Fitting Constrained Vector Autoregression Models

Tucker McElroy and David Findley

Abstract This paper expands the estimation theory for both quasi-maximum likelihood estimates (QMLEs) and Least Squares estimates (LSEs) for potentially misspecified constrained VAR(p) models. Our main result is a linear formula for the QMLE of a constrained VAR(p), which generalizes the Yule-Walker formula for the unconstrained case. We make connections with the known LSE formula and the determinant of the forecast mean square error matrix, showing that the QMLEs for a constrained VAR(p) minimize this determinant but not the component entries of the mean square forecast error matrix, as opposed to the unconstrained case. An application to computing mean square forecast errors from misspecified models is discussed, and numerical comparisons of the different methods are presented and explored.

Key words: ARIMA, forecasting, frequency domain, vector time series

1 Introduction

An extremely popular vector time series model is the Vector Autoregression of order p, or VAR(p) for short. Constraining a particular coefficient to be zero can affect the estimation of this model considerably, and is an important tool for assessing the impact of related series on short-term forecasting. This paper expands the estimation theory for both quasi-maximum likelihood estimates (QMLEs) and Least Squares estimates (LSEs) for potentially misspecified constrained VAR(p) models. Our main result is a linear formula for the QMLE of a constrained VAR(p), which generalizes the Yule-Walker formula for the unconstrained case; then we connect this with the

T. McElroy
U.S. Census Bureau, e-mail: tucker.s.mcelroy@census.gov

D. Findley
U.S. Census Bureau, e-mail: david.f.findley@census.gov

known LSE formula, concluding that the LSEs and QMLEs retain certain forecasting optimality properties even when the fitted model is misspecified.

The QMLE for a constrained VAR(p) minimizes the Total Innovation Variance (TIV) – i.e., the determinant of the forecast mean square error matrix – and the LSE is asymptotically equivalent to the QMLE. Hence, these estimates provide the best possible parameters – for the given model – with respect to TIV, even when the model is misspecified. TIV has a long history as an overall assessment of predictive capacity (Wilks (1932), Whittle (1953)), and is closely connected to the Kullback-Leibler divergence between model and truth; this determinant, once it is properly scaled, provides the data dependent portion of the maximized Gaussian likelihood function. The topic has been treated by many authors (including Akaike (1969, 1974)), summarized in Taniguchi and Kakizawa (2000); also see Maïnassara and Francq (2011).

Another feature of the QMLE for unconstrained VAR(p) models is that the resulting fitted model is always stable, whereas this need not be true for LSEs. Opinions vary over the desirability of this trait, as discussed in Lütkepohl (2006). If the true data process is stationary, then ensuring the stability of our fitted model is desirable. But if there may be co-integration or explosive behavior present in the data, then using the QMLEs would be misleading – instead we would prefer to use LSEs.

These results provide some motivation for considering QMLEs for fitting constrained VAR models; given that the formulas are just as simple and fast as the LSEs, and the properties are quite similar, practitioners may be interested in computing them. We also note that the same formulas used to compute QMLEs can be used to determine the pseudo-true values (PTVs) that arise when a misspecified constrained VAR(p) is fitted (via Whittle estimation or maximum likelihood estimation (MLE)) to a data process. A PTV is defined informally as that parameter vector (or vectors, as they may be non-unique) to which estimates converge in probability when the model is misspecified. Having a quick way to compute PTVs is helpful for simulation studies of the impact of model misspecification. For example, if one wanted to gauge the Mean Squared Error (MSE) of forecasting from a misspecified model, the PTVs could be plugged into the forecast filter, and the resulting forecast errors determined from analytical calculations (we discuss this application later in the paper).

Since the VAR(p) model is often applied to do forecasting, we also make some connections between the QMLEs for the constrained VAR(p) and the unconstrained case, where the estimates are given by the Yule-Walker (YW) formula. Whereas the YW estimates optimize each entry of the asymptotic one-step ahead forecast MSE matrix, the PTVs in the constrained case only minimize the determinant of this matrix, namely the TIV – which is a weaker property. This suggests that the best we can hope for in the constrained VAR(p) case is to improve forecast MSE in the entangled sense of TIV; while we may minimize TIV, we may not be minimizing the diagonal entries of the forecast MSE matrix! This new and somewhat surprising conclusion is explained in the paper.

Section 2 provides the general theory of the QMLE for constrained VAR models, with connections to the Yule-Walker equations, and the implications to forecasting

discussed. These results are compared to known formulas for the LSEs (Lütkepohl, 2006), with the outcome that we can make the same conclusions about LSEs asymptotically. Section 3 provides numerical illustrations of the LSE, MLE, and QMLE methods for the bivariate VAR(1), the point being to demonstrate how forecasting performance diverges between the methods when the model is misspecified. In this part of the paper we also discuss an application of PTVs to computing h-step ahead forecast MSE from a mis-specified model. A fuller version of this paper is McElroy and Findley (2013), which contains the proofs of results, as well as some additional examples.

2 Theoretical Results

In this section we provide a complete theory of QMLE fitting of constrained VAR models. We begin with some general results about the QMLE method discussed in Taniguchi and Kakizawa (2000), showing that it is sufficient to optimize the TIV. Then we specialize to constrained VAR models, providing an exact solution, and make comparisons to the LSE method.

2.1 General Theory of QMLE

We consider difference stationary processes, and generally follow the treatments of vector time series in Brockwell and Davis (1991), Taniguchi and Kakizawa (2000), and Lütkepohl (2006). Included in our framework are the popular co-integated VAR and VARIMA models used by econometricians, as well as structural VARIMA models. The formulas also cover the case of more unconventional processes that have long-range dependence. For notation we use an underline for every matrix, which for the most part are $m \times m$. The identity matrix is denoted by $\underline{1}_m$. Also in general capital letters refer to composite objects and lower case letters refer to components (such as coefficients); Latin letters refer to random variables/vectors, and Greek letters refer to deterministic quantities (like parameters). Matrix polynomial and power series functions are defined as $\underline{A}(x) = \sum_{k=0}^{p} \underline{a}_j x^j$ with $p < \infty$ or $p = \infty$ as the case may be. We use B for the backshift operator, which sends a time series back in time: $B\mathbf{X}_t = \mathbf{X}_{t-1}$, working on all components of the vector at once. Then the action of $\underline{A}(B)$ on $\mathbf{X}_t$ is understood by linear extension. Also we introduce the following convenient notation for any matrix power series $\underline{A}(x)$: $[\underline{A}]_\ell^j(x) = \sum_{k=\ell}^{j} \underline{a}_k x^k$.

Let us suppose that the data can be differenced to stationarity by application of a degree d differencing polynomial $\underline{\Delta}(B)$; its application to the observed time series $\{\mathbf{X}_t\}$ yields a covariance stationary time series $\{\mathbf{W}_t\}$, i.e., $\underline{\Delta}(B)\mathbf{X}_t = \mathbf{W}_t$. The operator $\underline{\Delta}(B)$ is referred to as the differencing operator, and in general contains both stable and unstable elements that are not easily separated. As discussed in Lütkepohl

(2006), the zeroes of $\det \underline{\Delta}(z)$ include some on the unit circle of the complex plane, and the rest outside.

