RESEARCH REPORT SERIES
(Statistics #2013-06)

Fitting Constrained Vector Autoregression Models

Tucker McElroy
David Findley

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: September 25, 2013

Fitting Constrained Vector Autoregression Models

Tucker McElroy[1] and David Findley[1]
U.S. Census Bureau

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; however, the QMLEs need not minimize the component entries of the mean square forecast error matrix, in contrast 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.

[1] Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100

Keywords. ARIMA, Forecasting, Frequency Domain, Goodness-of-Fit, Model Comparison, 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) – i.e., the minimizers of the Whittle likelihood – 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 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 (see Lütkepohl (2006) for a precise statement of these classical results). 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[1], 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

[1]Formally, the concept of stability states that a stationary solution to the autoregressive equation exists, and a sufficient condition is the causality criterion discussed in Theorem 11.3.1 of Brockwell and Davis (1991); in this paper we follow Lütkepohl (2006) in defining stability to mean the causality criterion, namely that the roots of the determinant of the autoregressive polynomial lie outside the unit circle of the complex plane.

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.

In order to motivate the difficulties inherent in enforcing stability and parameter constraints, we consider in Section 2 a bivariate VAR(1) initially, and describe how to fit via MLE to ensure a stable fit, and how to obtain the various constrained VAR(1) models via MLE. We then present the QMLE solution based on explicit algebra, which demonstrates some interesting discrepancies between the unconstrained and constrained cases. Then Section 3 provides the general theory of the QMLE for constrained VAR models, with connections to the Yule-Walker equations and the implications to forecasting. 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 4 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. Then Section 5 provides an application of the LSE and QMLE methods on a revision vintage data set, i.e., data that is revised over time, resulting in a sequence of vintages.

2 A Motivational Example

We begin by considering a bivariate VAR(1) model, which we wish to estimate such that the resulting fit is guaranteed to be stable (see Brockwell and Davis (1991) for a treatment of stability and invertibility for VARMA processes). We first provide a parametrization that facilitates the MLE approach, and secondly discuss the QMLE approach.

2.1 Maximum Likelihood Estimation

Maximum Likelihood Estimates (MLEs) have the attractive property of being asymptotically normal and efficient under some constraints on the model parameter space and the data process, as discussed in Taniguchi and Kakizawa (2000). Consider a bivariate VAR(1) process $\{X_t\}$ with $\Phi(z) = 1_2 - \Phi z$, where 1_2 is an identity matrix of dimension two, such that $\Phi(B)X_t = \epsilon_t$ and $\{\epsilon_t\}$ is white noise with covariance matrix Σ. We shall be interested in stable processes, wherein the zeroes of $\det(1_2 - \Phi z)$ lie outside the unit circle of the complex plane. Equivalently, all the eigenvalues of Φ have magnitude strictly less than unity.

Adopting the convention that $\Gamma(h) = \mathbb{E}[X_{t+h}X_t']$ (as in Brockwell and Davis (1991)) with $'$ denoting transpose, we obtain

$$\Gamma(h) = \Phi\,\Gamma(h-1) \quad h \geq 1, \qquad \Gamma(0) = \Phi\,\Gamma(0)\,\Phi' + \Sigma.$$

The parameters are the four entries of Φ, along with the three free entries of Σ; these are related to $\Gamma(0)$ via equation (2.1.32) of Lütkepohl (2006):

$$\text{vec}\Gamma(0) = (1_4 - \Phi \otimes \Phi)^{-1} \text{vec}\Sigma. \tag{1}$$

So from a knowledge of Φ and Σ, we can compute $\Gamma(h)$ quickly. Now consider the data written in vector form as $X = [X_1', \cdots, X_n']'$, which is assumed to have mean zero for simplicity. Here n is the sample size. The covariance matrix of this random vector is denoted Σ_X, and has a block form, with n^2 blocks filled by 2×2 matrices. The jkth block of Σ_X is given by $\Gamma(j-k)$, noting that $\Gamma(-h) = \Gamma'(h)$. Then the scaled log Gaussian likelihood is (up to an irrelevant constant)

$$X'\Sigma_X^{-1}X + \log|\Sigma_X|. \tag{2}$$

This objective function is simple to evaluate, given any viable candidate Φ. However, we can't just use arbitrary real numbers for the entries of Φ, since stability is a necessary condition. The set of coefficients Φ that guarantees a stable VAR(1) is a complicated four-dimensional manifold, which we now proceed to discuss.

The four entries of Φ are designated Φ_{11}, Φ_{12}, Φ_{21}, and Φ_{22}, and allowing these to be arbitrary real numbers describes all of $\mathbb{R}^4$; the actual subset of parameters corresponding to all stable VAR(1) processes will be called the *stability manifold*. We demonstrate that it is locally homeomorphic to $\mathbb{R}^4$, which is helpful for achieving unconstrained optimization of the likelihood; see Pinheiro and Bates (1996) for a discussion of the advantages of unconstrained optimization over a parameter set that is in bijective correspondence with Euclidean space. Two necessary conditions for stability are that $|\text{tr}\Phi| < 2$ and that $|\det\Phi| < 1$; however, these are not sufficient.

