Chapter 5

Graphical modeling of structural VARs

5.1 The Structural VAR, SVAR

In this chapter we describe structural VAR models and how they can be constructed using the methods of *graphical modeling*. The form of the structural VAR (SVAR) that we will use differs primarily from that of the canonical VAR model by including regression terms for the dependence between current variables in addition to dependence of current on past variables. These terms explain any correlation between the innovations of the canonical VAR so that the residual series, or structural model innovations, from the SVAR model will be uncorrelated with each other. Our aim is also that structural models be sparse, in the sense that they represent the dependence using a relatively small number of model terms or coefficients. It is reasonable to believe that in a model that reflects the true structure of a system of inter-related series, the current value of each series should depend on a relatively small number of other current and lagged values.

Identifiability is a major concern with structural models. For the canonical VAR model, it was necessary only to identify the order to identify a unique model. Structural modeling in econometrics has traditionally relied strongly on economic theory to specify the terms in the model, with constraints to ensure that it can be uniquely identified. This is because there are many possible structural models which could equally well represent the observable statistical properties of the time series. However, our approach to the identification of structural VAR models is largely empirical; we would use contextual information where available and relevant to help guide us to an appropriate model, but our procedure is mainly to use appropriate statistics to identify the sparse structure of a well-fitting model. It is based on the theory of graphical modeling, and the statistics it uses are the partial correlations between current and lagged values of the series.

5.2 The directed acyclic graph, DAG

It is natural to represent causal relationships among variables in a diagrammatic manner using *directed graphs* in which an arrow *links* each variable to the others which it affects causally. The books by Whittaker (1990), Lauritzen

(1996) and Edwards (2000) provide accessible accounts of general statistical modeling procedures that are based on these representations of dependence. This topic has been an extremely active subject of research in recent years, but we limit ourselves to considering the specific application to structural modeling of multivariate time series. An application in this area which makes reference to the wider literature is found in Awokuse and Bessler (2003).

The canonical VAR, presented in the previous chapter, models the current observations of the multivariate series at time t through the dependence on previous (lagged) observations of the time series themselves. A visually effective way to present this model for a set of three series is by the graph in Figure 5.1, where each of the six variables is a *node*.

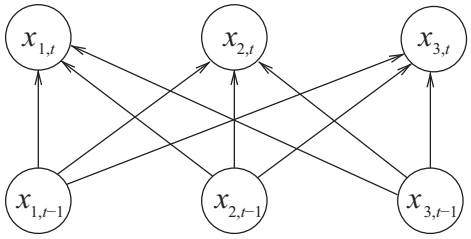

Figure 5.1 *Graphical representation of a saturated canonical VAR(1).*

In a VAR model, all current variables at time t, e.g., $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$, depend on all past variables: in the case of a VAR(1), the variables at time $t-1$. Such dependence in the graph is represented by the arrows or *directed edges* linking the nodes representing current and past variables, where the direction is suggested by the flow of time. We can simply indicate this complete dependence on the past by saying that the VAR is *saturated* on past variables. As discussed in the previous chapter and as is apparent from the graph, in the VAR there are no terms describing dependence between current variables and, because of that, typically the innovations are correlated. The dependence between the innovations in the VAR reflects the conditioning of the current variables upon all the past variables. Including explicit terms for the dependence between current variables provides an alternative way to capture the dependence between innovations. We use as an initial example the monthly Flour price series shown in Figure 1.2, for which the criteria AIC, HQC and SIC all selected an order 2 autoregression. The series were of simultaneous prices in the three cities of Buffalo, Minneapolis and Kansas City, and it is unsurprising that the innovations from the canonical VAR(2) were highly correlated, as shown by their correlation matrix:

$$\begin{pmatrix} 1.0000 & 0.9664 & 0.8700 \\ 0.9664 & 1.0000 & 0.8976 \\ 0.8700 & 0.8976 & 1.0000 \end{pmatrix}$$

By a structural VAR (SVAR) we will therefore mean a multivariate time

series model of the form

$$\Phi_0 x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + a_t, \tag{5.1}$$

where in contrast to the (canonical) VAR, the non-singular matrix Φ_0 has rows by which current values of the elements of x_t are related, and the elements of the residuals a_t are assumed to be uncorrelated. It is immediate that every model of the form (5.1) can be transformed to a canonical VAR by dividing through by Φ_0, i.e.,

$$\begin{aligned} x_t &= \Phi_0^{-1}\Phi_1 x_{t-1} + \Phi_0^{-1}\Phi_2 x_{t-2} + \cdots + \Phi_0^{-1}\Phi_p x_{t-p} + \Phi_0^{-1} a_t \quad (5.2)\\ &= \Phi_1^* x_{t-1} + \Phi_2^* x_{t-2} + \cdots + \Phi_p^* x_{t-p} + e_t. \quad (5.3) \end{aligned}$$

The result is known as the *reduced form* VAR (Lütkepohl, 1993, pp. 54, 325). The residuals a_t of the SVAR are therefore related to the innovations e_t of the VAR, and the general covariance matrix V of e_t to the diagonal covariance matrix D of a_t, by

$$\begin{aligned} \Phi_0 e_t &= a_t \quad (5.4)\\ \Phi_0 V \Phi_0' &= D. \quad (5.5) \end{aligned}$$

Although any given SVAR determines a unique corresponding VAR reduced form, there are many SVAR models that correspond to a given VAR, resulting from the possible choices of Φ_0 that diagonalize V in (5.5). It is therefore our hope that the sparse parameterization of the SVAR will help us to identify just one, or possibly a small number of these, which can adequately represent the structure of the series. Note that, because the transformation (5.2) from the SVAR to the VAR depends on the inverse of Φ_0, a sparse SVAR is likely to have a reduced form VAR with much less sparse coefficient matrices. However, the inverse of the innovation covariance matrix is $V^{-1} = \Phi_0' D^{-1} \Phi_0$, which will reflect in part any sparse aspects of Φ_0. This is actually a special case of a property arising in the general theory of graphical modeling which we will shortly introduce and apply to the identification of SVAR models.

The word *structural*, as we use it for the SVAR (5.1), has a meaning somewhat distinct from its original use for a structural model in econometrics (Hurwitz, 1962). In that context a structural model should be able to predict the effects of interventions which are of the nature of changes in a variable or parameter. In fact, our SVAR has the form of a *simultaneous equation model*; see Zellner and Theil (1962). We will refer to this as a SimEM, because the abbreviation SEM is generally used for the wider class of structural equation models.

In the general SimEM, the elements of a_t are not explicitly required to be uncorrelated, but the model is only generally identifiable if a large number of constraints are imposed on the coefficients, typically zero values. Without these constraints, many different sets of parameters would equally well represent the observable statistical properties of the data. The dependencies are

suggested by theoretical considerations underpinning the analyzed system of equations and are not empirically identified. For example, Zellner and Theil (1962) model the current relationships between six economic variables by using economic theory.

For our SVAR model we have assumed that the elements of a_t are uncorrelated, i.e., the covariance matrix D of a_t is diagonal, otherwise the model would certainly not be identifiable. However, we will also assume that

- each row of Φ_0 describes how one element of x_t depends on some (if any) of the other current elements;
- this dependence is recursive, not cyclical. This means that we could reorder the elements of x_t so that Φ_0 is upper triangular with unit diagonals: each element in turn depends on none, one or more elements that are *lower* in the ordering, taking the first as the highest. The ordering need not be unique; indeed, if there were no dependence at all, the ordering could be arbitrary.