The series $\{\mathbf{W}_t\}$ is assumed to be stationary with mean vector $\mathbf{m}$, and we further suppose that it is purely non-deterministic. Its lag h autocovariance matrix will be denoted

$$\Gamma(h) = \mathbb{E}[(\mathbf{W}_{t+h} - \mathbf{m})(\mathbf{W}_t - \mathbf{m})'].$$

The spectral density matrix of $\{\mathbf{W}_t\}$ is denoted by $\underline{F}(\lambda)$, and is defined via $\underline{F}(\lambda) = \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-i\lambda h}$. Hence we have the relation $\Gamma(h) = (2\pi)^{-1} \int_{-\pi}^{\pi} \underline{F}(\lambda)\, e^{i\lambda h}\, d\lambda$. We further assume that $\underline{F}(\lambda)$ has full rank for each λ, which will ensure that the forecast error covariance matrix, defined below, is nontrivial; this condition also implies that $\int_{-\pi}^{\pi} \log\det \underline{F}(\lambda)\, d\lambda > -\infty$.

We will consider any model for $\{\mathbf{W}_t\}$ that is invertible, such that a Wold Decomposition (Brockwell and Davis (1991) or Reinsel (1997)) exists, which means that – when the model is true – we can write

$$\mathbf{W}_t = \mathbf{m} + \underline{\Psi}(B)\mathbf{A}_t, \tag{1}$$

where the series $\{\mathbf{A}_t\}$ is mean zero and uncorrelated (but possibly dependent) over time with positive definite covariance matrix $\underline{\sigma}$. Here $\underline{\Psi}(B)$ is a causal power series with coefficient matrices $\underline{\psi}_k$. By the invertibility assumption, we mean the assumption that $\det \underline{\Psi}(z) \neq 0$ for $|z| \leq 1$ and

$$\int_{-\pi}^{\pi} \log\det \left[\underline{\Psi}\left(e^{-i\lambda}\right) \underline{\Psi}'\left(e^{i\lambda}\right) \right] d\lambda = 0. \tag{2}$$

Thus $\underline{\Psi}^{-1}(z)$ is well-defined for $|z| \leq 1$. If our model is correct for the data process, such that (1) holds exactly, then we can write $\mathbf{A}_t = \underline{\Psi}(B)^{-1}[\mathbf{W}_t - \mathbf{m}]$, showing that $\{\mathbf{A}_t\}$ is the linear innovations process of $\{\mathbf{W}_t\}$. The filter $\underline{\Psi}(B)^{-1}$ is called the innovations filter of $\{\mathbf{W}_t\}$.

However, in general any model that we propose is mis-specified, so we cannot assume that (1) holds exactly. Let us consider any causal invertible model, i.e., one with a Wold filter representation $\underline{\Psi}_\xi(B)$, such that this Wold filter is parameterized by a vector $\xi \in \Xi$ associated with the model coefficients, while accounting for any coefficient constraints. Invertibility means that $\det \Psi_\xi(z)$ is nonzero for $|z| \leq 1$ for all $\xi \in \Xi$, where Ξ is assumed to be an open convex set. The filter $\underline{\Psi}_\xi(B)$ therefore satisfies (2). In this paper we are principally interested in so-called separable models, where the parameter ξ does not depend on our parametrization of the innovation variance $\underline{\sigma}$, the covariance of the putative innovations $\{\mathbf{A}_t\}$; for the more general treatment of non-separable models, see Taniguchi and Kakizawa (2000). By specializing to separable models, we can obtain a more focused result.

So assume that ξ is parameterized separately from the distinct entries of the model's innovation covariance matrix. Let ζ denote the vector $\mathrm{vec}\underline{\sigma}$, so that $\underline{\sigma}_\zeta$ refers to our model's innovation covariance matrix. We require this matrix to belong to the set $\mathscr{S}_+$ of all positive definite matrices. Then the full vector of parameters can be written as $\vartheta = [\xi', \zeta']'$, so that the first set of parameters control the Wold

filter $\Psi_\xi(B)$, and the second set of parameters parametrize the innovation covariance matrix $\underline{\sigma}_\zeta$. Then the spectral density of this model can be written as

$$\underline{F}_\vartheta(\lambda) = \underline{\Psi}_\xi(e^{-i\lambda})\,\underline{\sigma}_\zeta\,\underline{\Psi}'_\xi(e^{i\lambda}),$$

and furthermore from (2),

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \underline{F}_\vartheta(\lambda)\,d\lambda = \log\det\underline{\sigma}_\zeta.$$

This last expression is guaranteed to be positive, since the matrix belongs to $\mathscr{S}_+$.

Now because $\underline{\Psi}_\xi(B)$ is invertible, the one-step ahead forecast filter for the differenced series $\{\mathbf{W}_t\}$ is well-defined, and is given by $B^{-1}[\underline{\Psi}_\xi]_1^\infty(B)\,\underline{\Psi}_\xi(B)^{-1}$, as described in McElroy and McCracken (2012). The forecast errors when using such a filter are then given by $\mathbf{E}_t = \underline{\Psi}_\xi(B)^{-1}(\mathbf{W}_t - \mathbf{m})$, whose covariance results in the following important matrix:

$$\Omega(\xi) = \mathbb{E}\left[\mathbf{E}_t\mathbf{E}_t'\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \underline{\Psi}_\xi(e^{-i\lambda})^{-1}\,\underline{F}(\lambda)\,\underline{\Psi}_\xi(e^{i\lambda})^\dagger\,d\lambda. \tag{3}$$

Here $\dagger$ is short for inverse transpose. Note that $\{\mathbf{E}_t\}$ may not be exactly a white noise, because our model is misspecified, or is imperfectly estimated. We label the above matrix as the Forecast Error Variance (FEV) matrix, denoted by $\Omega(\xi)$, the dependence on the parameter ξ being explicit. Note that the FEV is always positive definite, because of our assumption that $\underline{F}(\lambda)$ has full rank for all λ (this can be weakened to having less than full rank for a set of λs of Lebesgue measure zero, which allows us to embrace the possibility of co-integration).

It is reasonable to seek models and parameter values ξ such that the FEV is minimized in an appropriate sense. Because the diagonal entries of the FEV represent forecast mean squared errors (MSEs), it is plausible to minimize any of these diagonal entries, or perhaps the trace of $\Omega(\xi)$. Another approach would be to minimize the determinant of the FEV, although this quantity is difficult to interpret in terms of forecast performance. Note that $\det\Omega(\xi)$ is the TIV defined earlier, and is related to the Final Prediction Error (FPE) of Akaike (1969), a scaled version of the determinant of the estimated innovations variance matrix, based upon results of Whittle (1953). Historically, the work of Akaike (1969) forms the basis for using the FEV determinant as a fitting criterion for VAR models. Whittle (1953) refers to $\det\Omega(\xi)$ as the Total Prediction Variance, adopting terminology from Wilks (1932); we utilize the term Total Innovation Variance (TIV) instead, to emphasize its connection to the innovations process. There are many articles that discuss VAR model selection via the FPE criterion of Akaike (1969), and there have been numerous successful applications in industry and econometrics; see Akaike and Kitagawa (1999) for additional applications.

