr>
<td></td>
<td>[0.0005]</td>
<td>[0.0000]</td>
<td>[0.0000]</td>
<td>[0.0000]</td>
</tr>
<tr>
<td></td>
<td>$F(1, 81)$</td>
<td>$F(3, 73)$</td>
<td>$F(1, 80)$</td>
<td>$F(9, 76)$</td>
</tr>
<tr>
<td>$\Delta eq$</td>
<td>6.4152**</td>
<td>[no disaggregated data detected]</td>
<td>[no disaggregated data detected]</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>[0.0000]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F(6, 76)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta ep$</td>
<td>–</td>
<td>298.74**</td>
<td>47.627**</td>
<td>9.1731**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.0000]</td>
<td>[0.0000]</td>
<td>[0.0000]</td>
</tr>
<tr>
<td></td>
<td>$F(2, 74)$</td>
<td>$F(4, 77)$</td>
<td>$F(5, 80)$</td>
<td></td>
</tr>
<tr>
<td>$\Delta r$</td>
<td>9.0223**</td>
<td>7.7292**</td>
<td>6.9570**</td>
<td>[no disaggregated data detected]</td>
</tr>
<tr>
<td></td>
<td>[0.0000]</td>
<td>[0.0000]</td>
<td>[0.0001]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F(9, 73)$</td>
<td>$F(8, 68)$</td>
<td>$F(4, 77)$</td>
<td></td>
</tr>
<tr>
<td>$\Delta lr$</td>
<td>6.0420*</td>
<td>17.348**</td>
<td>9.7188**</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>[0.0161]</td>
<td>[0.0001]</td>
<td>[0.0000]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F(1, 81)$</td>
<td>$F(1, 75)$</td>
<td>$F(6, 75)$</td>
<td></td>
</tr>
<tr>
<td>$\Delta poil$</td>
<td>[no disaggregated data detected]</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Subsystem</td>
<td>[no disaggregated data detected]</td>
<td>1313.7**</td>
<td>248.68**</td>
<td>71.804**</td>
</tr>
<tr>
<td></td>
<td>[0.0000]</td>
<td>[0.0000]</td>
<td>[0.0000]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\chi^2(36)$</td>
<td>$\chi^2(42)$</td>
<td>$\chi^2(12)$</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Test statistics on the conditional model for the null hypothesis of weak exogeneity ($\alpha_{12} = 0$), equation by equation and by subsystem, for the United States, the euro area, the United Kingdom, and China.

<table>
<thead>
<tr>
<th>Equation</th>
<th>United States</th>
<th>Euro area</th>
<th>United Kingdom</th>
<th>China</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta y$</td>
<td>2.0362*</td>
<td>1.1289</td>
<td>1.5640</td>
<td>0.89407</td>
</tr>
<tr>
<td></td>
<td>[0.0363]</td>
<td>[0.4173]</td>
<td>[0.1283]</td>
<td>[0.6472]</td>
</tr>
<tr>
<td></td>
<td>$F(61, 21)$</td>
<td>$F(61, 15)$</td>
<td>$F(60, 21)$</td>
<td>$F(62, 23)$</td>
</tr>
<tr>
<td>$\Delta^2 p$</td>
<td>1.9866*</td>
<td>1.0024</td>
<td>2.1438*</td>
<td>1.7627</td>
</tr>
<tr>
<td></td>
<td>[0.0414]</td>
<td>[0.5306]</td>
<td>[0.0276]</td>
<td>[0.0664]</td>
</tr>
<tr>
<td></td>
<td>$F(61, 21)$</td>
<td>$F(61, 15)$</td>
<td>$F(60, 21)$</td>
<td>$F(62, 23)$</td>
</tr>
<tr>
<td>$\Delta eq$</td>
<td>2.0799*</td>
<td>1.4414</td>
<td>1.5427</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>[0.0324]</td>
<td>[0.2194]</td>
<td>[0.1358]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F(61, 21)$</td>
<td>$F(61, 15)$</td>
<td>$F(60, 21)$</td>
<td></td>
</tr>
<tr>
<td>$\Delta ep$</td>
<td>–</td>
<td>2.1887*</td>
<td>1.9507*</td>
<td>1.0833</td>
</tr>
<tr>
<td></td>
<td>[0.0471]</td>
<td>[0.0457]</td>
<td>[0.4302]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F(61, 15)$</td>
<td>$F(60, 21)$</td>
<td>$F(62, 23)$</td>
<td></td>
</tr>
<tr>
<td>$\Delta r$</td>
<td>1.8828</td>
<td>1.1300</td>
<td>1.9280*</td>
<td>1.1119</td>
</tr>
<tr>
<td></td>
<td>[0.0545]</td>
<td>[0.4164]</td>
<td>[0.0485]</td>
<td>[0.4014]</td>
</tr>
<tr>
<td></td>
<td>$F(61, 21)$</td>
<td>$F(61, 15)$</td>
<td>$F(60, 21)$</td>
<td>$F(62, 23)$</td>
</tr>
<tr>
<td>$\Delta lr$</td>
<td>3.0119**</td>
<td>1.4480</td>
<td>1.5433</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>[0.0034]</td>
<td>[0.2164]</td>
<td>[0.1356]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F(61, 21)$</td>
<td>$F(61, 15)$</td>
<td>$F(60, 21)$</td>
<td></td>
</tr>
<tr>
<td>$\Delta poil$</td>
<td>1.5703</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>[0.1258]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$F(61, 21)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsystem</td>
<td>2.0446**</td>
<td>1.5966**</td>
<td>1.8573**</td>
<td>1.0876</td>
</tr>
<tr>
<td></td>
<td>[0.0000]</td>
<td>[0.0100]</td>
<td>[0.0001]</td>
<td>[0.3330]</td>
</tr>
<tr>
<td></td>
<td>$F(366, 104)$</td>
<td>$F(366, 68)$</td>
<td>$F(360, 104)$</td>
<td>$F(248, 82)$</td>
</tr>
</tbody>
</table>
Table 4: Test statistics on the marginal model for the null hypothesis of weak exogeneity ($\alpha_{21} = 0$), equation by equation and by subsystem, for the United States, under different assumptions about $\alpha_{22}$.