Proposition 1 *The stability manifold for the bivariate VAR(1) is given by all matrices Φ such that*

$$|tr\Phi| < 1 + \det\Phi \qquad \textit{and} \qquad |\det\Phi| < 1.$$

The compact expression in Proposition 1 can be rewritten as the following inequalities:

$$\begin{aligned} \Phi_{12}\,\Phi_{21} &< (1-\Phi_{11})\,(1-\Phi_{22}) \\ \Phi_{12}\,\Phi_{21} &< (1+\Phi_{11})\,(1+\Phi_{22}) \\ -1 + \Phi_{11}\,\Phi_{22} < \Phi_{12}\,\Phi_{21} &< 1 + \Phi_{11}\,\Phi_{22}. \end{aligned}$$

The property that parameter spaces are manifolds has practical ramifications for statistical inference, which we briefly remind the reader about. In frequentist estimation of the model, non-linear optimization routines are required to find the maximizers of the Gaussian likelihood (or the minimizers of the Whittle likelihood, if this is preferred). As Pinheiro and Bates (1996) argue persuasively, it is preferable to use unconstrained optimization routines, which in turn require that

the parameter set can be written as the image – by some homeomorphism (continuous bijection with continuous inverse) – of Euclidean space. Optimization then occurs within this pre-image Euclidean space in an unconstrained fashion, and the full scope of the parameter manifold is searched out by the algorithm – this is guaranteed by the surjectivity of the homeomorphism. For example, when fitting a simple iid Gaussian model with variance σ^2, it is convenient to express the parameter manifold $(0, \infty)$ as the image of $\mathbb{R}$ via some convenient homeomorphism φ (e.g., the exponential function), and perform unconstrained optimization of a "pre-parameter" ϑ which is related to the desired parameter via $\sigma^2 = \varphi(\vartheta)$. This is a commonly used device.

On the other hand, with Bayesian estimation it is advantageous to elicit priors that have compact support on the closure of the parameter manifold. If this is not the case, then posterior estimates of the parameters may correspond to non-stable processes; in order to correct the deficiency, some authors have advocated projection or trimming of undesirable draws. This procedure is computationally inefficient at best. It is more elegant and advantageous to elicit the priors to fit exactly to the parameter space the modeler wishes to explore. Identification of the homeomorphism φ that maps Euclidean space to the stability manifold gives a fairly natural method of eliciting priors: a multivariate standard uncorrelated normal distribution could be placed on the Euclidean pre-parameter space, and the induced distribution obtained by application of φ is the definition of the prior on the parameter manifold. In the above example of the iid Gaussian model, where the homeomorphism is taken to be the exponential function, the prior on σ^2 is just the log-normal distribution.

Returning to our VAR(1) example, some of the parameter manifolds corresponding to constrained models are of interest. Sometimes a data set warrants setting certain coefficients of Φ equal to zero. For example, if $\Phi_{21} = 0$ then the restrictions reduce to $|\Phi_{11}| < 1$, $|\Phi_{22}| < 1$, and Φ_{12} unconstrained. The parameter manifold is a subset of $\mathbb{R}^3$, which is a cylinder set with base given by a square. A similar manifold is obtained for the restriction $\Phi_{12} = 0$. If one of the diagonal entries is zero, say $\Phi_{11} = 0$, then the relevant inequality is

$$-1 < \Phi_{12}\,\Phi_{21} < 1 - |\Phi_{22}|.$$

Thus, for any value of $|\Phi_{22}| \geq 0$, the parameters Φ_{12} and Φ_{21} are restricted to the interior of a hyperbola given by the above inequality.

In summary, this subsection describes how maximum likelihood estimation or Bayesian estimation of a VAR(1) could be carried out so as to guarantee all parameter estimates belong to the stability manifold. Important constrained sub-models can be easily set up and estimated. However, it seems difficult to generalize this particular example to higher order VAR models or higher variable dimension.

2.2 Quasi-Maximum Likelihood Estimation

We consider the same bivariate process $\{X_t\}$, and we refer to the first component $\{X_t^1\}$ as the first series, and the second component $\{X_t^2\}$ will be denoted as the second series. It is shown later in this paper that the QMLEs minimize the empirical TIV, which for an unconstrained VAR(1) model (called model A) is given by the determinant of the matrix Ω, defined to be $\mathbb{E}[(X_t - \Phi X_{t-1})(X_t - \Phi X_{t-1})']$, or the covariance of the one step ahead forecast errors. This quantity can be re-expressed as

$$\Omega = \Gamma(0) - \Phi\,\Gamma'(1) - \Gamma(1)\,\Phi' + \Phi\,\Gamma(0)\,\Phi';$$

this is called the Forecast Error Variance (FEV) matrix (see below for general explication). Here $\Gamma(0)$ and $\Gamma(1)$ are understood to be empirical estimates of covariance matrices, i.e., $\Gamma(h) = n^{-1}\sum_{t=1}^{n-h} X_{t+h}X_t'$ for $h \geq 0$. Actually, the same arguments apply verbatim if we seek to determine the PTVs, when we let the $\Gamma(h)$ denote the Data Generating Process' (DGP) autocovariance instead. But to focus our discussion, we let the $\Gamma(h)$ be the empirical estimates for now.

The unconstrained VAR(1) is fitted by minimizing $|\Omega|$ (the rationale is explained in Section 3 below), and the solution is given by the multivariate Yule-Walker (YW) equations: $\Phi = \Gamma(1)\,\Gamma^{-1}(0)$, as discussed below in generality. But if we wish to fit a constrained VAR(1), the YW solution is no longer relevant. For example, consider the constraint that $\Phi_{12} = 0$. The YW estimate for this entry is given by

$$\frac{\Gamma_{22}(0)\,\Gamma_{12}(1) - \Gamma_{12}(0)\,\Gamma_{22}(1)}{\Gamma_{22}(0)\,\Gamma_{11}(0) - \Gamma_{12}(0)\,\Gamma_{21}(0)}.$$

For any j, k, h, write $\rho_{jk}(h) = \Gamma_{jk}(h)/\Gamma_{jk}(0)$. Then we see that the YW estimate satisfies the desired constraint iff $\rho_{12}(1) = \rho_{22}(1)$, which is a probability zero event.