We now provide a treatment of the connection of QMLE and TIV minimization for separable models (they need not be VAR at this point, but rather any separable model with causal invertible Wold representation), which connects Gaussian maxi-

mum likelihood estimation to minimization of the TIV. The Kullback-Leibler (KL) discrepancy between a true process' spectrum $\underline{F}$ and a putative model spectrum $\underline{F}_\vartheta$ is defined via

$$D(\underline{F}_\vartheta, \underline{F}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \underline{F}_\vartheta(\lambda) + \mathrm{tr}\left\{\underline{F}_\vartheta(\lambda)^{-1}\ \underline{F}(\lambda)\right\} d\lambda.$$

See Taniguchi and Kakizawa (2000) for more exposition. This formula is also valid when the multivariate periodogram $\underline{I}(\lambda) = n^{-1}\sum_{t=1}^n \mathbf{W}_t e^{-i\lambda t}\sum_{t=1}^n \mathbf{W}_t' e^{i\lambda t}$ is substituted for $\underline{F}$, yielding $D(\underline{F}_\vartheta, \underline{I})$. This quantity is related to -2 times the multivariate Gaussian log likelihood, and is more convenient to work with in empirical applications, since no matrix inversions are required for its calculation. In fact, empirical estimates based on this criterion have similar asymptotic properties to Gaussian maximum likelihood estimates.

The definition of a QMLE is a parameter ϑ_I such that $\vartheta \mapsto D(\underline{F}_\vartheta, \underline{I})$ is minimized. The definition of a PTV is a parameter ϑ_F such that $\vartheta \mapsto D(\underline{F}_\vartheta, \underline{F})$ is minimized. The general theory of Taniguchi and Kakizawa (2000) shows that, under suitable conditions on the process and the model (requiring the uniqueness of ϑ_F), that QMLEs are consistent and asymptotically normal for PTVs, and are also efficient when the model is correctly specified. In this case, the PTVs are identical with the true parameters of the process: since $\underline{F} \in \{\underline{F}_\vartheta : \vartheta \in \Xi \times \mathscr{S}_+\}$, there exists some $\widetilde{\vartheta}$ such that $\underline{F} = \underline{F}_{\widetilde{\vartheta}}$, and the PTVs are identical with this $\widetilde{\vartheta}$.

Because QMLEs and MLEs are asymptotically equivalent when the underlying process is Gaussian, PTVs are informative about what parameter estimates are converging to when models are misspecified; this, along with their asymptotic efficiency under correct model specification – and their relative ease of computation – motivates interest in QMLEs (and also PTVs). Now the above formula for KL is general, but in the case of a separable model we have an alternative formula:

$$\begin{aligned} D(\underline{F}_\vartheta, \underline{F}) &= \log\det \underline{\sigma}_\zeta + \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left\{\underline{\sigma}_\zeta^{-1}\,\Psi_\xi(e^{-i\lambda})^{-1}\,\underline{F}(\lambda)\,\Psi_\xi'(e^{i\lambda})^{-1}\right\} \\ &= \log\det \underline{\sigma}_\zeta + \mathrm{tr}\left\{\underline{\sigma}_\zeta^{-1}\,\Omega(\xi)\right\}. \end{aligned} \tag{4}$$

This derivation uses (3) and an interchange of integration and trace. In fact, this derivation does not assume any particular model structure for $\underline{F}$, so we can also obtain an alternative formula for $D(\underline{F}_\vartheta, \underline{I})$ as $\log\det \underline{\sigma}_\zeta + \mathrm{tr}\left\{\underline{\sigma}_\zeta^{-1}\,\widehat{\Omega}(\xi)\right\}$, where $\widehat{\Omega}(\xi)$ is an empirical version of the FEV defined via

$$\widehat{\Omega}(\xi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \underline{\Psi}_\xi(e^{-i\lambda})^{-1}\,\underline{I}(\lambda)\,\underline{\Psi}_\xi(e^{i\lambda})^{\dagger}\, d\lambda.$$

We can then determine the PTVs and QMLEs by the same mathematics: by the appropriate simplification of the derivation of Magnus and Neudecker (1988, p. 317), for any fixed $\xi \in \Xi$ the FEV matrix $\Omega(\xi)$ minimizes $\zeta \mapsto D(\underline{F}_{\xi,\zeta}, \underline{F})$ over all parametrizations such that $\underline{\sigma}_\zeta \in \mathscr{S}_+$. This is appropriate for PTVs; for QMLEs, we

have $\widehat{\Omega}(\xi)$ minimizing $\zeta \mapsto D(\underline{F}_{\xi,\zeta}, \underline{I})$. Recall that the FEV is in $\mathscr{S}_+$ by our full rank assumption on $\underline{F}$; in the case of the QMLEs the empirical FEV can violate this only in the trivial case that the data equals the zero vector[1]. Then from (4) we obtain

$$D\left(F_{\xi,\mathrm{vec}\Omega(\xi)}, \underline{F}\right) = \log\det\Omega(\xi) + m.$$

This is a concentration of the likelihood, analogously to the procedure with univariate time series, and relates KL to TIV. If we minimize the above expression with respect to ξ, and then compute $\Omega(\xi)$ for that optimal ξ, then we have produced the PTV ϑ. Of course, the dimension m is irrelevant to this problem, as is the presence of the logarithm. Therefore, the PTV ξ_F, which we assume exists uniquely in Ξ, satisfies

$$\xi_F = \arg\min_{\xi\in\Xi} \det\Omega(\xi) \qquad \zeta_F = \mathrm{vec}\,\Omega(\xi_F).$$

Our parameter space should be taken to be a compact convex subset Ω of $\Xi \times \mathrm{vec}(\mathscr{S}_+)$ that contains $\vartheta_F = [\xi_F', \mathrm{vec}'\Omega(\xi_F)]'$. In the next section we will demonstrate the existence and uniqueness of such PTVs for constrained VAR models. The treatment for QMLEs follows identically: the concentrated empirical KL equals m plus the log determinant of the empirical FEV, and hence

$$\xi_I = \arg\min_{\xi\in\Xi} \det\widehat{\Omega}(\xi) \qquad \zeta_I = \mathrm{vec}\,\Omega(\xi_I).$$

In summary, we see that the QMLEs and PTVs for ξ are computed by minimizing the empirical and theoretical TIVs, respectively, and then plugging these parameters back into the empirical/theoretical FEV matrix. So whereas the TIV seems to be a non-intuitive quantity in terms of forecast performance, it is actually the right objective function if we wish to obtain statistically efficient parameter estimates in the correct model case. Theorem 3.1.2 of Taniguchi and Kakizawa (2000) gives a central limit theorem for the QMLEs; also see (3.4.25) in Lütkepohl (2006) for the special case of a VAR model, assuming the model is correctly specified.