The dependence between variables in an SVAR is conveniently illustrated by the sparse SVAR shown in Figure 5.2. We contrast this with the saturated VAR in the graph in Figure 5.1, noting that it includes directed edges linking current variables. The number of coefficients in this SVAR is the number, 8, of links, whereas the VAR in Figure 5.1 has 12 coefficients, including the 3 correlation coefficients between current variables, which are not explicit in the figure. These correlations are modeled in the SVAR by the links between current variables.

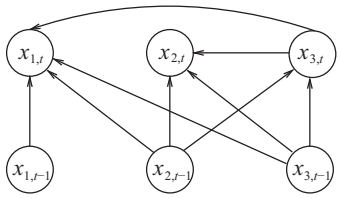

Figure 5.2 *An example of a graph representing a sparse SVAR(1).*

The resulting graph should be acyclic, i.e., it should *not* have any cyclic dependence between current variables. The graph in Figure 5.3 gives an example of the cyclic dependence that we exclude: it is possible by following the arrows from one variable eventually to return to the same variable; thus $x_{1,t}$ affects $x_{3,t}$, which affects $x_{2,t}$, which affects $x_{1,t}$.

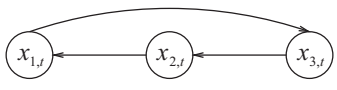

Figure 5.3 *A graph representing cyclic dependence.*

The diagram representing the SVAR, such as that in Figure 5.2, consisting of *nodes* with *directed edges*, is an example of a *directed acyclic graph* (DAG). Nodes (or their associated variables) that have a directed edge leading to a given node are known as the *parents* of that node, and the node itself as their *child*. Thus, in Figure 5.2, $x_{2,t-1}$, $x_{3,t-1}$ and $x_{3,t}$ are the parents of $x_{2,t}$. In general, a DAG completely determines the distribution of a set of variables, given, for each node, its distribution conditional upon its parents. Because we are assuming the normal linear model, this means that the linear regression of each current variable upon its parents fully determines the properties of the multivariate time series x_t. As we will see in Section 5.6, it also greatly simplifies the estimation of the model, allowing the use of ordinary least squares separately for the regression equation of each current variable. It is essential for this purpose that the graph be acyclic. However, we do reconsider in Section 5.9 the possibility of describing the dependence between current variables using a simultaneous equation model, for which an acyclic directed graph is not an appropriate description.

To confirm that the SVAR represented by a DAG does completely determine the series properties, consider starting at some initial time t_0 with a given set of values (say zeros) for the past variables $x_{t_0-1}, \ldots, x_{t_0-p}$. Then x_t may be simulated using the regression equation for each component current variable on other current and lagged values, provided this is carried out in the order determined by the DAG. For the SVAR illustrated in Figure 5.2, $x_{3,t}$ is first generated from $x_{2,t-1}$ and $x_{3,t-1}$, then $x_{2,t}$ from $x_{2,t-1}$, $x_{3,t-1}$ and $x_{3,t}$ and finally $x_{1,t}$ from $x_{1,t-1}$, $x_{2,t-1}$, $x_{3,t-1}$ and $x_{3,t}$. In this case it does not matter in which order $x_{1,t}$ and $x_{2,t}$ are generated, only that an acyclic ordering exists. Successive values x_{t_0}, x_{t_0+1}, x_{t_0+2}, ... may be determined in this way, and, provided the reduced form of the model is stationary, the process so generated will reach the stationary equilibrium. This thought experiment also draws attention to the fact that the DAG as exemplified in Figure 5.2 is partial. To be complete, the same edges between the current components of x_t should also be shown between the components of x_{t-1}, and also the component variables of x_{t-2} should be shown as predictive for x_{t-1}. This extension could be continued indefinitely into the past and the future, but, of course, it is not necessary to do so. The DAG showing only the links between current variables and from lagged to current variables is sufficient for a process model that is stationary or invariant to a time shift.

The key to constructing an SVAR with a sparse structure that adequately represents the observed series is the determination of the (or an) ordering of the dependence of current variables using appropriate statistical summaries of the data. The theory of graphical models assists us once more; in the next section we describe how to estimate a conditional independence graph which can enable us to achieve this aim. To conclude this section, we remark that a saturated SVAR model is one which, for the chosen ordering of current variables, has no sparse structure, i.e., each current variable is regressed upon all those current variables that are lower in the ordering, and on all the lagged

values up to the specified order. Such a model is exactly equivalent to the saturated VAR model, so it confers no advantage over the VAR, and any one such model is equivalent to any other based upon a different ordering of the current variables.

5.3 The conditional independence graph, CIG

We introduce this section by developing a DAG model for the dependence between the innovations of a multivariate time series, using the Flour prices as an example. Given that we have identified the order of a VAR model for the series and estimated its parameters, the innovations are effectively observable (subject to some sampling variability of the parameter estimates). We will go on to show how this can be extended to a DAG representing an SVAR model for the Flour prices themselves. This strategy is possible because, given a DAG representation of a multivariate times series x_t, such as illustrated in Figure 5.2, the innovations e_t from a canonical VAR fitted to the series can be described by a DAG having the subgraph restricted to the current variables. We are simply conditioning on all the lagged values; it is not necessary to know which lagged values are actually linked to current values. Figure 5.4 shows the subgraph of Figure 5.2 to illustrate this point, which we will shortly apply to the Flour price series.

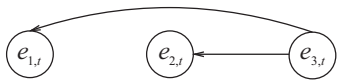

Figure 5.4 *The DAG representing the innovations of the SVAR(1) described by the DAG in Figure 5.2.*

In Section 5.1 we displayed the correlation matrix between the innovations of the Flour price series. We noted that the innovations were all highly correlated, suggesting that any one series of innovations can be predicted from the others. However, it is often the case with highly correlated sets of variables that some variables do not make a significant contribution to a prediction in the presence of other predictors, although they are correlated with the predicted variable. Taking each variable in turn as the predicted or response, and taking all the other variables together as predictors, we can determine which are relevant in the presence of all the others and, just as important, which are not. A graph may then be constructed in which the nodes (as for a DAG) correspond to the set of variables, and any two nodes (variables) are connected by an edge only if the first variable is dependent on the second when all the other variables are also included as predictors. In the general context of graphical modeling, we say that the edge between two nodes is *absent* if and only if the corresponding pair of variables are independent *conditional upon all the remaining variables in the set*. The resulting graph is then called a *conditional independence graph* (CIG). We note that the edges are not directed because by

this definition the relationship between two nodes is symmetric, even though we introduced the idea by considering one variable as predicted and the other as one of a set of predictors. The CIG clearly represents structure of the variables that differs from that specified in a DAG, but we will see how it can be used to draw conclusions about possible DAG representations by exploiting the conditional independence property. Most important, it can be directly estimated from the data, with no requirement to specify an ordering of the variables.

In our context of the normal linear model, the conditional independence graph is fully determined by the covariance matrix V of the variables (or by its correlation matrix). In order to construct the CIG from this matrix, it is not necessary to determine the linear predictor of each variable in turn, from all the others, though in principle this could be done. The conditional independence between two normal variables, given the remainder, is determined by their partial correlation being zero.

We introduced partial correlations in Section 2.9. In a system of several variables, the partial correlation between two of them is just their correlation once the linear dependence of each of them on the remaining variables has been subtracted. This definition extends to non-normal variables, but it is convenient to retain the notation of multivariate normality where subtracting the linear dependence on other variables is equivalent to conditioning upon them. More formally, given a set of random variables $X = X_1, \ldots, X_n$, the partial correlation $\pi(X_i, X_j)$ between two random variables X_i and X_j included in X is given by

$$\pi(X_i, X_j) = \rho(X_i, X_j \mid X \setminus \{X_i, X_j\}), \tag{5.6}$$

where $X \setminus \{X_i, X_j\}$ is the set of all variables except X_i and X_j. We have the following connection with prediction, that in the minimum mean square error linear predictor of X_i from $X \setminus X_i$, the coefficient of X_j is zero if and only if $\pi(X_i, X_j) = 0$ and symmetrically the coefficient of X_i is zero in predicting X_j.