We now proceed to derive the solution for this constrained model, which we henceforth call model B. The component functions of model B's FEV matrix are

$$\begin{aligned}
\Omega_{11} &= \Gamma_{11}(0) - 2\Gamma_{11}(1)\Phi_{11} + \Gamma_{11}(0)\Phi_{11}^2 \\
\Omega_{12} &= \Gamma_{12}(0) - \Gamma_{21}(1)\Phi_{11} - \Gamma_{11}(1)\Phi_{21} - \Gamma_{12}(1)\Phi_{22} \\
&\quad + \Gamma_{11}(0)\Phi_{11}\Phi_{21} + \Gamma_{12}(0)\Phi_{11}\Phi_{22} \\
\Omega_{22} &= \Gamma_{22}(0) - 2\Gamma_{21}(1)\Phi_{21} - 2\Gamma_{22}(1)\Phi_{22} \\
&\quad + \Gamma_{11}(0)\Phi_{21}^2 + 2\Gamma_{12}(0)\Phi_{21}\Phi_{22} + \Gamma_{22}(0)\Phi_{22}^2.
\end{aligned}$$

It is noteworthy that Ω_{11} only depends on Φ_{11}, and not on Φ_{21} or Φ_{22}, while Ω_{22} does not depend on Φ_{11} at all. Of course $\Omega_{21} = \Omega_{12}$, because Ω is symmetric by construction. Also $|\Omega| = \Omega_{11}\Omega_{22} - \Omega_{12}^2$. Noting that there are only three free parameters – namely Φ_{11}, Φ_{21}, and Φ_{22} – the gradient of $|\Omega|$

becomes

$$\frac{\partial|\Omega|}{\partial\Phi_{11}} = 2\left(\Gamma_{11}(0)\Phi_{11} - \Gamma_{11}(1)\right)\Omega_{22} + 2\left(\Gamma_{21}(1) - \Gamma_{11}(0)\Phi_{21} - \Gamma_{12}(0)\Phi_{22}\right)\Omega_{12} \tag{3}$$

$$\frac{\partial|\Omega|}{\partial\Phi_{21}} = -2\left(\Gamma_{21}(1) - \Gamma_{11}(0)\Phi_{21} - \Gamma_{12}(0)\Phi_{22}\right)\Omega_{11} + 2\left(\Gamma_{11}(1) - \Gamma_{11}(0)\Phi_{11}\right)\Omega_{12} \tag{4}$$

$$\frac{\partial|\Omega|}{\partial\Phi_{22}} = -2\left(\Gamma_{22}(1) - \Gamma_{12}(0)\Phi_{21} - \Gamma_{22}(0)\Phi_{22}\right)\Omega_{11} + 2\left(\Gamma_{12}(1) - \Gamma_{12}(0)\Phi_{11}\right)\Omega_{12}. \tag{5}$$

We set (3), (4), and (5) equal to zero and solve; with B_j for $1 \leq j \leq 8$ defined in the Appendix, the eventual solution is

$$\begin{aligned}
\Phi_{11} &= \Gamma_{11}(1)/\Gamma_{11}(0) \\
\Phi_{21} &= \frac{B_1B_2 - B_4B_6}{B_1B_3 - B_5B_6} \\
\Phi_{22} &= \frac{B_1B_7 + B_6B_8}{B_1B_3 - B_5B_6}.
\end{aligned}$$

Plugging back into the component functions of the FEV provides values of Ω_{jk} at the optima. We obtain

$$\begin{aligned}
\Omega_{11} &= B_1 \\
\Omega_{12} &= B_1\,\frac{B_4B_3 - B_5B_2}{B_1B_3 - B_5B_6} \\
\Omega_{22} &= \Gamma_{22}(0) - 2\Phi_{22}\Gamma_{22}(1) - \Phi_{21}^2\Gamma_{11}(0) + \Phi_{22}^2\Gamma_{22}(0).
\end{aligned}$$

From here, the TIV $|\Omega|$ can be easily computed. We also see that $\Phi_{11} = \rho_{11}(1)$ optimizes Ω_{11}, but Ω_{22} and Ω_{12} do not depend upon Φ_{11}. So the optima only minimize the TIV $|\Omega|$, and do not optimize all of the component functions. This differs from the unconstrained case of model A, where it can be shown (see treatment below) that the YW solutions optimize not only the TIV, but each component function of the FEV, namely Ω_{11}, Ω_{12}, and Ω_{22}.

We can now show an interesting connection between the QMLEs for model A and B. The Ω_{11} entries of each model's FEV matrix are equal to one another iff model A is actually the same as model B, i.e., the constraint $\Phi_{12} = 0$ happens to hold for the YW estimate. The derivation is as follows: Ω_{11} (evaluated at the optima) for model B is equal to B_1, and for model A is equal to

$$\Gamma_{11}(0) - \left[\Gamma_{11}(1)\;\Gamma_{12}(1)\right]\Gamma^{-1}(0)\begin{bmatrix}\Gamma_{11}(1)\\ \Gamma_{12}(1)\end{bmatrix} = B_1 - \Gamma_{11}(0)B_6^2/|\Gamma(0)|.$$

Hence the equality of FEVs for the first series holds iff $B_6 = 0$ iff $\rho_{12}(1) = \rho_{22}(1)$ iff Φ_{12} for the YW estimate equals zero. This result is actually intuitive: model B is saying that the second series can not contribute meaningfully to the forecast performance of series one – otherwise Φ_{12} would have been allowed to be nonzero. Model A generally will perform better – its FEV is lower by the

amount $\Gamma_{11}(0)B_6^2/|\Gamma(0)|$ over model B – but fails to be an improvement in exactly the case that Φ_{12} is estimated to be zero!

The other constrained model of interest is obtained by fixing $\Phi_{11} = 0$, which says that the first series should have no impact on the forecast performance of the first series. We call this model C. With derivations similar to those of model B, the optima are

$$\begin{aligned}\Phi_{12} &= \Gamma_{21}(1)/\Gamma_{22}(0) \\ \Phi_{21} &= \frac{C_1C_7 + C_2C_8}{C_1C_4 - C_2C_6} \\ \Phi_{22} &= \frac{C_1C_3 - C_2C_5}{C_1C_4 - C_2C_6},\end{aligned}$$

where the C_j are defined in the Appendix. Plugging these back into the component functions of the FEV yields

$$\begin{aligned}\Omega_{11} &= C_1 \\ \Omega_{12} &= C_1\,\frac{C_4C_5 - C_3C_6}{C_1C_4 - C_2C_6} \\ \Omega_{22} &= \Gamma_{22}(0) - 2\Phi_{21}\Gamma_{12}(1) - \Phi_{22}^2\Gamma_{22}(0) + \Phi_{21}^2\Gamma_{11}(0).\end{aligned}$$