2.2 Constrained Versus Unconstrained VAR Models

2.2.1 Properties of the Unconstrained Case: Full Optimization

The previous subsection treated general separable models. We now focus on unconstrained VAR models as a further special case. Let $\underline{\phi}$ be a $m \times mp$ dimensional matrix

[1] For any vector a, we have $a'\widehat{\Omega}(\xi)a = (2\pi n)^{-1}\int_{-\pi}^{\pi} |a'\Psi^{-1}(e^{-i\lambda})\sum_{t=1}^n \mathbf{W}_t e^{-i\lambda t}|^2\, d\lambda$, so that the expression equals zero iff $a'\Psi^{-1}(e^{-i\lambda})\cdot\sum_{t=1}^n \mathbf{W}_t e^{-i\lambda t} = 0$ almost everywhere with respect to λ; because both terms in this product are polynomials in $e^{-i\lambda}$, the condition is equivalent to one or the other of them being zero. In the one case that $a'\Psi^{-1}(e^{-i\lambda}) = 0$, we at once deduce that a is the zero vector; in the other case, we have that the discrete Fourier Transform $\sum_{t=1}^n \mathbf{W}_t e^{-i\lambda t} = 0$ for almost every λ, which can only be true if the data is zero-valued.

consisting of the concatenation of the coefficient matrices of $\underline{\Phi}(z) = \underline{1}_m - \sum_{j=1}^p \underline{\phi}_j z^j$. In terms of the notation of the previous section, $\xi = \text{vec}\,\underline{\phi}$ and $\underline{\Psi}_\xi(B) = \underline{\Phi}(B)^{-1}$. The invertibility assumption given above then dictates that $\underline{\Phi}(z)$ must belong to the set $\digamma_p$ of matrix polynomials such that the zeroes of $\det \underline{\Phi}(z)$ satisfy $|z| > 1$.

It will be convenient to introduce a notation for the transposed autocovariance: let $\underline{R}_{1:p+1,1:p+1}$ denote a $m(p+1)$ dimensional square matrix, which is block-Toeplitz with jkth block matrix given by $\Gamma(k-j) = \Gamma'(j-k)$. We can partition $\underline{R}_{1:p+1,1:p+1}$ into its upper left $p \times p$ block $\Gamma(0)$ and its lower right mp dimensional block $\underline{R}_{2:p+1,2:p+1}$, which is also block-Toeplitz (and equal to $\underline{R}_{1:p,1:p}$). The remaining portions are denoted $\underline{R}_{1,2:p+1}$ and $\underline{R}_{2:p+1,1}$. Then it can be shown that

$$\begin{aligned}\Omega(\xi) &= \Gamma(0) - \sum_{j=1}^p \underline{\phi}_j \Gamma(-j) - \sum_{k=1}^p \Gamma(k)\,\underline{\phi}_k' + \sum_{j,k=1}^p \underline{\phi}_j \Gamma(k-j)\,\underline{\phi}_k' \\ &= \Gamma(0) - \underline{\phi}\, R_{2:p+1,1} - R_{1,2:p+1}\,\underline{\phi}' + \underline{\phi}\, R_{1:p,1:p}\,\underline{\phi}' \end{aligned} \tag{5}$$

Our treatment looks at PTVs, but if we replace the true autocovariances $\Gamma(h)$ by sample estimates (the inverse Fourier Transforms of the periodogram I) and write $\widehat{\Omega}(\xi)$, we can apply the same mathematics as derived below, and obtain an identical treatment of QMLEs.

Let us first examine the case of an unconstrained VAR(p) model: we show that the PTV is the solution to the Yule-Walker (YW) equations (a known result), and also that the PTV minimizes each entry of the FEV matrix, not merely its determinant, the TIV (a new result). Noting that by definition ξ_F is a zero of the derivative of the TIV, we compute it via the chain rule:

$$\frac{\partial}{\partial \xi_\ell} \det \Omega(\xi) = \sum_{r,s} \Omega_{(r,s)}(\xi)\, \frac{\partial \Omega_{rs}(\xi)}{\partial \xi_\ell}.$$

See Mardia, Kent, and Bibby (1979). Here $\Omega_{(r,s)}$ is the co-factor of Ω, while Ω_{rs} is just the r,sth entry of the FEV matrix. The chain rule tells us that a *sufficient* condition for the gradient of the FPE to be zero, is that the gradients of Ω_{rs} are zero. That is, it is sufficient to find a solution that optimizes all the coefficient functions of the FEV. This is a stronger property than just minimizing $\det \Omega$, since there might be solutions that minimize the FPE but do not minimize all of the component functions. In the case of a VAR(p) this stronger property holds, which is remarkable and useful. The following result is a slight elaboration, for the perspective of KL discrepancy minimization, of the results of Whittle (1963) for case of full rank $\{\mathbf{W}_t\}$.

Proposition 1 *Let $\{\mathbf{W}_t\}$ be stationary and invertible, with full rank spectral density matrix. Then the PTV $\widetilde{\underline{\phi}}$ for a fitted VAR(p) satisfies the Yule-Walker equations*

$$\sum_{j=1}^p \widetilde{\underline{\phi}}_j \Gamma(k-j) = \Gamma(k),\ 1 \le k \le p, \tag{6}$$

or $\underline{\widetilde{\phi}}\,\underline{R}_{1:p,1:p} = \underline{R}_{1,2:p+1}$. *Furthermore, the corresponding polynomial* $\widetilde{\Phi}(z) \in \mathcal{F}_p$ *and* $\xi_F = vec\,\underline{\widetilde{\phi}}$ *uniquely minimizes* $\xi \mapsto \det \Omega(\xi)$, *with the FEV given by (5). The PTV also minimizes* $\xi \mapsto \Omega_{rs}(\xi)$ *for every* $1 \le r,s \le m$. *The PTV for the FEV is*

$$\underline{\sigma}_{\zeta_F} = \Omega(\xi_F) = \Gamma(0) - \underline{R}_{1,2:p+1}\,\underline{R}^{-1}_{1:p,1:p}\,\underline{R}_{2:p+1,1}. \tag{7}$$

A parallel result holds for the QMLEs, in the manner described at the beginning of this subsection. That is, the sample autocovariances are defined for $0 \le h \le n-1$ by

$$\widehat{\Gamma}(h) = n^{-1} \sum_{t=1}^{n-h} \left(\mathbf{W}_{t+h} - \overline{\mathbf{W}}\right) \left(\mathbf{W}_t - \overline{\mathbf{W}}\right)',$$

and $\widehat{\Gamma}(-h) = \widehat{\Gamma}'(h)$; it is easily seen that these quantities are related to the periodogram via

$$\underline{I}(\lambda) = \sum_{h=-n+1}^{n-1} \widehat{\Gamma}(h) e^{-ih\lambda}.$$

We assume that $p \le n-1$. Then the QMLEs satisfy the empirical YW equations, obtained by replacing $\Gamma(h)$ in (6) by $\widehat{\Gamma}(h)$, and so forth. Convergence of QMLEs to PTVs is guaranteed by results in Taniguchi and Kakizawa (2000).

2.2.2 Properties of Optimization for Constrained Models

Now let us consider the case where the VAR model has some constraints. We next provide an explicit solution for the PTV and QMLE when elements of ξ are constrained, which is a novel result.

Note that $\xi = \operatorname{vec} \underline{\phi}$ is the full vector of parameters. If some of these are constrained, we can write

$$\operatorname{vec} \underline{\phi} = J\,\psi + a \tag{8}$$

for a matrix J that is $m^2 p \times r$, where $r \le m^2 p$; here a is an r-vector. The vector ψ consists of all free parameters in $\underline{\phi}$. Unfortunately, there is no guarantee that the PTVs/QMLEs for such a constrained VAR will result in a stable model, and we've found through numerical experiments that this can indeed occur. The structure of J is arbitrary (only that its entries are known quantities, and not parameters), so the case that multiple entries of $\underline{\phi}$ are the same can also be entertained by (8).