As previously presented in Section 2.9, for a set of random variables $X = X_1, \ldots, X_n$ with covariance matrix V=Var(X), we can directly construct the whole matrix of partial correlations as

$$\pi(X_i, X_j) = \frac{-W_{i,j}}{\sqrt{W_{i,i}W_{j,j}}}, \tag{5.7}$$

where $W_{i,j}$ are the elements of the matrix $W = V^{-1}$.

We can represent a summary of all the non-zero partial correlations by a diagram in which the nodes representing the random variables are linked if their partial correlation is different from zero. Such a diagram is called a *partial correlation graph*. If the variables are normally distributed, a zero partial correlation corresponds to conditional independence, i.e.,

$$\pi(X_i, X_j) = 0 \Leftrightarrow X_i \perp X_j \mid X \setminus \{X_i, X_j\}, \tag{5.8}$$

and in such a case the diagram is also the conditional independence graph (CIG). Figure 5.5 shows the CIG that we identify below for the Flour price series innovations. The implications of the CIG for determining a DAG representation of the dependence between the variables will be considered in the next section. The important point is that we can estimate the CIG using the sample partial correlation matrix between the variables and identifying those entries which, subject to sampling variability, may be judged to be zero.

We need then a test procedure to decide whether a non-zero partial correlation can be considered to be zero. A large value for a sample partial correlation suggests a non-zero partial correlation, so we need to determine critical values, i.e., thresholds above which a value is to be considered significantly different from zero, for a prescribed level of significance. The matrix of sample partial correlations is readily constructed from the sample covariance matrix of the data and the critical values for these are determined as follows:

1. Given a random sample of a set of normal random variables $X = \{X_1, \ldots, X_m\}$ with covariance matrix V, let the *data matrix* $\mathbf{X}$ consist of columns of length n of the mean corrected samples of X. The sample covariance matrix is then $\widehat{V} = \frac{1}{n}\mathbf{X}'\mathbf{X}$. The sample inverse covariance matrix is computed as $\widehat{W} = \widehat{V}^{-1}$ and by its entries, $\widehat{W}_{i,j}$, the sample partial correlations can be calculated as

$$\widehat{\pi}(X_i, X_j) = \frac{-\widehat{W}_{i,j}}{\sqrt{\widehat{W}_{i,i}\widehat{W}_{j,j}}}. \tag{5.9}$$

2. Under the hypothesis that $\pi(X_i, X_j) = 0$, the ratio

$$\frac{\widehat{\pi}(X_i, X_j)\sqrt{n-m+1}}{\sqrt{1-\widehat{\pi}(X_i, X_j)^2}} \tag{5.10}$$

   is distributed as a t_{n-m+1} variable where $n-m+1$ is the number of degrees of freedom, because this ratio is the t-value of the estimate of the coefficient β_j in the regression

$$X_i = \beta_1 X_1 + \cdots + \beta_{i-1}X_{i-1} + \beta_{i+1}X_{i+1} + \cdots + \beta_m X_m + a_i \tag{5.11}$$

   of X_i on $X \setminus \{X_i\}$.
3. We reject the null hypothesis that $\pi(X_i, X_j) = 0$ at level α if

$$|\widehat{\pi}(X_i, X_j)| > \frac{t_{\alpha/2,n-m+1}}{\sqrt{t^2_{\alpha/2,n-m+1} + (n-m+1)}}, \tag{5.12}$$

   where $t_{\alpha/2,n-m+1}$ is the corresponding critical value of the t_{n-m+1} distribution.

The estimated CIG is then constructed with links between the nodes X_i

and X_j only if this test is rejected. In large samples, say when $n - m > 30$, (5.12) can be well approximated by

$$|\widehat{\pi}(X_i, X_j)| > \frac{z_{\alpha/2}}{\sqrt{z^2_{\alpha/2} + (n - m + 1)}}, \tag{5.13}$$

where z is the standard normal distribution.

In the graphical modeling literature, other tests for significance of partial correlations are used. For example, Whittaker (1990, p. 189) considers a test procedure based on the large sample properties of maximum likelihood estimates which leads to the asymptotically equivalent large sample distribution

$$-n \ \log[1 - \widehat{\pi}(X_i, X_j)^2] \sim \chi^2_1.$$

Tests like the ones described above strictly apply only to testing a single partial correlation where the probability of making a type I error, i.e., of *wrongly* concluding that a value is non-zero, is α. Because we test several partial correlations simultaneously, we are in a multiple testing situation where the probability of making at least one type I error is greater than α. Similarly, the overall probability of making at least one type II error, of *wrongly failing* to conclude that a value is non-zero, is greater than it would be for any single test. In practice, we will use these tests to screen all the partial correlations to identify a preliminary CIG structure. This is similar to the use of the partial autocorrelation function for identifying the AR model order of a univariate time series. The graphs so identified will be used to formulate models which will then be fitted and tested rigorously. We have found that a good strategy to cope with the problem of multiple testing in identifying a CIG is to consider the presence of edges with different significance levels, e.g., 0.01, 0.05 and 0.1, and to indicate these by the lines used in presenting the graph. The use of the 0.01 level reduces the overall type I error and that of the 0.1 level reduces the overall type II error. A recent approach to controlling the error levels is given by Drton and Perlman (2008).

The difference between a CIG and a partial correlation graph is due to the assumption of normality of the set of variables X in the CIG. This gives the presence of edges in a CIG the stronger meaning of conditional dependence rather than partial correlation, as in the latter type of graph. However, from the modeling point of view, they both give an indication of the explanatory variables to include in a linear model: when the variables we are dealing with are not Gaussian, the null hypothesis of the test is not independence but lack of linear predictability.

To illustrate the construction of a CIG, let us consider the example of the Flour price series innovations. The sample partial correlation matrix is

$$\widehat{\pi} = \begin{pmatrix} 1.0000 & & \\ 0.8532 & 1.0000 & \\ 0.0231 & 0.4483 & 1.0000 \end{pmatrix}$$

where the diagonal elements, corresponding to the partial correlations of each variable with itself, are equal to 1 and where only the lower triangular part is needed as the matrix is symmetric.

We can observe that $\widehat{\pi}_{1,2}$, corresponding to the partial correlation of the innovations of $x_{1,t}$ and $x_{2,t}$, given the innovations of $x_{3,t}$, is large (0.8532) while $\widehat{\pi}_{1,3}$ is small (0.0231). We can formally test which partial correlations are significantly different from zero by using (5.13). There are three series of innovations, indicated, respectively, as $e_{1,t}$, $e_{2,t}$ and $e_{3,t}$, and each one of them has 98 observations, so $n = 3$ and $m = 98$. If we set the significance at a level of $\alpha = 0.05$, then $z_{0.05/2} = 1.96$, and we can set the threshold for the significance of partial autocorrelations as

$$|\widehat{\pi}(X_i, X_j)| > \frac{1.96}{\sqrt{1.96^2 + (98 - 3 + 1)}} = 0.196.$$

This confirms our initial impression that $\widehat{\pi}_{1,2}$ is significantly different from zero as $0.8532 > 0.196$ while $\widehat{\pi}_{1,3}$ is not, because $0.0231 < 0.196$. The computed threshold also indicates that $\widehat{\pi}_{2,3}$ is significantly different from zero.

We could consider other levels of significance, e.g., $\alpha = 0.01$ and $\alpha = 0.1$ with corresponding threshold values $|\widehat{\pi}(X_i, X_j)| > 0.254$ and $|\widehat{\pi}(X_i, X_j)| > 0.165$. Hence both $\widehat{\pi}_{1,2}$ and $\widehat{\pi}_{2,3}$ are significant even at a 0.01 level of significance, while $\widehat{\pi}_{1,3}$ is not significant even at a 0.1 level of significance.