What if we now compare forecast performance of the first series according to models B and C, which are non-nested? Each model's Ω_{11} values are equal to one another iff

$$\Gamma_{11}^2(1)/\Gamma_{11}(0) = \Gamma_{21}^2(1)/\Gamma_{22}(0).$$

This relationship is more meaningful when we consider the asymptotic forecast performance, so that the $\Gamma(h)$ matrices represent the DGP's autocovariances. If the true underlying process is indeed a VAR(1), then the above condition is equivalent to

$$\frac{\Phi_{11}^2\,\Gamma_{11}^2(0) + \Phi_{12}^2\,\Gamma_{21}^2(0) + 2\,\Phi_{11}\,\Phi_{12}\,\Gamma_{11}(0)\,\Gamma_{21}(0)}{\Gamma_{11}(0)} = \frac{\Phi_{21}^2\,\Gamma_{11}^2(0) + \Phi_{22}^2\,\Gamma_{21}^2(0) + 2\,\Phi_{21}\,\Phi_{22}\,\Gamma_{11}(0)\,\Gamma_{21}(0)}{\Gamma_{22}(0)}.$$

There are many VAR(1) processes that satisfy this, but one example is furnished by letting Φ be rank one, such that the second column is a multiple of the first, and also imposing that $\Gamma_{11}(0) = \Gamma_{22}(0)$. In this case, the difference of the first and second series will be a moving average process of order one, and the relationship of each series to the lagged series is exactly the same.

This section has provided simple algebraic formulas for the QMLEs of constrained VAR(1) models in terms of the sample autocovariances, much like the case of the LSEs. A general formula for the LSEs is given in Lütkepohl (2006), which we summarize in the next section, making comparisons with the QMLE case. However, as we increase the dimension and the VAR order, the formulas for the QMLEs become more complicated, and in fact instead of a direct formula we obtain coupled linear equations for the VAR parameters and the innovation covariance matrix, just as in the LSE case.

3 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.

3.1 General Theory of QMLE

We consider difference stationary processes (which may be non-Gaussian), 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 (together with the non-deterministic assumption) 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

$$\mathbf{W}_t = \mathbf{m} + \underline{\Psi}(B)\mathbf{A}_t, \tag{6}$$

where the series $\{\mathbf{A}_t\}$ is mean zero and uncorrelated (but possibly dependent) over time with positive definite covariance matrix $\underline{\sigma}^2$. 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

$$\int_{-\pi}^{\pi} \log\det\left[\underline{\Psi}\left(e^{-i\lambda}\right)\,\underline{\Psi}'\left(e^{i\lambda}\right)\right]\,d\lambda = 0. \tag{7}$$

Thus $\underline{\Psi}^{-1}(z)$ is well-defined for $|z| \leq 1$. If our model is correct for the data process, such that (6) holds exactly, then we can write $\mathbf{A}_t = \underline{\Psi}(B)^{-1}\left[\mathbf{W}_t - \mathbf{m}\right]$, 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 (6) 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 (7). 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 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 $\mathcal{S}_+$ 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 (7),

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

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

[2]There cannot be zero eigenvalues in the covariance matrix, as this will interfere with estimation of the model.

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:

$$\Omega(\xi) = \mathbb{E}\left[\mathbf{E}_t\mathbf{E}_t'\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \underline{\Psi}_\xi(e^{-i\lambda})^{-1}\,\underline{F}(\lambda)\,\underline{\Psi}_\xi(e^{i\lambda})^{\dagger}\,d\lambda. \tag{8}$$

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 maximum 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

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

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 $\underline{F} \in \{\underline{F}_\vartheta : \vartheta \in \Xi \times S_+\}$, 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:

$$\begin{aligned} D(\underline{F}_\vartheta, \underline{F}) &= \log\det \underline{\sigma}_\zeta + \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{tr}\left\{\underline{\sigma}_\zeta^{-1}\, \Psi_\xi(e^{-i\lambda})^{-1}\, \underline{F}(\lambda)\, \Psi'_\xi(e^{i\lambda})^{-1}\right\} \\ &= \log\det \underline{\sigma}_\zeta + \operatorname{tr}\left\{\underline{\sigma}_\zeta^{-1}\, \Omega(\xi)\right\}. \end{aligned} \tag{9}$$

This derivation uses (8) 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 $\log\det \underline{\sigma}_\zeta + \operatorname{tr}\left\{\underline{\sigma}_\zeta^{-1}\, \widehat{\Omega}(\xi)\right\}$, 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 $\underline{\sigma}_\zeta \in S_+$. 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 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[3]. Then from (9) we obtain

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

[3]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.

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

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

Our parameter space should be taken to be a compact convex subset Ω of $\Xi \times \operatorname{vec}(\mathcal{S}_+)$ that contains $\vartheta_F = [\xi_F', \operatorname{vec}' \Omega(\xi_F)]'$. 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

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

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.

3.2 Constrained Versus Unconstrained VAR Models

3.2.1 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 consisting of the concatenation of the coefficient matrices of $\underline{\Phi}(z) = 1_m - \sum_{j=1}^p \underline{\phi}_j z^j$. In terms of the notation of the previous section, $\xi = \operatorname{vec} \underline{\phi}$ 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

$$\begin{aligned}\Omega(\xi) &= \Gamma(0) - \sum_{j=1}^{p} \underline{\phi}_j \, \Gamma(-j) - \sum_{k=1}^{p} \Gamma(k) \, \underline{\phi}'_k + \sum_{j,k=1}^{p} \underline{\phi}_j \, \Gamma(k-j) \, \underline{\phi}'_k \\ &= \Gamma(0) - \underline{\phi} \, R_{2:p+1,1} - R_{1,2:p+1} \, \underline{\phi}' + \underline{\phi} \, R_{1:p,1:p} \, \underline{\phi}' \end{aligned} \tag{10}$$