<table>
<thead>
<tr>
<th>Equation</th>
<th>Restriction</th>
<th>[with target size in angled brackets &lt; · &gt;]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>${\alpha_{21} = 0}$</td>
<td>${\alpha_{22} = 0}$</td>
</tr>
<tr>
<td></td>
<td>$\alpha_{22} \equiv 0$</td>
<td>$\alpha_{22}$</td>
</tr>
<tr>
<td></td>
<td>$&lt; 0% &gt;$</td>
<td>$&lt; 100% &gt;$</td>
</tr>
</tbody>
</table>

\[\Delta y^*\]

|          | 0.52067 | 2.5862 | 3.7595* | 0.52067 |
|          | [0.5960] | [0.0961] | [0.0277] | [0.5960] |
|          | $F(2, 85)$ | $F(2, 24)$ | $F(2, 76)$ | $F(2, 85)$ |

\[\Delta^2 p^*\]

|          | 4.8109* | 3.9737* | 6.5474** | 9.6221** |
|          | [0.0105] | [0.0323] | [0.0025] | [0.0002] |
|          | $F(2, 85)$ | $F(2, 24)$ | $F(2, 68)$ | $F(2, 71)$ |

\[\Delta ep^*\]

|          | 1.2756 | 0.058953 | 1.0669 | 2.1998 |
|          | [0.2845] | [0.9429] | [0.3489] | [0.1172] |
|          | $F(2, 85)$ | $F(2, 24)$ | $F(2, 80)$ | $F(2, 84)$ |

| Subsystem | VARX*(2,2) | 13.647* | 11.866 |
| Subsystem | [0.0338] | [0.0650] |
|           | $\chi^2(6)$ | $\chi^2(6)$ |

| Subsystem | VARX*(3,2) | 8.5265 | 7.8067 |
| Subsystem | [0.2020] | [0.2526] |
|           | $\chi^2(6)$ | $\chi^2(6)$ |

Table IV]
Table 5: Impulse indicator saturation at the 0.1% target level, equation by equation and by subsystem, for the United States, the euro area, the United Kingdom, and China.