All this information is efficiently conveyed in the CIG, where we link variables with significant partial correlations, i.e., between $e_{t,1}$ and $e_{t,2}$ and between $e_{t,2}$ and $e_{t,3}$. In general, we will represent the strength of the significance of the partial correlations by the thickness of the lines representing the edges, so in this case we could indicate a strong partial correlation (significant at $\alpha = 0.01$) with a thick continuous line. We will indicate a mid-strength partial correlation (significant at $\alpha = 0.05$) with a continuous line and a weak partial correlation (significant at $\alpha = 0.1$) with a broken line. No link will appear for a partial correlation not significant at any of the levels α considered.

In this example, all the significant partial correlations are strong, and consequently all the edges are represented by thick continuous lines. Figure 5.5 shows the resulting graph for the innovations of the Flour price series. Because the three variables are all normal, the presence of edges indicates conditional dependence among the variables and the graph is a CIG.

A note of caution should be attached to estimation of the CIG. The significance of an estimated partial correlation is equivalent to that of a t-value

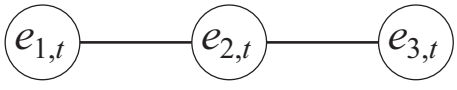

Figure 5.5 *Estimated conditional independence graph of the VAR(2) innovations of the Flour price series.*

in a regression and can only be interpreted as indicating whether or not a variable should be included, given the other variables in the model. If a variable (equivalently a link in the CIG) is removed because of a low absolute t-value and the regression re-fitted, the t-values of the coefficients of the remaining variables will in general change, possibly substantially if there is high collinearity in the data. So a second variable with a relatively low absolute t-value in the original regression may acquire a high absolute t-value when the first is removed. This will be an important consideration when we come to fitting the DAG structures that we identify on the basis of a CIG. Nevertheless, we will take the edges in the estimated CIG as a first indication of the dependency within the set of variables considered. We will go on to model these dependencies using regression equations described by a directed acyclic graph. But having fitted a directed acyclic graph model, we will carry out diagnostic correlation checks on the residuals, designed to detect inadequacy in the model. So if we note significant correlation remaining, we will use this to indicate possible extensions of the model with new edges. This approach to modeling dependence is in the spirit of estimation and diagnostic checking advocated by Box and Jenkins (1970).

In the next section, we illustrate how a CIG, such as shown in Figure 5.5, enables us to draw conclusions about possible DAG representations of the variables.

5.4 Interpretation of CIGs

In this section, we will restrict ourselves to the non-time series context of a set of random variables such as observed time series innovations, for which we have a complete CIG. We will show first how the information in a CIG can in some examples help us to directly determine a DAG representation of the variables. We then explain how, more generally, we can determine which DAG representations are consistent with a given CIG, and so by exploring these possibilities, find a suitable model.

We start with the example of the CIG in Figure 5.5 for which the predictive interpretations of the graph may be set out explicitly. Consider in turn the interpretation for $e_{1,t}$, $e_{2,t}$ and $e_{3,t}$. The general prediction equation for the first of these is

$$e_{1,t} = \beta_{1,2}e_{2,t} + \beta_{1,3}e_{3,t} + \alpha_{1,t}, \tag{5.14}$$

where $\alpha_{1,t}$ is independent of the predictors $e_{2,t}$ and $e_{3,t}$. But because there is no link between $e_{1,t}$ and $e_{3,t}$, i.e., their partial autocorrelation is zero, we may *omit* $e_{3,t}$ from this equation and write

$$e_{1,t} = \beta_{1,2}e_{2,t} + \alpha_{1,t}. \tag{5.15}$$

For the second variable,

$$e_{2,t} = \beta_{2,1}e_{1,t} + \beta_{2,3}e_{3,t} + \alpha_{2,t}, \tag{5.16}$$

for which there is no simplification because $e_{2,t}$ is linked to both $e_{1,t}$ and $e_{3,t}$. However, the third variable is similar to the first, giving

$$e_{3,t} = \beta_{3,2} e_{2,t} + \alpha_{3,t}. \tag{5.17}$$

The key to constructing a DAG representation for the variables is that they must be ordered in such a way that each variable is only predicted by others which are lower in the ordering. This is necessary so that their joint distribution may be expressed as the product of conditional distributions: for example, for random variables X, Y and Z, the joint distribution may be expressed as

$$f(x, y, z) = f(x) f(y|x) f(z|x, y). \tag{5.18}$$

In the normal linear context (with mean corrected variables) this means specifying a marginal variance for X, the regression equation for Y on X with a residual variance and the regression equation for Z on X and Y with a residual variance.

Applying this to $e_{1,t}$, $e_{2,t}$ and $e_{3,t}$, we require the marginal variance of $e_{1,t}$, a regression of $e_{2,t}$ on $e_{1,t}$ and a regression of $e_{3,t}$ on $e_{1,t}$ and $e_{2,t}$. But we know from (5.17) that $e_{1,t}$ can be omitted from this last equation, giving

$$\begin{aligned} e_{1,t} &= a_{1,t} \\ e_{2,t} &= \theta_{2,1} e_{1,t} + a_{2,t} \\ e_{3,t} &= \theta_{3,2} e_{2,t} + a_{3,t}. \end{aligned} \tag{5.19}$$

Note that we have used different notation for the coefficients and error terms from those in (5.14), (5.16) and (5.17) above, because $a_{1,t}$, $a_{2,t}$ and $a_{3,t}$ are now *orthogonal* residuals, a property not shared by the regression errors in (5.14), (5.16) and (5.17), which are in general correlated. The DAG representation for these regressions is shown in Figure 5.6.

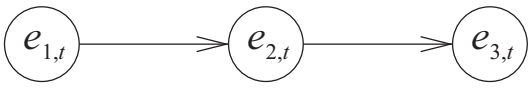

Figure 5.6 *A possible DAG representation of the VAR(2) innovations of the Flour price series.*

However, we may instead take $e_{2,t}$ lowest in our ordering of the variables, followed by $e_{1,t}$ then $e_{3,t}$. This simply interchanges the roles of these variables in the first two equations of (5.19) and is represented in Figure 5.7.

A third possibility is to reverse the original ordering, which would reverse the direction of both arrows in Figure 5.6. However, these are the only three DAGs consistent with the CIG in Figure 5.5. For example, the DAG shown in Figure 5.8, obtained by reversing both arrows in Figure 5.7, *cannot* be deduced from the CIG.

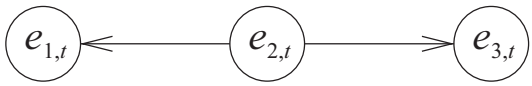

Figure 5.7 *A further DAG representation of the VAR(2) innovations of the Flour price series.*

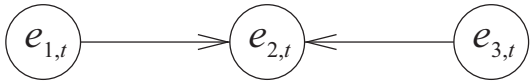

Figure 5.8 *An incompatible DAG representation of the VAR(2) innovations of the Flour price series.*

The dependence of $e_{2,t}$ on $e_{1,t}$ and $e_{3,t}$, as shown, could only be obtained by taking $e_{2,t}$ as highest in the ordering, but then a dependence of either $e_{3,t}$ on $e_{1,t}$, as shown in Figure 5.9, or the reverse, could not be avoided. The DAG in Figure 5.9 could be used to represent the joint distribution of the variables, but fails to take advantage of the simplicity of structure revealed in the CIG.