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 2 *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*

$$\sum_{j=1}^{p} \widetilde{\underline{\phi}}_j \Gamma(k-j) = \Gamma(k), \;\; 1 \le k \le p, \tag{11}$$

or $\widetilde{\underline{\phi}} \, \underline{R}_{1:p,1:p} = \underline{R}_{1,2:p+1}$. Furthermore, the corresponding polynomial $\widetilde{\Phi}(z) \in \mathcal{F}_p$ and $\xi_F = vec \, \widetilde{\underline{\phi}}$ uniquely minimizes $\xi \mapsto \det \Omega(\xi)$, with the FEV given by (10). The PTV also minimizes $\xi \mapsto \Omega_{rs}(\xi)$ for every $1 \le r, s \le m$. The PTV for the FEV is

$$\underline{\sigma}_{\zeta_F} = \Omega(\xi_F) = \Gamma(0) - \underline{R}_{1,2:p+1} \, \underline{R}^{-1}_{1:p,1:p} \, \underline{R}_{2:p+1,1}. \tag{12}$$

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 $p \leq n-1$. Then the QMLEs satisfy the empirical YW equations, obtained by replacing $\Gamma(h)$ in (11) by $\widehat{\Gamma}(h)$, and so forth. Convergence of QMLEs to PTVs is guaranteed by results in Taniguchi and Kakizawa (2000).

3.2.2 Properties of Optimization for Constrained Models

Now let us consider the case where the VAR model has some constraints. For example, we might consider the special case discussed in Section 2. There we showed explicitly that the YW solution was different from the QMLE, and that the QMLE does not optimize all the components of the FEV matrix – it only optimizes the FPE. 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 to be zero, we can write

$$\operatorname{vec} \underline{\phi} = J\,\psi + a \tag{13}$$

for a matrix J that is $m^2p \times r$, where $r \leq m^2p$. 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 (13).

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 $\widetilde{\underline{\phi}}$, but it is computed in terms of the PTV $\widetilde{\psi}$, and $\xi_F = \operatorname{vec} \widetilde{\underline{\phi}} = J\,\widetilde{\psi} + a$. Likewise, $\widetilde{\underline{\sigma}} = \underline{\sigma}_{\zeta_F} = \Omega(\xi_F)$ by the previous subsection's general results. Now we can state our result.

Proposition 3 *Let $\{\mathbf{W}_t\}$ be stationary and invertible, with full rank spectral density matrix. Then the PTV $(\widetilde{\psi}, \widetilde{\underline{\sigma}})$ for a fitted constrained VAR(p) with constraints of the form (13) satisfies*

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

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 (13) 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

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

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

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

and $\xi_I = \text{vec}\, \widehat{\underline{\phi}}_{QMLE} = J\, \widehat{\psi}_{QMLE} + a$. These estimates need not result in a stable fitted model (see Section 4).

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 3. In the special case that the true process is a constrained VAR(p) (i.e., the specified model is correct), then $\mathbf{W}_t = \sum_{j=1}^{p} \widetilde{\underline{\phi}}_j \mathbf{W}_{t-j} + \epsilon_t$ and (11) is true. Also, plugging into (10) yields (12), so that Proposition 2 holds for this case. The formula (12) 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 (12) 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

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

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 $\mathrm{vec}\widehat{\underline{\phi}}_{LSE} = J\,\widehat{\psi}_{LSE} + a$, we find the equivalent expression

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

Now $n^{-1}\, Z\, Z' \approx \widehat{\underline{R}}_{1:p,1:p}$ and $n^{-1}\, W\, Z' \approx \widehat{\underline{R}}'_{1,2:p+1}$; 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

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

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 (11)).

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.

4 Numerical Illustrations

4.1 Finite-Sample Results

For constrained bivariate VAR(1) models, the chief fitting methods are MLE, QMLE, or LSE. In Section 2 we provided explicit formulas for QMLEs in terms of covariances, and here we implement these procedures together with the LSE method on four bivariate VAR(1) processes described below. 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.

Recall that model A was the unconstrained bivariate VAR(1), while model B involved the constraint $\Phi_{12} = 0$, and model C involved the constraint $\Phi_{11} = 0$. We investigate fitting models A, B, and C to each process via QMLE and LSE, noting that 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 Σ are in all cases equal to 1_2. These quantities are computed using the formulas of Section 2.2, but using the true autocovariances instead of the empirical estimates. We see that all the Φ PTVs are stable (the absolute trace is always less than two) 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 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.

4.2 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

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

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 subsection, 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 Σ. Into this formula, we would substitute the appropriate PTVs for $\underline{\phi}_1$ and the true DGP 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.

5 Data Application

For an empirical application of the constrained VAR methodology, we present an analysis of revised housing starts data. The choice of this application arises from the *a priori* consideration that time series revisions should follow a stable stationary time series model, because revisions tend not to be easily forecastable. An excellent overview of the source of revisions, as well as an analysis of their dynamics, is given in Jacobs and van Norden (2011); also see the numerous references therein. The analysis here is less sophisticated than that of Jacobs and van Norden (2011), which uses somewhat elaborate state space modeling; our purpose is to offer a realistic application of the methods in Section 3.2, and we do not claim that our model is ideal for forecasting or understanding vintage structure.

Given those provisos, we consider the time series of New Residential Construction, published by the U.S. Census Bureau at `http://www.census.gov/construction/nrc/historical_data/historic_releases.html` focusing on Table 3: New Privately-Owned Housing Units Started. We examine the monthly time series of seasonally unadjusted data, in totals, for the four housing regions of Northeast, Midwest, South, and West. The data is published in four vintages: a preliminary release for a given month, followed by two revisions, and then a final vintage. In practice, the second revision rarely differs at all from the final revision (this is because we are focused on data revisions; revisions due to seasonal adjustment would involve updates for several years worth of vintages). We consider the specific months of January 2008 through December 2012, restricting to those months for which full vintages are available – producing a sample of 57 time points.