We next state PTVs and QMLEs for $\underline{\phi}$ together with $\underline{\sigma}_\zeta$, with each formula being dependent on the other – similarly to the OLS solution discussed in Lütkepohl (2006). The PTV for $\underline{\phi}$ is still denoted by $\underline{\widetilde{\phi}}$, but it is computed in terms of the PTV $\widetilde{\psi}$, and $\xi_F = \operatorname{vec} \underline{\widetilde{\phi}} = J\,\widetilde{\psi} + a$. Likewise, $\underline{\widetilde{\sigma}} = \underline{\sigma}_{\zeta_F} = \Omega(\xi_F)$ by the previous subsection's general results. Now we can state our result.

Proposition 2 *Let* $\{\mathbf{W}_t\}$ *be stationary and invertible, with full rank spectral density matrix. Then the PTV* $(\widetilde{\psi}, \underline{\widetilde{\sigma}})$ *for a fitted constrained VAR(p) with constraints of the form (8) satisfies*

$$\widetilde{\psi} = \left(J' \left[\underline{R}_{1:p,1:p} \otimes \widetilde{\underline{\sigma}}^{-1}\right] J\right)^{-1} J' \left\{\left[\underline{R}'_{1,2:p+1} \otimes \widetilde{\underline{\sigma}}^{-1}\right] vec(\underline{1}_m) - \left[\underline{R}_{1:p,1:p} \otimes \widetilde{\underline{\sigma}}^{-1}\right] a\right\}$$
$$\widetilde{\underline{\sigma}} = \Omega\left(\xi_F\right).$$

Remark 1 The fitted constrained VAR models need not satisfy the Riccati equations, which take the form $\Gamma(0) = \underline{\phi}\,\underline{R}_{1:p,1:p}\,\underline{\phi}' + \underline{\sigma}$, and hence the resulting fitted VAR model need not correspond to a stationary process. This phenomenon arises due to taking unconstrained optimization of the TIV over all $\psi \in \mathbb{R}^r$, whereas only some subset of this space, in general, corresponds to stable VAR processes. It is interesting that enforcing certain kinds of constraints of the type given by (8) essentially forces the PTVs into a region of instability. The broader problem of enforcing stability is not studied in this paper.

Remark 2 In general we cannot substitute the formula for $\widetilde{\underline{\sigma}}$ into the formula for $\widetilde{\psi}$ and simplify, because the algebra is intractable. In the special case that J is the identity and $a = 0$ (the unconstrained case), the formula for $\widetilde{\psi}$ simplifies to

$$\left[\underline{R}^{-1}_{1:p,1:p} \otimes \widetilde{\underline{\sigma}}\right] \left[\underline{R}'_{1,2:p+1} \otimes \widetilde{\underline{\sigma}}^{-1}\right] \operatorname{vec}(\underline{1}_m) = \operatorname{vec}\left(\underline{R}_{1,2:p+1}\,\underline{R}^{-1}_{1:p,1:p}\right),$$

which is the YW equation. To solve the coupled system, one could propose initial guesses (such as the YW solutions) and iteratively solve the formulas on a computer, hoping for contraction towards the PTV solution pair.

Substituting empirical estimates for the autocovariances, the same mathematics produces formulas for the QMLEs. The empirical counterpart of the asymptotic story is exactly similar. We denote the parameter estimates by

$$\widehat{\psi}_{QMLE} = \left(J' \left[\widehat{\underline{R}}_{1:p,1:p} \otimes \widehat{\underline{\sigma}}^{-1}_{QMLE}\right] J\right)^{-1}$$
$$\cdot J' \left\{\left[\widehat{\underline{R}}'_{1,2:p+1} \otimes \widehat{\underline{\sigma}}^{-1}_{QMLE}\right] \operatorname{vec}(\underline{1}_m) - \left[\widehat{\underline{R}}_{1:p,1:p} \otimes \widehat{\underline{\sigma}}^{-1}_{QMLE}\right] a\right\}$$
$$\widehat{\underline{\sigma}}_{QMLE} = \Omega\left(\xi_I\right),$$

and $\xi_I = \operatorname{vec} \widehat{\underline{\phi}}_{QMLE} = J\,\widehat{\psi}_{QMLE} + a$. These estimates need not result in a stable fitted model (see Section 3).

Suppose that the true process is a VAR(p), and we fit a constrained VAR(p) model. Then the QMLEs and PTVs can be computed iteratively via the formulas of Proposition 2. In the special case that the true process is a constrained VAR(p) (i.e., the specified model is correct), then $\mathbf{W}_t = \sum_{j=1}^{p} \widetilde{\underline{\phi}}_j \mathbf{W}_{t-j} + \varepsilon_t$ and (6) is true. Also, plugging into (5) yields (7), so that Proposition 1 holds for this case. The formula (7) for the FEV is the same as would be obtained using the constrained VAR formula, because the unconstrained model reduces to the constrained model asymptotically. We can use the empirical version of (7) to estimate the FEV consistently, and substitute into the formula for $\widehat{\psi}_{QMLE}$; however, these estimates are only consistent for the true parameters under a correct model hypothesis, and need not

tend to the PTVs in the case that the model is wrong. Also see the discussion of the estimation of the FEV via LSE methodology in Lütkepohl (2006).

A formula for LSEs for the constrained VAR(p) is given in Lütkepohl (2006), which we translate into our own notation. Omitting mean effects, we let Z be a $pm \times (n-p)$ dimensional matrix, with columns given by $[Z_p, Z_{p+1}, \cdots, Z_{n-1}]$ and $Z_t = [\mathbf{W}_t', \mathbf{W}_{t-1}', \cdots, \mathbf{W}_{t-p+1}']'$. Note that when p is fairly large, some data is being "thrown away." Also let W be $m \times (n-p)$ dimensional, given by $W = [\mathbf{W}_{p+1}, \mathbf{W}_{p+2}, \cdots, \mathbf{W}_n]$.

The method requires some plug-in estimate of the innovation variance, which we generically denote by $\widehat{\underline{\sigma}}$; this might be estimated by a separate method, and then plugged in below, as described in Lütkepohl (2006). The LSE formula for ψ is then

$$\widehat{\psi}_{LSE} = \left(J' \left[ZZ' \otimes \widehat{\underline{\sigma}}^{-1} \right] J \right)^{-1} J' \left\{ \left[ZW' \otimes \widehat{\underline{\sigma}}^{-1} \right] \text{vec}(\underline{1}_m) - \left[ZZ' \otimes \widehat{\underline{\sigma}}^{-1} \right] a \right\}.$$

If we were to plug in the QMLE for the innovation covariance matrix, the similarities to the QMLE formula are striking. The above formula can be re-expressed in an equivalent form. Letting $\text{vec}\widehat{\underline{\phi}}_{LSE} = J\,\widehat{\psi}_{LSE} + a$, we find the equivalent expression

$$J' \,\text{vec}\left(\widehat{\underline{\sigma}}^{-1} \left[\widehat{\underline{\phi}}_{LSE} ZZ' - WZ' \right] \right) = 0.$$

Now $n^{-1} ZZ' \approx \widehat{\underline{R}}_{1:p,1:p}$ and $n^{-1} WZ' \approx \widehat{\underline{R}}'_{1,2:p+1}$; the relations would have been exact, except for some missing terms due to the data that gets thrown away by the LSE method. This approximation error is $O_P(1/n)$, and has no impact on the asymptotic behavior. On the other hand, we can re-express the QMLEs as

$$J' \,\text{vec}\left(\widehat{\underline{\sigma}}^{-1}_{QMLE} \left[\widehat{\underline{\phi}}_{QMLE} \widehat{\underline{R}}_{1:p,1:p} - \widehat{\underline{R}}_{1,2:p+1} \right] \right) = 0.$$

Notice that the expression in square brackets is identically zero if and only if the QMLE satisfies the Yule-Walker equations (and when J is the identity – i.e., no constraints in play – the above equation reduces to (6)).