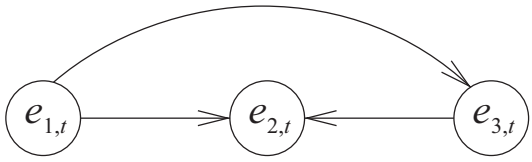

Figure 5.9 *A DAG representation of the VAR(2) innovations of the Flour price series that fails to take advantage of the simplicity of the CIG.*

5.5 Properties of CIGs

The previous illustration with just three variables was particularly simple. For a general CIG, there are several properties that are useful for interpreting its structure and determining the possible DAG models by which this structure may be explained. A useful notion is the set of *neighbors* of a node (or variable). These are simply the variables to which that node is linked. The *pairwise Markov property* of a CIG is that by which it is constructed—any two variables that are not linked are conditionally independent given all the remaining variables. There are two further properties that may be deduced from this:

- The *local Markov property* is that each variable, given its neighbors, is conditionally independent of all the remaining variables.
- The *global Markov property*. This generalizes the pairwise property to the case when we consider two groups (or blocks) of variables. If there are no

links between any pair of variables, one taken from each block, then the two blocks are conditionally independent of each other given all the remaining variables.

The local property is a special case of the global property, which may be deduced from the pairwise property. In the context of the global property, the two blocks are said to be *separated* by the remaining variables, and the result is known as the separation theorem. Its general proof is not simple, but it is very straightforward if all the variables are assumed to be jointly normal. Whittaker (1990) provides a general proof with reference to original sources and acknowledgment of the developers of these ideas. We illustrate the properties using the CIG shown in Figure 5.10. The neighbors of X_2 are X_1, X_3 and X_4, so the local Markov property states that conditional upon these, X_2 is independent of X_5 and X_6. This independence is not just pairwise, between the pair X_2, X_5 and the pair X_2, X_6, but between X_2 and the variables $\{X_5, X_6\}$ taken jointly, which is in general a stronger statement (although not for multivariate normal variables). The global Markov property states that $\{X_1, X_2\}$ are independent of $\{X_5, X_6\}$ conditional upon $\{X_3, X_4\}$. This has the useful consequence that X_2 alone is independent of $\{X_5, X_6\}$ conditional upon $\{X_3, X_4\}$. These properties then enable us to build up one

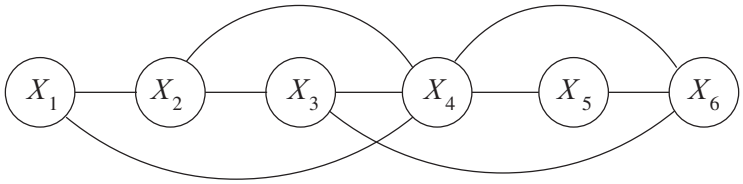

Figure 5.10 *A CIG illustrating local and block independence.*

possible model for these variables, represented by the DAG in Figure 5.11, based upon a choice of ordering from X_1 as the highest through X_6 as the lowest. There is no simplification of the relationship between X_4, X_5 and X_6, but X_5 can be omitted as an explanatory variable for X_3, since it is separated from $\{X_1, X_2, X_3\}$ by $\{X_4, X_6\}$. By the earlier argument, both X_5 and X_6 can be omitted as explanatory variables for X_2. The only explanatory variables for X_1 are X_2 and X_4, because these are its neighbors.

Given the CIG between a set of variables, the task of formulating a DAG representation of the relationship between them is considerably enhanced by use of a theorem, known as the *moralization theorem* or *moralization rule*, by which we may derive the CIG that is implied by any proposed DAG. It is then possible to postulate DAG representations and check whether they are consistent with the given CIG. The theorem gives the following simple steps for deriving the CIG:

1. For each node of the DAG, insert an undirected edge between all pairs of its parent nodes, unless they are already linked by an edge.

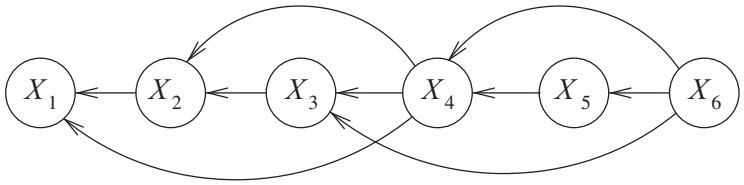

Figure 5.11 *A DAG representation of a model that may be deduced from the CIG in Figure 5.10.*

2. Replace each directed edge in the DAG by an undirected edge.

The first step is described as *marrying the parents*, and hence the construction of the CIG is called *moralization* of the graph (Lauritzen and Spiegelhalter, 1988).

As an example, consider the DAG in Figure 5.11. In this example, moralization introduces no new edges and the resulting CIG is just that in Figure 5.10. However, consider next a trimmed version of this DAG, as shown in Figure 5.12. Moralization of the parents of X_1 will lead to a link between X_2 and X_4. Also, moralization of the parents of X_4 leads to a link between X_5 and X_6. The resulting CIG is therefore the same as that in Figure 5.10. In practice, it is not always straightforward to formulate a DAG, or set of DAGs, that are consistent with a given CIG, and there are simple examples of CIGs with which no DAG is consistent. This can occur due to omission of important variables from the set considered. In the multivariate normal case, it is always possible to fit a regression of each variable on those lower in any chosen ordering, to obtain a model represented by a DAG and which will reproduce the partial correlation graph from which the CIG is constructed. However, there is no reason why such a DAG should reflect any sparse structure in the CIG.

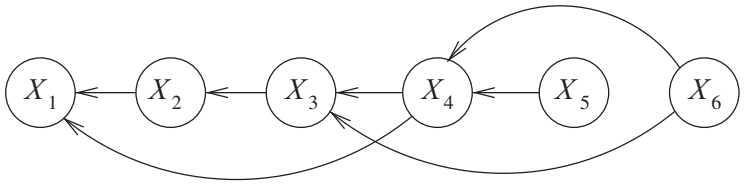

Figure 5.12 *A DAG representation of a reduced model for which the corresponding CIG is also that shown in Figure 5.10.*

In general, however, the moralization theorem can be very helpful in selecting one or more DAG representations of the variables and excluding certain possibilities. For example, from the DAG in Figure 5.11, it is possible to remove either the edge between X_4 and X_6 or that between X_5 and X_6 and retain consistency with the CIG in Figure 5.10, but not to remove both together. For the Flour price innovations, the moralization theorem tells us

immediately that the DAG in Figure 5.8 is not consistent with the CIG in Figure 5.5, because it implies a moralization link between $e_{1,t}$ and $e_{3,t}$. In the case of multivariate normal variables, it is possible to determine some simple quantitative rules for the magnitudes of links in the partial correlation graph that arise due to moralization. These would, for example, help to suggest whether the edge between X_4 and X_6 or that between X_5 and X_6 might be a moralization link in the CIG in Figure 5.10. We set out these rules in Section 5.8, but reproduce the most simple of them here. The others extend this rule to situations where the three variables have parents and children.

The DAG on the left of Figure 5.13 represents a variable z dependent on the independent variables x and y. The CIG on the right represents the moralized CIG. The partial correlation of the moralized link in this graph is given in terms of the other two by the product rule, which is developed further in Section 5.8:

$$\pi(x, y) \;=\; -\pi(x, z) \times \pi(y, z). \tag{5.20}$$

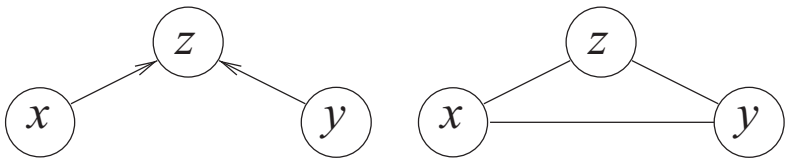

Figure 5.13 *A simple directed graph and its corresponding conditional independence graph.*

In practice, the selection between different possible DAG explanations of an estimated CIG is assisted by comparing the goodness of fit of the estimated models and testing for significance of the coefficients that correspond to links in the competing models. We illustrate this procedure in the next section.