For notation, let Y_t^k denote data published at time $t+k$ about time period t (focusing on one of the four geographical regions for the moment). So the first vintage Y_t^1 for time period t is actually published one month later. So $\{Y_t^k : t \in \mathbb{Z}\}$ represents the kth vintage time series. Consider the time series $\{Y_t^2 - Y_t^1 : t \in \mathbb{Z}\}$, $\{Y_t^3 - Y_t^2 : t \in \mathbb{Z}\}$, and so forth; these represent the revision to preliminary (i.e., the first vintage), the revision to the second vintage, and so forth. Summing these time series gives a total revision: the sum of m such series would yield $\{Y_t^{m+1} - Y_t^1 : t \in \mathbb{Z}\}$, giving the revision of the first to $m+1$th vintage from the preliminary. Of course, the real series

we can construct do not extend into the infinite past or future, but the constructed data can be conceptualized as a sample drawn from the above revision vintage series.

Let $\{\mathbf{X_t}\}$ consist of the first, second, and third revision vintage series, and take the dimension to be m. So $\mathbf{X}_t^k = Y_t^{k+1} - Y_t^k$ for $1 \leq k \leq m$. Because the third revision vintage consists mostly of zeroes, for each of the four regions, we eliminate this vintage revision from analysis, restricting to the first two; hence $m = 2$. We analyze each region separately, because once we difference the vintages all trend, cyclical, and seasonal structure is eliminated, and it is hard to believe that there would be correlations between regions associated with survey reporting errors. That is, geographical linkages can only occur through trend, cyclical, or seasonal latent components, and these being removed from the data, there is no reason to analyze the geographies multivariately.

Now as we move down through the components of $\{\mathbf{X_t}\}$, we can expect the samples to have more zero (or small) values, because very little additional revising is occurring. Moreover, empirical experience suggests that the components of $\mathbf{X_t}$ can be cross-correlated, because once a particular data value is revised in a particular direction – due to editing mistakes, late reporting, etc. – it often happens that further vintages are revised in the same direction. Whether or not there is interesting serial correlation is an empirical matter that we explore here.

Due to the construction of the revision vintage series, we expect the dynamics to be stationary; trends or other nonstationarities would enable a high degree of predictability in revisions, thereby obviating the need to publish faulty preliminary estimates – a scenario that is not credible. We therefore focus on stationary models. We apply the VAR methodology of Section 3.2, applying the QMLE methodology to the data, although LSE could also have been used. We first compute a sample mean (which is broadly consistent) from the 57 data points, and then fit constrained VAR models to the mean-centered data. Results from fitting different constrained models are presented in Table 5, where the constraints defining Models II and III varies by region, as suggested by the region's data. Standard errors are computed using (3.4.25) of Lütkepohl (2006) – which states that the asymptotic variance is $\Gamma(0)^{-1} \otimes \sigma$ – and are given in parentheses. We can apply a Gaussian Likelihood Ratio (GLR) test, as defined in Taniguchi and Kakizawa (2000), to discriminate among nested competitors in each case. For each of the four regional datasets, Model III – which involves either two or three restrictions of the entries of $\underline{\phi}_1$ – is preferred via the GLR statistic at significance level .05.

For the forecasting of vintages, we consider the present time to be December 2012, which means preliminary estimates are available for December 2012, November 2012 and October 2012, while first revisions are available for November 2012 and October 2012 and secondary revisions for only October 2012. This means that the secondary revision for November 2012 and December 2012 are unknown, and the first revision for December 2012 is also unknown. We now describe how these quantities can be estimated from forecasts of the revision vintage differences. Letting t denote December 2012, we first aim to estimate Y_{t-1}^3, or the secondary revision for November 2012. We

have

$$\widehat{Y}^3_{t-1} = \widehat{\mathbf{X^2_{t-1}}} + Y^2_{t-1}, \tag{14}$$

which is computed by forecasting the second revision difference one step ahead, and adding the known Y^2_{t-1}, which is the second revised value for November 2012. For any of the constrained VAR(1) models, the h-step ahead forecast is given by adding the mean to the product of Φ^h with $\mathbf{X_{t-2}}$, the last (fully) available value of the revision series.

The next value to estimate is Y^2_t, or the first revision for December 2012. We must forecast $\mathbf{X_{t-1}}$ one step ahead, but note that only $\mathbf{X^1_{t-1}}$ is available – we must impute using the previous estimate (14) to get $\mathbf{X^2_{t-1}}$. So we have

$$\widehat{Y}^2_t = \widehat{\mathbf{X^1_t}} + Y^1_t, \tag{15}$$

where in computing the revision forecast from $\mathbf{X_{t-1}}$, we must first estimate $\mathbf{X^2_{t-1}}$ via $\widehat{Y}^3_{t-1} - Y^2_{t-1}$, using (14).

Finally, we need to estimate Y^3_t, or the second revision for December 2012. Now the formula is

$$\widehat{Y}^3_t = \widehat{\mathbf{X^2_t}} + \widehat{Y}^2_t, \tag{16}$$

utilizing (15). In this manner the vintage triangle of unknown values can be forecasted and imputed. However, we cannot obtain estimates of future values at time $t+1$, unless we model one of the raw vintage series together with the vintage differences.

The projected values, along with the true values subsequently published, are presented in Table 6. Note that the rows correspond to times t, $t-1$, and $t-2$, while the columns are the vintages 1,2, or 3. Projection is based upon the numbers in the lower triangle, in bold, whereas the estimates in the upper triangle are in parentheses besides the true publications. In the case of the NE the forecasted revisions have quite a bit of error, although the data values were somewhat anomalous in this case. The error is tolerable for the other three regions' vintage triangle, providing potentially useful results.

Appendix

A.1 Derivation of Optima for Models B and C.