So, if we use the QMLE for the innovation variance in the LSE approach – or another estimate that is consistent for the PTV – then the LSEs are approximate solutions to the above QMLE equation. This tells us that their asymptotic behavior is the same, so that LSEs obey the same Central Limit Theorem as the QMLEs, indicated in Taniguchi and Kakizawa (2000), *even when* the VAR model is misspecified.

3 Numerical Illustrations

3.1 Finite-Sample Results

For constrained bivariate VAR(1) models, the chief fitting methods are MLE, QMLE, or LSE. Explicit formulas are given in Section **??**, which we here implement on four bivariate VAR(1) processes described below. Let Φ denote the first coefficient matrix $\underline{\phi}_1$, with jkth entry denoted Φ_{jk}. The Φ matrices for the four examples are

$$\begin{bmatrix} 1/2 & 1/3 \\ 1/3 & 1/2 \end{bmatrix} \quad \begin{bmatrix} 2/3 & 0 \\ 1 & 1/3 \end{bmatrix} \quad \begin{bmatrix} .95 & 0 \\ 1 & 1/2 \end{bmatrix} \quad \begin{bmatrix} -.25 & .5 \\ -1 & 1.25 \end{bmatrix},$$

and in each case the innovation variance matrix is the identity. All four processes are stable.

We investigate fitting three models – denoted A, B, and C – to each process via QMLE and LSE. Model A is the unconstrained VAR(1), while model B has the constraint that $\Phi_{12} = 0$, and model C has the constraint that $\Phi_{11} = 0$. So model B is a misspecification for the first and fourth processes, while model C is a misspecification for all four processes. For model A the PTVs correspond to the true values, but for models B and C they can be quite different due to mis-specification. The PTVs for Φ, for the four processes respectively, are

$$\begin{bmatrix} .6739 & 0 \\ 1/3 & 1/2 \end{bmatrix} \quad \begin{bmatrix} 2/3 & 0 \\ 1 & 1/3 \end{bmatrix} \quad \begin{bmatrix} .95 & 0 \\ 1 & 1/2 \end{bmatrix} \quad \begin{bmatrix} .4244 & 0 \\ -1 & 1.25 \end{bmatrix},$$

for model B, and for model C are given by

$$\begin{bmatrix} 0 & .5942 \\ 1/3 & 1/2 \end{bmatrix} \quad \begin{bmatrix} 0 & .5373 \\ 0 & .6915 \end{bmatrix} \quad \begin{bmatrix} 0 & .4914 \\ 0 & .9668 \end{bmatrix} \quad \begin{bmatrix} 0 & .1954 \\ .2443 & .7721 \end{bmatrix}.$$

The PTVs for $\underline{\sigma}$ are in all cases equal to $\underline{1}_2$; see additional discussion in McElroy and Findley (2013). These quantities are computed from the formulas of Proposition 2. We see that all the Φ PTVs are stable for the first three processes, but is unstable for model B fitted to the fourth process. However, for model C all PTVs are stable for all four processes; the double zero for the second and third process with model C is quite interesting.

It is interesting to examine the PTVs in the cases of model B and model C, fitted to the first process. Although these models are mis-specified, their mis-specification in some sense chiefly pertains to the forecast performance of the first component of the bi-variate series; actually, their PTVs for the second component of the bi-variate series are correct! That is, utilizing the mis-specified models B and C has no impact on the asymptotic forecast performance of the second component series.

The example given by the fourth process begs the question: how often do unstable PTV fits arise in practice? We drew a sample of a million bivariate VAR(1) processes

Parameter Estimates						
Parameters	Models					
n=100	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.481 .328	.649 0	0 .574	.487 .331	.654 0	0 .579
	.329 .478	.335 .471	.317 .488	.332 .483	.332 .483	.332 .483
Max $\lvert\zeta\rvert$	.808	.652	.734	.817	.658	.742
Min $\lvert\zeta\rvert$	.157	.467	.245	.159	.481	.258
Prob Unstable	0	0	0	0	0	0
n=200	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.490 .331	.661 0	0 .585	.493 .333	.665 0	0 .587
	.332 .489	.336 .485	.326 .494	.334 .492	.334 .492	.334 .492
Max $\lvert\zeta\rvert$	.821	.662	.748	.826	.665	.752
Min $\lvert\zeta\rvert$	.158	.485	.253	.159	.491	.260
Prob Unstable	0	0	0	0	0	0
n= 400	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.495 .332	.667 0	0 .588	.496 .333	.668 0	0 .589
	.332 .494	.334 .492	.328 .497	.333 .496	.333 .496	.333 .496
Max $\lvert\zeta\rvert$	.826	.667	.753	.828	.668	.755
Min $\lvert\zeta\rvert$	.163	.492	.256	.163	.496	.259
Prob Unstable	0	0	0	0	0	0

Table 1 Model fitting results for sample sizes 100, 200, 400 from the VAR(1) with $\Phi_{11} = 1/2$, $\Phi_{12} = 1/3$, $\Phi_{21} = 1/3$, $\Phi_{22} = 1/2$, and $\underline{\sigma} = \underline{1}_2$. Models A, B, C are used, corresponding to unconstrained VAR(1), a VAR(1) with $\Phi_{12} = 0$, and a VAR(1) with $\Phi_{11} = 0$ respectively. Mean values for parameter estimates are reported for Φ, as well as the maximal and minimal absolute eigenvalues. Unless both of these are less than one, the fit is unstable, and the proportion of unstable fits is reported.

Parameter Estimates						
Parameters	Models					
n=100	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.647 -.003	.647 0	0 .530	.654 -.003	.654 0	0 .230
	.997 .324	.997 .324	-.031 .685	1.007 .328	1.007 .328	1.007 .328
Max $\lvert\zeta\rvert$	.642	.648	.646	.649	.654	.665
Min $\lvert\zeta\rvert$	.350	.324	.111	.354	.328	.338
Prob Unstable	0	0	0	0	0	0
n=200	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.658 -.002	.657 0	0 .534	.661 -.002	.661 0	0 .234
	.998 .329	.998 .329	-.015 .688	1.002 .331	1.002 .331	1.002 .331
Max $\lvert\zeta\rvert$	.643	.657	.668	.646	.661	.674
Min $\lvert\zeta\rvert$	.353	.329	.071	.355	.331	.343
Prob Unstable	0	0	0	0	0	0
n= 400	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.662 -.001	.661 0	0 .536	.664 -.001	.663 0	0 .236
	.999 .331	.999 .331	-.009 .689	1.001 .332	1.001 .332	1.001 .332
Max $\lvert\zeta\rvert$	.646	.661	.679	.647	.663	.678
Min $\lvert\zeta\rvert$	.351	.331	.049	.352	.332	.346
Prob Unstable	0	0	0	0	0	0

Table 2 Model fitting results for sample sizes 100, 200, 400 from the VAR(1) with $\Phi_{11} = 2/3$, $\Phi_{12} = 0$, $\Phi_{21} = 1$, $\Phi_{22} = 1/3$, and $\underline{\sigma} = \underline{1}_2$. Models A, B, C are used, corresponding to unconstrained VAR(1), a VAR(1) with $\Phi_{12} = 0$, and a VAR(1) with $\Phi_{11} = 0$ respectively. Mean values for parameter estimates are reported for Φ, as well as the maximal and minimal absolute eigenvalues. Unless both of these are less than one, the fit is unstable, and the proportion of unstable fits is reported.