5.6 Estimation and selection of DAGs

We will continue to work in the context of a random sample from a multivariate normal distribution, the innovations from the Flour price series providing an appropriate example which we use for illustration. The main ideas of this section are readily extended to the time series context, including lagged values, in the next section. As in Section 5.3, let the data matrix $\mathbf{X}$ with elements $x_{i,j}$ consist of columns $x_1, x_2, \ldots, x_m$ of the mean corrected samples of the variables $X = \{X_1, X_2, \ldots, X_m\}$. Any proposed DAG representing these variables specifies a regression equation for each variable in terms of a subset of the remaining variables, with the assumption that the errors from any one regression are uncorrelated with those from another. To fit the DAG, we therefore carry out ordinary least squares (OLS) regression of each data column x_i on the explanatory columns specified by the DAG. This provides estimates of the coefficients corresponding to each link in the DAG, and their t value,

and columns a_i of residuals constituting a matrix A. Figure 5.14 shows one of the DAG models considered for the Flour price series innovations, with the estimated coefficients and t values (in parentheses) adjacent to the links.

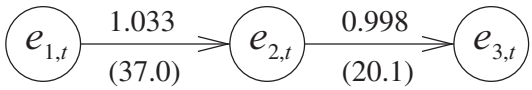

Figure 5.14 *A fitted DAG model of the VAR(2) innovations of the Flour price series, with the estimated coefficients and t values (in parentheses) adjacent to the links.*

The estimated coefficients are clearly highly significant. The next step is to assess the adequacy of this model as a representation of the relationships between the variables. For this purpose, we first use the likelihood of the model and its associated information criteria. In practice, we will use the deviance of the model, i.e., minus twice the log likelihood, which is given by

$$n\left(\log s_1 + \log s_2 + \log s_3\right) \tag{5.21}$$

where s_1, s_2 and s_3 are the sample variances of the residuals (*not* corrected for degrees of freedom)

$$s_i = \frac{1}{n}\sum_{t=1}^{n} a_{i,t}^2. \tag{5.22}$$

The difference between this and the likelihood of the saturated model, which also includes $e_{1,t}$ in the regression for $e_{3,t}$, we will call the deviance reduction D. We will compare this with the difference in degrees of freedom, k, between the proposed model and the saturated model, which is just 1 in this example. For a formal test of the null hypothesis that coefficients of the omitted links in the DAG are all zero, we can refer D to the chi-squared distribution on k degrees of freedom. The information criteria that we will use are defined by $D+2k$ for the AIC, $D+2(\log\log n)k$ for the HQC and $D+(\log n)k$ for the SIC. The values of these for this example are displayed in Table 5.1. The deviance difference is less than the degrees of freedom (df), so there is no reason not to prefer the proposed model to the saturated model. The information criteria are all negative, indicating preference for the proposed model. The further step

Table 5.1 *Likelihood assessment of the DAG in Figure 5.14.*

df k	deviance D	AIC	HQC	SIC
1	0.05	−1.95	−2.99	−4.53

to assess the adequacy of the fitted model is to check the cross-correlations between the residuals, as shown in Table 5.2.

The correlation between $a_{1,t}$ and $a_{2,t}$ is necessarily zero as a result of the regression, but the other two are not. The approximate standard error of

Table 5.2 *Correlations between residuals of the DAG fitted to the Flour price innovation series.*

	$a_{1,t}$	$a_{2,t}$	$a_{3,t}$
$a_{1,t}$	1.000		
$a_{2,t}$	0.000	1.000	
$a_{3,t}$	0.006	−0.022	1.000

these two correlations is bounded by $1/\sqrt{n} = 0.101$, and we deem them to be acceptable, i.e., not suggestive of any model inadequacy.

In Section 5.4 we showed that the DAG in Figure 5.7 was also consistent with the estimated CIG of these variables, and so was the DAG in Figure 5.6 with the edge directions reversed. If we fit these, we find that they give exactly the same deviance and information criteria as in Table 5.1. We say that these models are likelihood equivalent—they imply exactly the same distributional properties of the variables, although the residual series are different because of the different orderings of the variables in the DAG. All saturated models are also likelihood equivalent to each other, so any one of them may be specified to obtain the saturated model deviance. However, on fitting the model shown in Figure 5.8, which is *not* consistent with the CIG, a deviance difference of 139.6 shows how extremely inadequate is that representation of the dependence between the variables. This very simple example has been a platform for introducing the main ideas, which we will now demonstrate in the time series context.

5.7 Building a structural VAR, SVAR

There are six possible saturated DAGs representing the relationships between three variables, corresponding to the number of distinct orderings. Consideration of the CIG of the Flour prices series innovations reduced the number under consideration to three. Each of those three models extends naturally to an SVAR(2) model for the original series, in which the dependency between the current variables is exactly the same as for that between the innovations, but with the additional dependence on all values up to lag 2. Thus the innovations model in Figure 5.14 extends to the SVAR in Figure 5.15. When fitted, the estimated coefficients between current variables in the extended SVAR(2) are exactly the same as for the innovations model, and so are the deviance and information criteria.

To continue the analysis, we now estimate the CIG between all the current values and the variables lagged up to the model order of 2. Recall that this order was determined by application of the AIC to the saturated VAR. The implication is that, given the variables up to lag 2, no variables at a greater lag have predictive value. However, we must restrict the lags used to construct the CIG to those up to the model order. If we were to include values up to greater lags, the partial correlations would in general be reduced in magnitude (see

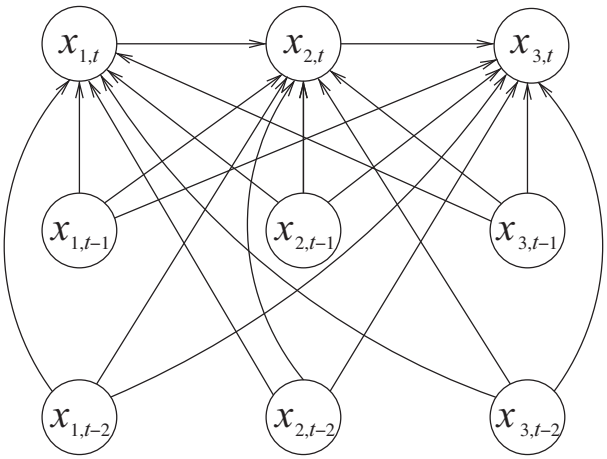

Figure 5.15 *The SVAR(2) series model equivalent to the innovations model in Figure 5.14.*

Chapter 9). In fact, we need only the partial correlations between the current variables and between the current and lagged variables. These correspond to the dependencies which are considered for inclusion in the SVAR. We do not need the partial correlations between any pair of lagged values, although these will be produced as part of the calculations.

To estimate the CIG, we assemble a data matrix $\boldsymbol{X}$ in which the columns in general consist of $\{x_{i,t-u}; t = p+1, \ldots, n\}$ for $i = 1, \ldots, m$ and $u = 0, \ldots, p$, where p is the model order. The estimated CIG is then constructed exactly as before, and from this we extract the required values. For the Flour price series, these are shown in Table 5.3.