First consider model B; the derivations for model C are similar in nature. We note that Ω_{22} times (4) plus Ω_{12} times (3) is equal to $-2|\Omega|$ times $\Gamma_{21}(1) - \Gamma_{11}(0)\Phi_{21} - \Gamma_{12}(0)\Phi_{22}$, which will also be zero. By the definition of Ω as FEV, it must be positive definite for all values of Φ, and hence $|\Omega| > 0$ must always hold. Hence we obtain the condition

$$\Gamma_{21}(1) - \Gamma_{11}(0)\Phi_{21} - \Gamma_{12}(0)\Phi_{22} = 0. \tag{A.1}$$

This relation can be substituted back into (3) and (4), which produces

$$(\Gamma_{11}(1) - \Gamma_{11}(0)\Phi_{11})\,\Omega_{22} = 0 \qquad (\Gamma_{11}(1) - \Gamma_{11}(0)\Phi_{11})\,\Omega_{12} = 0.$$

Note that if $\Omega_{22} = 0$, then necessarily $|\Omega| \leq 0$. So this cannot be. Therefore we must have the relation

$$\Gamma_{11}(1) - \Gamma_{11}(0)\Phi_{11} = 0. \tag{A.2}$$

Note then that (A.1) and (A.2) imply both (3) and (4) are zero, and they also imply

$$\Phi_{11} = \Gamma_{11}(1)/\Gamma_{11}(0) \tag{A.3}$$

$$\Phi_{22} = \Gamma_{21}(1)/\Gamma_{21}(0) - \Phi_{21}\Gamma_{11}(0)/\Gamma_{21}(0). \tag{A.4}$$

Substituting these solutions back into the formulas for Ω_{11} and Ω_{12}, utilizing (A.1), yields

$$\begin{aligned}
\Omega_{11} &= \Gamma_{11}(0) - \Gamma_{11}^2(1)/\Gamma_{11}(0) \\
\Omega_{12} &= \Gamma_{12}(0) - \Gamma_{12}(1)\Gamma_{21}(1)/\Gamma_{12}(0) - \Phi_{21}\left(\Gamma_{11}(1) - \Gamma_{12}(1)\Gamma_{11}(0)/\Gamma_{21}(0)\right).
\end{aligned}$$

These quantities can now be utilized in (5), along with (A.3) and (A.4) to produce the system $B_1\left(B_2 - \Phi_{21}B_3\right) = \left(B_4 - \Phi_{21}B_5\right)B_6$, where the B_j are defined via

$$\begin{aligned}
B_1 &= \Gamma_{11}(0) - \Gamma_{11}^2(1)/\Gamma_{11}(0) \\
B_2 &= \Gamma_{22}(1) - \Gamma_{22}(0)\Gamma_{21}(1)/\Gamma_{21}(0) \\
B_3 &= \Gamma_{12}(0) - \Gamma_{22}(0)\Gamma_{11}(0)/\Gamma_{21}(0) \\
B_4 &= \Gamma_{12}(0) - \Gamma_{12}(1)\Gamma_{21}(1)/\Gamma_{21}(0) \\
B_5 &= \Gamma_{11}(1) - \Gamma_{12}(1)\Gamma_{11}(0)/\Gamma_{21}(0) \\
B_6 &= \Gamma_{12}(1) - \Gamma_{12}(0)\Gamma_{11}(1)/\Gamma_{11}(0) \\
B_7 &= \Gamma_{21}(1) - \Gamma_{11}(0)\Gamma_{22}(1)/\Gamma_{21}(0) \\
B_8 &= \Gamma_{11}(0) - \Gamma_{21}(1)\Gamma_{11}(1)/\Gamma_{21}(0).
\end{aligned}$$

Solving this system finally yields

$$\Phi_{21} = \frac{B_1B_2 - B_4B_6}{B_1B_3 - B_5B_6}, \tag{A.5}$$

which together with (A.3) and (A.4) provides a complete solution. A final formula for Φ_{22} is given by

$$\Phi_{22} = \frac{B_1B_7 + B_6B_8}{B_1B_3 - B_5B_6.} \tag{A.6}$$

For model C, we define constants C_j as follows:

$$\begin{aligned}
C_1 &= \Gamma_{11}(0) - \Gamma_{21}^2(1)/\Gamma_{22}(0) \\
C_2 &= \Gamma_{11}(1) - \Gamma_{21}(0)\Gamma_{21}(1)/\Gamma_{22}(0) \\
C_3 &= \Gamma_{12}(1) - \Gamma_{22}(1)\Gamma_{11}(0)/\Gamma_{21}(0) \\
C_4 &= \Gamma_{21}(0) - \Gamma_{22}(0)\Gamma_{11}(0)/\Gamma_{21}(0) \\
C_5 &= \Gamma_{12}(0) - \Gamma_{22}(1)\Gamma_{11}(1)/\Gamma_{21}(0) \\
C_6 &= \Gamma_{21}(1) - \Gamma_{11}(1)\Gamma_{22}(0)/\Gamma_{21}(0) \\
C_7 &= \Gamma_{22}(1) - \Gamma_{22}(0)\Gamma_{12}(1)/\Gamma_{21}(0) \\
C_8 &= \Gamma_{22}(0) - \Gamma_{22}(1)\Gamma_{21}(1)/\Gamma_{21}(0).
\end{aligned}$$

These values allow us to express the optimizers of model C.

A.2 Proofs of Results

Proof of Proposition 1. The roots of the polynomial $1 - \Phi z$ are

$$\zeta = \frac{\mathrm{tr}\Phi \pm \sqrt{\mathrm{tr}^2\Phi - 4\,\det\Phi}}{2\,\det\Phi}.$$

These roots are complex conjugate iff $(\Phi_{11} - \Phi_{22})^2 < -4\Phi_{12}\,\Phi_{21}$, in which case the square magnitude of them is equal to $1/\det\Phi$. The condition of conjugacy ensures that $\det\Phi > \mathrm{tr}^2\Phi/4$, so that the determinant is guaranteed to be positive. Hence in this case, stability – which means that $|\zeta| > 1$ – is equivalent to the condition that $\det\Phi < 1$.