Parameter Estimates						
Parameters	Models					
n=100	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.925 -.003	.922 0	0 .483	.934 -.002	.932 0	0 .425
	1.000 .488	1.000 .488	-.260 1.063	1.009 .495	1.009 .495	1.009 .495
Max $\lvert\zeta\rvert$	.913	.922	.926	.923	.932	.946
Min $\lvert\zeta\rvert$	.501	.488	.149	.506	.495	.451
Prob Unstable	0	0	0	.0030	.0014	.0044
n=200	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.938 -.002	.936 0	0 .487	.942 -.001	.941 0	0 .434
	1.000 .494	.999 .495	-.159 1.029	1.004 .498	1.004 .498	1.004 .498
Max $\lvert\zeta\rvert$	.932	.936	.946	.937	.941	.954
Min $\lvert\zeta\rvert$	.501	.495	.095	.503	.498	.456
Prob Unstable	0	0	0	0	0	0
n= 400	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	.945 -.001	.943 0	0 .490	.947 -.001	.946 0	0 .439
	1.000 .497	1.000 .497	-.095 1.006	1.002 .499	1.002 .499	1.002 .499
Max $\lvert\zeta\rvert$	.942	.943	.957	.944	.946	.958
Min $\lvert\zeta\rvert$	.500	.497	.059	.501	.499	.459
Prob Unstable	0	0	0	0	0	0

Table 3 Model fitting results for sample sizes 100, 200, 400 from the VAR(1) with $\Phi_{11} = .95$, $\Phi_{12} = 0$, $\Phi_{21} = 1$, $\Phi_{22} = 1/2$, and $\underline{\sigma} = \underline{1}_2$. Models A, B, C are used, corresponding to unconstrained VAR(1), a VAR(1) with $\Phi_{12} = 0$, and a VAR(1) with $\Phi_{11} = 0$ respectively. Mean values for parameter estimates are reported for Φ, as well as the maximal and minimal absolute eigenvalues. Unless both of these are less than one, the fit is unstable, and the proportion of unstable fits is reported.

Parameter Estimates						
Parameters	Models					
n=100	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	-.257 .504	.409 0	0 .183	-.260 .509	.413 0	0 .408
	-.989 1.231	-.947 1.199	.243 .753	-.999 1.243	-.999 1.243	-.999 1.243
Max $\lvert\zeta\rvert$	.707	1.199	.808	.714	1.243	.687
Min $\lvert\zeta\rvert$	.278	.409	.055	.281	.413	.598
Prob Unstable	0	98.76 %	0	0	100.00 %	0
n=200	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	-.251 .501	.419 0	0 .189	-.253 .504	.421 0	0 .406
	-.994 1.240	-.973 1.224	.244 .763	-.999 1.247	-.999 1.247	-.999 1.247
Max $\lvert\zeta\rvert$	.722	1.224	.819	.726	1.247	.671
Min $\lvert\zeta\rvert$	.269	.418	.056	.271	.421	.608
Prob Unstable	0	100.00 %	0	0	100.00 %	0
n= 400	QMLE Model A	QMLE Model B	QMLE Model C	LSE Model A	LSE Model B	LSE Model C
Φ	-.251 .500	.422 0	0 .193	-.251 .502	.423 0	0 .405
	-.997 1.246	-.985 1.237	.243 .768	-1.000 1.249	-1.000 1.249	-1.000 1.249
Max $\lvert\zeta\rvert$	.737	1.237	.825	.739	1.249	.659
Min $\lvert\zeta\rvert$	.258	.422	.057	.259	.423	.616
Prob Unstable	0	100.00 %	0	0	100.00 %	0

Table 4 Model fitting results for sample sizes 100, 200, 400 from the VAR(1) with $\Phi_{11} = -1/4$, $\Phi_{12} = 1/2$, $\Phi_{21} = -1$, $\Phi_{22} = 5/4$, and $\underline{\sigma} = \underline{1}_2$. Models A, B, C are used, corresponding to unconstrained VAR(1), a VAR(1) with $\Phi_{12} = 0$, and a VAR(1) with $\Phi_{11} = 0$ respectively. Mean values for parameter estimates are reported for Φ, as well as the maximal and minimal absolute eigenvalues. Unless both of these are less than one, the fit is unstable, and the proportion of unstable fits is reported.

by allowing each entry of Φ to be an independent normal variable, and found that 34% of these processes were stable; of those, the proportion having stable PTVs arising from fitting model B was only 26%. This indicates that a high proportion of stable VAR processes may have unstable PTVs when constrained models are utilized.

We next proceeded to simulate from these four processes, fitting all three models via both QMLE and LSE methodologies. The results are summarize in Tables 1, 2, 3 and 4. There we present the mean values of the estimates of Φ, computed over 5000 simulations of the given VAR processes, with sample sizes of 100, 200, and 400. We also present mean values of the maximum and minimum absolute eigenvalues of Φ. Only rarely did unstable estimates arise in practice for the first three processes: this was assessed by computing the proportion of simulations wherein the maximum eigenvalue exceeded one. This only occurred for the LSE estimates in the case of sample size 100; the QMLE method always resulted in stable fits, and the LSE estimates become "increasingly stable" as sample size was increased. For the fourth process, models A and C produced stable fits in finite sample, but virtually all the time model B produced an unstable VAR, as expected.

3.2 Gauging Forecast MSE

We now describe an application of the calculation of PTVs. Suppose that we wished to study the impact of model misspecification on forecast performance, as a function of an underlying process; see Schorfheide (2005) for motivation and discussion. So we suppose that the true $\underline{F}$ is known for the process we are studying, and some misspecified model is fit to the data. McElroy and McCracken (2012) provides expressions for the multi-step forecast error from a misspecified model; the forecast error process is

$$-[\underline{\Delta}^{-1}(B)\underline{\Psi}(B)]_0^{h-1}\underline{\Psi}^{-1}(B)\,\mathbf{W}_t$$

if we are forecasting h steps ahead. Now the parameter estimates would enter into the coefficients of $\underline{\Psi}$. Asymptotically, these estimates will converge to the PTVs. The variance of the corresponding error process (where parameter estimates have converged to the PTVs) is given by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}[\underline{\Delta}^{-1}(z)\underline{\Psi}(z)]_0^{h-1}\,\underline{\Psi}^{-1}(z)\,\underline{F}(\lambda)\,\underline{\Psi}^{\dagger}(\bar{z})\,[\underline{\Psi}'(\bar{z})\underline{\Delta}^{\dagger}(\bar{z})]_0^{h-1}\,d\lambda.$$