The critical values for significance at the levels 0.10, 0.05 and 0.01 are, respectively, 0.171, 0.202 and 0.262. These are based on transforming the critical

Table 5.3 *Partial autocorrelations, up to lag 2, between the Flour price series.*

	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$
$x_{1,t}$			
$x_{2,t}$	**0.853**		
$x_{3,t}$	0.023	**0.448**	
$x_{1,t-1}$	**0.452**	**−0.497**	0.130
$x_{2,t-1}$	**−0.288**	**0.522**	**−0.402**
$x_{3,t-1}$	−0.012	**−0.299**	**0.658**
$x_{1,t-2}$	**0.478**	**−0.301**	−0.132
$x_{2,t-2}$	**−0.412**	**0.264**	0.054
$x_{3,t-2}$	−0.036	0.036	0.058

values of the normal distribution, as described in Section 5.3, rather than the t distribution, not only because the degrees of freedom are sufficiently high, but because the small sample theory underlying the t distribution does not fully extend to the context of multivariate time series regression. However, the asymptotic use of the normal distribution is justified; see Reale and Tunnicliffe Wilson (2002) and Tunnicliffe Wilson and Reale (2008). It will be noted that there are 13 significant values according to these critical values, shown in bold type in the table, and they are all significant at the 0.01 level. Note that the partial correlations between current variables are necessarily identical to those between the innovations.

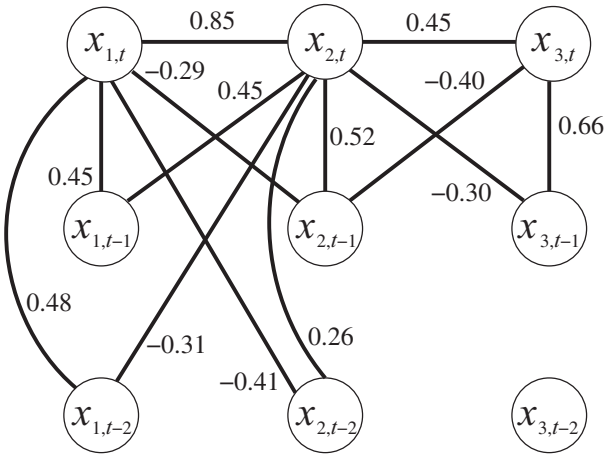

Figure 5.16 *The estimated CIG constructed for the Flour price series with partial correlations shown adjacent to the links.*

Figure 5.16 shows the resulting CIG. The links between current and lagged values provide further information beyond that in the CIG for the innovations alone for identifying the directions of dependence between current variables in the DAG representation of an SVAR model for the series. They also inform us about the links between current and lagged values in the DAG:

1. A link is certainly required between $x_{2,t}$ and $x_{3,t}$ in this DAG. There is strong evidence that this is *not* in the direction $x_{2,t} \leftarrow x_{3,t}$, because this would lead us to expect moralization links in the CIG, between $x_{3,t}$ and all the other parents of $x_{2,t}$ in a possible DAG representation. Three strong candidates for such parents, which have no links with $x_{3,t}$, are the lagged values $x_{1,t-1}$, $x_{1,t-2}$ and $x_{2,t-2}$, and we would have to suppose that their links with $x_{2,t}$ all arose from moralization if our argument were to fail. Our working hypothesis is therefore to assume the direction of this link to be $x_{2,t} \rightarrow x_{3,t}$.
2. The link between $x_{2,t}$ and $x_{3,t-1}$, with a partial correlation of -0.299, is

now a strong candidate for explanation as a moralization link. Applying the product rule, this partial correlation is very close to -0.295, which is the negative product of the partial correlations of 0.448 and 0.658 between, respectively, the hypothesised parents $x_{2,t}$ and $x_{3,t-1}$ and their common child $x_{3,t}$.

The result of these arguments is that $x_{3,t}$ depends in a simple manner on only $x_{2,t}$, $x_{2,t-1}$ and its own lagged value $x_{3,t-1}$. The series $x_{1,t}$ and $x_{2,t}$ are *not* dependent on $x_{3,t}$ at all, and their interdependence appears to be qualitatively symmetrical in the CIG of Figure 5.16. There are two likelihood equivalent DAG interpretations, in both of which all the links between these two series and their lagged values are included. The only difference lies in the choice of direction of the dependence between $x_{1,t}$ and $x_{2,t}$. However, in the model where this direction is chosen to be $x_{1,t} \to x_{2,t}$, the coefficient of the link $x_{1,t-1} \to x_{1,t}$ in the estimated DAG has a t value of 0.95. We therefore choose the direction $x_{1,t} \to x_{2,t}$ and exclude the link $x_{1,t-1} \to x_{1,t}$, which therefore suggests that the corresponding link in the CIG is due to moralization. This gives the more parsimonious SVAR as that represented in Figure 5.17.

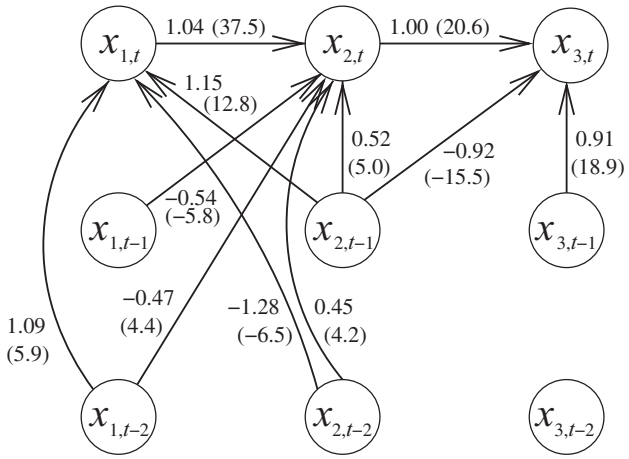

Figure 5.17 *The DAG representing a parsimonious SVAR for the Flour price series, with estimated model coefficients, and t values in parentheses, adjacent to the links.*

This model was again estimated by OLS regression of each current variable in turn, upon the explanatory variables indicated by the DAG, with the regression vectors taken from the same data matrix of current and lagged values used in the construction of the CIG. The estimated coefficients and their t values are shown adjacent to the corresponding links in Figure 5.17. Re-estimation of the model using full maximum likelihood only has a small gain in efficiency for a series of this length. This comes from making use of the relatively small amount of extra information in the first two months of

data that were necessarily trimmed from the vectors of current variables in the data matrix. The maximum likelihood estimates, only slightly different from the OLS estimates, are actually shown in (5.23), which displays the coefficient matrices Φ_0, Φ_1 and Φ_2 of the SVAR as used in the model equation (5.1).

$$\begin{matrix} \Phi_0 & & & \Phi_1 & & & \Phi_2 & & \end{matrix}$$
$$\begin{pmatrix} 1 & 0 & 0 \\ -1.04 & 1 & 0 \\ 0 & -1.01 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1.16 & 0 \\ -0.54 & 0.52 & 0 \\ 0 & -0.91 & 0.89 \end{pmatrix} \begin{pmatrix} 1.14 & -1.13 & 0 \\ -0.45 & 0.43 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.23}$$

Table 5.4 gives the values of the deviance difference and information criteria relative to the saturated model as estimated by OLS. This gives 0.11 for the p-value of testing the null hypothesis that the omitted coefficients are zero. The information criteria also favor the chosen SVAR.

Table 5.4 *Likelihood assessment of the DAG in Figure 5.17.*

df k	deviance D	AIC	HQC	SIC
10.00	15.69	−4.31	−14.77	−30.16

The SVAR model, for these series, lends itself better to interpretation than the canonical VAR. First, the representation for $x_{3,t}$ can be re-arranged with only a minor approximation as a simple regression on $x_{2,t}$ with a univariate AR(1) disturbance:

$$\begin{aligned} x_{3,t} &= x_{2,t} + n_t \\ n_t &= 0.91 n_{t-1} + a_t. \end{aligned} \tag{5.24}$$

Second, again with only minor approximation, we can represent

$$x_{1,t} = x_{1,t-2} + (x_{2,t-1} - x_{2,t-2}) + a_{1,t}, \tag{5.25}$$

in which the current value of $x_{1,t}$ is the value two months previously, but with a correction equal to the change in price of $x_{2,t}$ over the previous two months. Third, we can approximate

$$x_{2,t} = x_{1,t} + 0.5\,(x_{2,t-1} + x_{2,t-2}) - 0.5\,(x_{1,t-1} + x_{1,t-2}) + a_{2,t}, \tag{5.26}$$

in which the current value of $x_{2,t}$ is the current value of $x_{1,t}$ corrected by the difference between the average price of the two series over the previous two months.