<table>
<thead>
<tr>
<th>Equation</th>
<th>United States</th>
<th>Euro area</th>
<th>United Kingdom</th>
<th>China</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta y$</td>
<td>9.5024** [0.0002]</td>
<td>[no impulse indicators detected]</td>
<td>[no impulse indicators detected]</td>
<td>261.28** [0.0000]</td>
</tr>
<tr>
<td>$F(2, 80)$</td>
<td>1980(2), 1982(1)</td>
<td></td>
<td></td>
<td>14 dummies</td>
</tr>
<tr>
<td>$\Delta^2 p$</td>
<td>[no impulse indicators detected]</td>
<td>4.3977** [0.0067]</td>
<td>4.7081* [0.0330]</td>
<td>17.058** [0.0000]</td>
</tr>
<tr>
<td>$F(3, 73)$</td>
<td>1981(1), 1986(3), 1986(4)</td>
<td>$F(1, 80)$</td>
<td>$F(4, 81)$</td>
<td></td>
</tr>
<tr>
<td>$\Delta eq$</td>
<td>21.850** [0.0000]</td>
<td>18.816** [0.0000]</td>
<td>[no impulse indicators detected]</td>
<td>–</td>
</tr>
<tr>
<td>$F(1, 81)$</td>
<td>1987(4)</td>
<td>$F(1, 75)$</td>
<td>1987(1)</td>
<td></td>
</tr>
<tr>
<td>$\Delta ep$</td>
<td>– [no impulse indicators detected]</td>
<td>23.271** [0.0000]</td>
<td>52.951** [0.0000]</td>
<td></td>
</tr>
<tr>
<td>$F(1, 80)$</td>
<td>1992(4)</td>
<td>$F(10, 75)$</td>
<td>10 dummies</td>
<td></td>
</tr>
<tr>
<td>$\Delta r$</td>
<td>61.369** [0.0000]</td>
<td>26.137** [0.0000]</td>
<td>16.987** [0.0001]</td>
<td>+∞** [0.0000]</td>
</tr>
<tr>
<td>$F(8, 74)$</td>
<td>8 dummies</td>
<td>$F(2, 74)$</td>
<td>$F(1, 80)$</td>
<td>1982(4)</td>
</tr>
<tr>
<td></td>
<td>1981(2), 1992(3)</td>
<td>1985(1)</td>
<td>18 dummies</td>
<td></td>
</tr>
<tr>
<td>$\Delta lr$</td>
<td>15.243** [0.0000]</td>
<td>[no impulse indicators detected]</td>
<td>9.1190** [0.0034]</td>
<td>–</td>
</tr>
<tr>
<td>$F(3, 79)$</td>
<td>1980(2), 1983(3), 1984(2)</td>
<td>$F(1, 80)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta poil$</td>
<td>11.813** [0.0009]</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$F(1, 81)$</td>
<td>1990(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsystem</td>
<td>150.24** [0.0000]</td>
<td>28.996** [0.0001]</td>
<td>58.818** [0.0000]</td>
<td>796.10** [0.0000]</td>
</tr>
<tr>
<td>$\chi^2 (18)$</td>
<td>1980(4), 1981(4), 1987(4)</td>
<td>$\chi^2 (6)$</td>
<td>$\chi^2 (12)$</td>
<td>$\chi^2 (40)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1987(1)</td>
<td>1985(1), 1987(1)</td>
</tr>
</tbody>
</table>
Second, in spite of the problems with combining conditional models from different conditional-marginal factorizations, the current implementation of GVARs may “work” in certain restrictive cases, as when the underlying VAR of \((x'_{it}, x'_{it})\) has a block-diagonal error covariance matrix. In that situation, the coefficients on \(x'_{it}\) in the VARX* for \(x_{it}\) would be zero. That implies that the conditional model for \(x_{it}\) would be the same as its marginal model, which itself is the same as the equations for \(x_{it}\) in the VAR of \((x'_{it}, x'_{it})\). The GVAR would simply be the VAR of \((x'_{it}, x'_{it})\).

Third, weak exogeneity is generally not sufficient to validate impulse response analysis. Depending upon the way in which impulse responses are calculated, strong exogeneity, super exogeneity, or even super super strong exogeneity may be required; see Ericsson, Hendry, and Mizon (1998). For instance, impulse responses are interpretable as being generated by a shift in the intercepts, hence requiring super exogeneity for that class of interventions.

Fourth, the reduced-form normalization of the cointegrating vectors in the VARX* may be problematic for some purposes if no domestic variables enter a given cointegrating vector; cf. the assumptions in Harbo, Johansen, Nielsen, and Rahbek (1998).

Fifth, forecasting may be affected by observed (and future) breaks in structure. Clements and Hendry (1998, 1999, 2002) and Hendry (2006) discuss techniques for robustifying forecasts, particularly forecasts from vector error correction models (and hence GVARs), which are naturally sensitive to shifts in equilibrium mean.

9 Conclusions

A global vector autoregression is an ingenious structure for capturing international linkages between country- or region-specific error correction models. A GVAR embodies a standardized economically appealing choice of variables for each country or region; it treats long-run properties through cointegration analysis in a systematic fashion; and it permits flexible dynamic specification through vector error correction modeling. The current paper re-examines the theoretical and empirical underpinnings for GVARs, focusing on data aggregation, exogeneity, and parameter constancy. This paper proposes extensions and improvements in these areas, with the aim of achieving an even more robust approach to GVAR modeling. Recent developments in computer-automated model selection allow implementation of tests that historically have been viewed as infeasible. To illustrate these refinements to GVAR modeling, tests of data aggregation, exogeneity, and parameter constancy are calculated for the GVAR in DdPS, focusing on the VARX* specifications for the United States, the euro area, the United Kingdom, and China.
References