This matrix depends on the data process in a double fashion: first through $\underline{F}$ in the center of the integrand, and again through the PTVs involved in $\underline{\Psi}$, which are previously computed as described in Section 3. As an example, consider the bivariate VAR(1) models A, B, C of the previous sub-section, fitted to any of the first three true processes described above (we ignore the fourth process, because the forecasting formulas do not apply to unstable model fits). The h-step ahead forecast error variance matrix simplifies to

$$\Gamma(0) - \underline{\phi}_1^h \Gamma(-h) - \Gamma(h)\,\underline{\phi}_1'^h + \underline{\phi}_1^h \Gamma(0)\,\underline{\phi}_1'^h.$$

Observe that this is a symmetric matrix, and its minimal value at $h = 1$ is given by the innovation variance matrix $\underline{\sigma}$. Into this formula, we would substitute the appropriate PTVs for $\underline{\phi}_1$ and the true process' autocovariances for $\Gamma(h)$ and $\Gamma(0)$. The resulting entries of the forecast error variance matrix are plotted in Figure 1 with $1 \leq h \leq 100$, with matrix entries for the first diagonal in Red (Solid), the second diagonal in Green (Dotted-Dashed), and the off-diagonal in Blue (Dashed). Some of these plots are identical, which occurs when model B is actually correctly specified.

For the first process, going across the top row of Figure 1, we note that model A is correctly specified, and both diagonal entries of the forecast variance matrix are the same due to symmetry of Φ. Mis-specification, as shown for models B and C of the top row, has no impact on the second diagonal (Dotted-Dashed), but increases the first diagonal (Solid) of the MSE matrix for short horizons. The reason for this behavior is that the PTVs for models B and C are still correct for the second component of the bi-variate series, as mentioned above.

For the second process, both models A and B are correctly specified, and hence the MSE plots are identical. Now there is a large discrepancy in forecast performance between the first component series (Solid) and the second (Dotted-Dashed). The final panel for model C shows an interesting feature: forecast performance at low horizons is actually worse than at longer horizons, which can happen for a mis-specified model. The third process has a similar story, although model C fares competitively in the long run with the correctly specified models.

Disclaimer

This article is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

References

1. Akaike, H.: Fitting autoregresive models for prediction. Annals of the Institute of Statistical Mathematics **21**, 243–247 (1969)
2. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**, 716–723 (1974)
3. Akaike, H., Kitagawa, G.: The Practice of Time Series Analysis. Springer, New York (1999)
4. Brockwell, P., Davis, R. Time Series: Theory and Methods, 2nd Ed. Springer, New York (1991)
5. Golub, G., Van Loan, C.: Matrix Computations. Johns Hopkins University Press, Baltimore (1996)
6. Jacobs, J., Van Norden, S.: Modeling data revisions: measurement error and dynamics of "true" values. Journal of Econometrics **161**, 101–109 (2011)
7. Magnus, J., Neudecker, H.: Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York (1999)

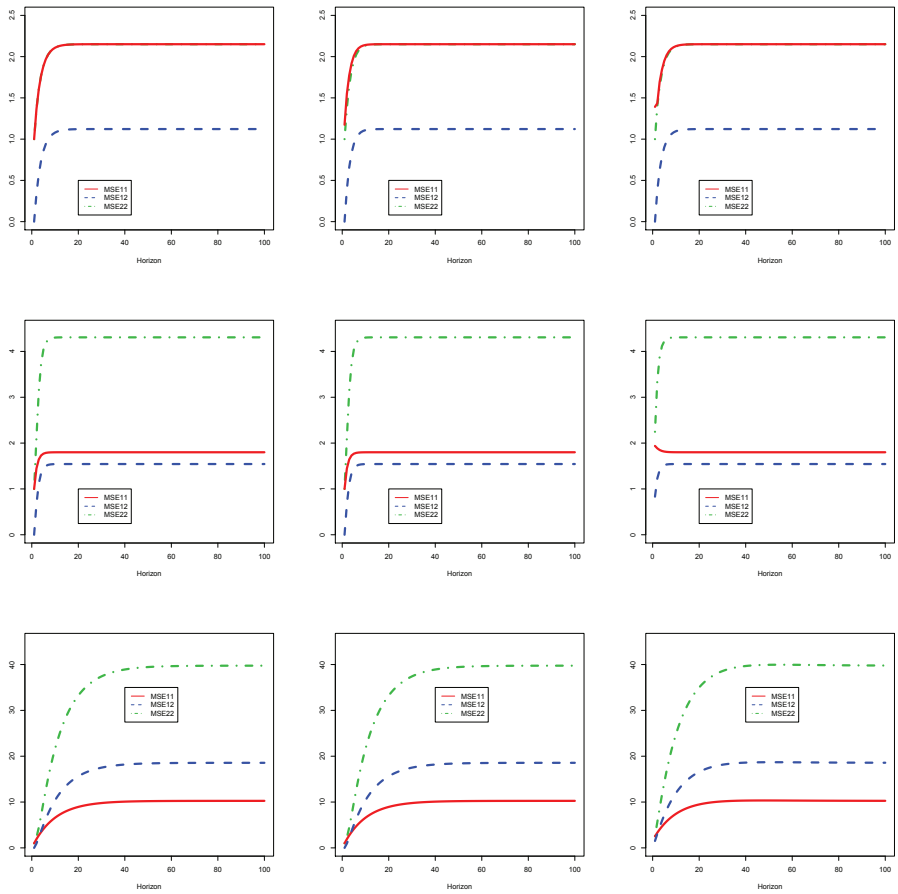

Fig. 1 Asymptotic forecast MSE as a function of forecast horizon. In each panel, the entries of the FEV matrix are plotted, with the first diagonal entry in Red (Solid), the second diagonal entry in Green (Dotted-Dashed), and the off-diagonal in Blue (Dashed). The first row of panels corresponds to Process 1 of Section 3, while the second row of panels corresponds to Process 2 and the third row to Process 3. The first column of panels corresponds to Model A of Section 2, while the second column of panels corresponds to Model B and the third column to Model C.

8. Maïnassara, B., Francq, C.: Estimating structural VARMA models with uncorrelated but non-independent error terms. Journal of Multivariate Analysis **102**, 496–505 (2011)
9. McElroy, T., Findley, F.: Fitting Constrained Vector Autoregression Models. U.S. Census Bureau Research Report, (2013).
10. McElroy, T., McCracken, M.: Multi-Step Ahead Forecasting of Vector Time Series. Federal Reserve Bank of St. Louis, Working Papers, 2012-060A (2012)
11. Pinheiro, J., Bates, D.: Unconstrained parameterizations for variance-covariance matrices. Statistics and Computing **6**, 289–296 (1996)

12. Reinsel, G.: Elements of Multivariate Time Series Analysis, 2nd Ed. Springer-Verlag, New York (1996)
13. Schorfheide, F.: VAR forecasting under misspecification. Journal of Econometrics **128**, 99–136 (2005)
14. Taniguchi, M., Kakizawa, Y.: Asymptotic Theory of Statistical Inference for Time Series. Springer-Verlag, New York (2000)
15. Whittle, P.: The analysis of multiple stationary time series. Journal of the Royal Statistical Society, Series B **15**, 125–139 (1953)
16. Whittle, P.: Prediction and Regulation. English Universities Press (1963)
17. Wilks, S.: Certain generalizations in the analysis of variance. Biometrika **24**, 471–494 (1932)