A structural model such as we have constructed does not take into account the fact that the Flour price series also reflect cycles of general economic activity which appear to be evident in their plots, and the model we have fitted does not capture any such cyclical behavior. However, the model has

a simple interpretable structure. With just 11 coefficients, 10 fewer than a saturated model (and three variance parameters), it competes well with other published models for these series. A structural VARMA(1,1) model has also been developed for these series in Oxley et al. (2009). This also has a simple interpretation, but the construction of a VARMA model is in general much more difficult, and we do not pursue this line of model development in this book.

The construction of the SVAR model is not completed until we have carried out sensible checks. As before, we check the cross-correlations between current residuals, which are shown in Table 5.5. We also check the lagged cross-correlations shown in Figure 5.18, with error limits of $\pm 2/\sqrt{n}$. These limits give a good indication of the adequacy of the cross-correlations, although they do not allow for the effects of parameter estimation.

Table 5.5 *Correlations between residuals of the final SVAR(2) for the Flour price series.*

	$a_{1,t}$	$a_{2,t}$	$a_{3,t}$
$a_{1,t}$	1.000		
$a_{2,t}$	0.000	1.000	
$a_{3,t}$	−0.019	−0.023	1.000

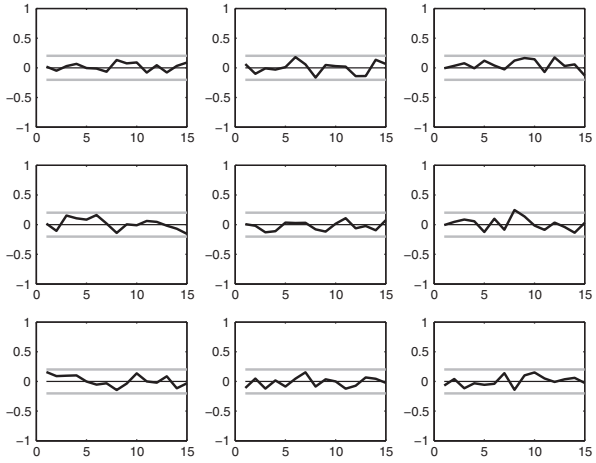

Figure 5.18 *Residual auto- and cross-correlations of the final SVAR(2) model for the Flour price series. Nominal two standard error limits about zero are shown by the gray lines.*

There are no indications of model inadequacy to be gleaned from these checks. Our final comment on this model is that it is stationary, though the

reciprocal roots of the operator in the reduced VAR model are 0.9879, 0.9228, 0.9083, −0.4643, 0.2680 and 0.0. The first of these is close to the boundary of stationarity. We pointed out in Chapter 2 that a sparse VAR(p) model fitted by solving Yule–Walker equations does not necessarily reproduce any lagged correlations and it is not necessarily stationary. The same is true of the sparse SVAR models that we have explored by graphical modeling and fitted by OLS. However, what we have aimed to do is to reproduce all the partial correlations up to lag p, by selecting terms in the sparse model that capture the pattern of zero and non-zero partial correlations. As a final check, in Table 5.6 we compare the partial correlations that may be computed for the fitted model with the sample values used to identify the model. The two entries that arise by moralization are between $x_{1,t}$ and $x_{1,t-1}$, and between $x_{2,t}$ and $x_{3,t-1}$. The general impression is that the model captures the sample partial correlations quite well. There is strong correlation both between and within these series. However, the graphical modeling approach we have followed has overcome the potential problems of multicollinearity and generated a model that has simple and interpretable structure.

Table 5.6 *Each pair of columns shows, on the left, the sample partial autocorrelations up to lag 2 between the Flour price series, and on the right, the non-zero partial correlations of the selected model shown alongside. The bold numbers indicate a non-zero model value, with the corresponding sample value to its left.*

	$x_{1,t}$		$x_{2,t}$		$x_{3,t}$	
$x_{1,t}$						
$x_{2,t}$	**0.853**	**0.861**				
$x_{3,t}$	0.023		**0.448**	**0.458**		
$x_{1,t-1}$	**0.452**	**0.482**	**−0.497**	**−0.442**	0.130	
$x_{2,t-1}$	**−0.288**	**−0.329**	**0.522**	**0.524**	**−0.402**	**−0.368**
$x_{3,t-1}$	−0.012		**−0.299**	**−0.308**	**0.658**	**0.672**
$x_{1,t-2}$	**0.478**	**0.486**	**−0.301**	**−0.385**	−0.132	
$x_{2,t-2}$	**−0.412**	**−0.439**	**0.264**	**0.337**	0.054	
$x_{3,t-2}$	−0.036		0.036		0.058	

Another check that is advisable, where appropriate, is to compare the spectral properties of the fitted model with smoothed spectral estimates of the series, and a final check is to hold back some data at the end of the series to compare with out-of-sample forecasts of the same points made using the model. We will illustrate these checks as they are applied to the further example in Section 5.10.

We end this section by summarizing the steps by which we build our SVAR:

- Use the AIC to determine the order of a saturated VAR model which well represents the series.
- Construct the sample CIG of the estimated innovation series derived from this VAR model.

- Use this to explore possible DAG representations of the innovation series.
- Estimate these DAG models and check their likelihood criteria and residual cross-correlations.
- Construct the sample CIG of the series and its values lagged up to the determined order.
- Use this to explore possible DAG representations of an SVAR model for the series, basing the selected dependencies between current variables on those explored for the innovation series.
- Estimate these DAG models and check their likelihood criteria and lagged residual auto- and cross-correlations. If necessary, include or exclude terms from the model to achieve an adequate sparse model.
- Check that the CIG implied by the model reasonably well matches the sample CIG; apply other comparisons of the model and sample spectral properties and examine an out-of-sample set of forecasts from the model to assess its consistency in this respect.

5.8 Properties of partial correlation graphs

The main aim of this chapter is to show how the estimated partial correlation matrix may inform us in selecting the dependencies between the current and lagged terms of a structural vector autoregressive model. An important tool has been the rule for moralization of an hypothesized DAG representation of these variables, to derive the implied CIG of the model, for comparison with the estimated CIG. One *converse* application of the rule is that a pattern in the CIG, of the form shown on the left of Figure 5.19, *excludes* the pattern in the DAG shown on the right of the same figure as a possible interpretation. The three DAGs in which one or both of the directions of the links are reversed

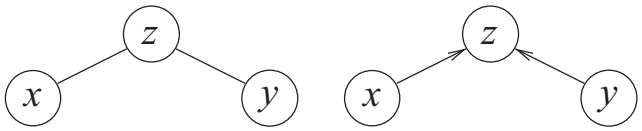

Figure 5.19 *A simple conditional independence graph and the directed graph interpretation that it* excludes.

are consistent with the CIG that is shown. Because the direction of a link from the past is determined by the arrow of time, this simple result can often give a useful indication of the direction of a link between current variables, as shown in the example of the Flour price series.

It will be useful before our next example in Section 5.10, to look further at the relationship between the model structure and the partial correlation matrix that underlies the moralization rule. In this section, we list some quantitative features which may assist in its intelligent application, particularly in