Now in the case that $\det\Phi \leq \mathrm{tr}^2\Phi/4$, we obtain two real roots of the polynomial $1 - \Phi z$. In this case we require $\zeta^2 > 1$, which holds iff

$$2\overset{2}{\det}\Phi < \mathrm{tr}^2\Phi - 2\,\det\Phi \pm \mathrm{tr}\Phi\,\sqrt{\mathrm{tr}^2\Phi - 4\,\det\Phi}.$$

These two constraints simplify to the equivalent condition that $\mathrm{tr}^2\Phi < (1 + \det\Phi)^2$. Now putting both cases together yields the equivalent conditions that $|\mathrm{tr}\Phi| < 1 + \det\Phi$ and that $|\det\Phi| < 1$. □

Proof of Proposition 2. To show this result requires some notation. For any matrix, we use a dot in a subscript to denote a free index that is summed over. Let ξ_ℓ correspond to the p, qth entry of the jth coefficient matrix $\underset{\sim}{\phi}_j$, denoted as $\xi_\ell = \underset{\sim}{\phi}_j^{pq}$. Then $\frac{\partial\Omega_{rs}}{\partial\underset{\sim}{\phi}_j^{pq}}$ equals

$$-\delta_{\{r=p\}}\Gamma_{qs}(j) - \delta_{\{s=p\}}\Gamma_{qr}(j) + \delta_{\{r=p\}}\left[\Gamma_{q\cdot}(j-1)\;\Gamma_{q\cdot}(j-2)\cdots\right]\left[\underset{\sim}{\phi}'\right]_{\cdot s} + \delta_{\{s=p\}}\underset{\sim}{\phi}_{r\cdot}\begin{bmatrix}\Gamma_{\cdot q}(1-j)\\ \Gamma_{\cdot q}(2-j)\\ \vdots\end{bmatrix},$$

where δ_A equals one if the set A is true and is zero otherwise. By collecting terms, this can be expressed as

$$\underline{\phi}\underline{R}_{1:p,1:p} = \underline{R}_{1,2:p+1}, \tag{A.7}$$

i.e., the YW equations. Therefore the unique $\underline{\phi}$ satisfying (A.7) is also a critical point of all the component functions Ω_{rs}, and hence $\det\Omega$ as well. The formula (11) follows by writing out (A.7). That the solution corresponds to a polynomial in F_p is shown in Whittle (1963). The formula for the FEV follows by evaluation.

We also need to establish that $\widetilde{\underline{\phi}}$ satisfying the YW equations is not only a critical point, but a minimizer of the FPE. Suppose that $\underline{\phi}$ is any other parameter matrix with corresponding $\xi = \mathrm{vec}\underline{\phi}$, and consider

$$\underline{\Phi}(B)\mathbf{W}_t = \widetilde{\underline{\Phi}}(B)\mathbf{W}_t + \sum_{k=1}^{p}\left(\underline{\phi}_k - \widetilde{\underline{\phi}}_k\right)\mathbf{W}_{t-k}.$$

The two summands on the right hand side are uncorrelated with one another due to (11). Then defining the covariance matrix $\Sigma_{\underline{\phi}-\widetilde{\underline{\phi}}} = \mathbb{E}\left\{(\underline{\Phi}-\widetilde{\underline{\Phi}})(B)\mathbf{W}_t\right\}\left\{(\underline{\Phi}-\widetilde{\underline{\Phi}})(B)\mathbf{W}_t\right\}'$, it follows that $\Omega(\xi) = \Omega_{\widetilde{\xi}} + \Sigma_{\underline{\phi}-\widetilde{\underline{\phi}}}$. Clearly $\Sigma_{\underline{\phi}-\widetilde{\underline{\phi}}} \geq 0$ and $\Sigma_{\underline{\phi}-\widetilde{\underline{\phi}}} = 0$ would imply $(\underline{\Phi}-\widetilde{\underline{\Phi}})(B)\mathbf{W}_t = 0$ with probability one, in contradiction to the full rank property of the spectral density of $\{\mathbf{W}_t\}$. Therefore $\Sigma_{\underline{\phi}-\widetilde{\underline{\phi}}} \gneqq 0$ and $\left|\Omega_{\widetilde{\xi}}\right| < |\Omega(\xi)|$ follows. This shows that the YW solution $\widetilde{\underline{\phi}}$ minimizes FPE, and moreover is unique. □

Proof of Proposition 3. To derive the result we work with the KL directly, given by (9), which in the case of a VAR(p) takes the form

$$\begin{aligned} D(\underline{F}_\vartheta, \underline{F}) = \log\det\underline{\sigma}_\zeta + \mathrm{tr}\left(\underline{\sigma}_\zeta^{-1}\,\Gamma(0)\right) - \mathrm{tr}\left(\underline{\sigma}_\zeta^{-1}\,\Phi\,\underline{R}_{1,2:p+1}\right) - \mathrm{tr}\left(\underline{R}_{2:p+1,1}\,\Phi'\,\underline{\sigma}_\zeta^{-1}\right) \\ + \mathrm{tr}\left(\underline{\sigma}_\zeta^{-1}\,\Phi\,\underline{R}_{1:p,1:p}\,\Phi'\right). \end{aligned}$$

Now we unpack each of these expressions further:

$$\begin{aligned} \mathrm{tr}\left(\underline{\sigma}_\zeta^{-1}\,\Phi\,\underline{R}_{1,2:p+1}\right) &= [\mathrm{vec}(1_m)]'\,\left[\underline{R}'_{1,2:p+1}\otimes\underline{\sigma}_\zeta^{-1}\right]\,[J\,\psi + a] \\ \mathrm{tr}\left(\underline{\sigma}_\zeta^{-1}\,\Phi\,\underline{R}_{1:p,1:p}\,\Phi'\right) &= \left[a' + \psi'\,J'\right]\,\left[\underline{R}_{1:p,1:p}\otimes\underline{\sigma}_\zeta^{-1}\right]\,[J\,\psi + a]. \end{aligned}$$

It is then evident that KL has a quadratic format, and that the gradient with respect to ψ equals

$$-2\,\left[\underline{R}'_{1,2:p+1}\otimes\underline{\sigma}_\zeta^{-1}\right]\;J + 2\,J'\,\left[\underline{R}_{1:p,1:p}\otimes\underline{\sigma}_\zeta^{-1}\right]\,[J\,\psi + a].$$

Setting this to zero and solving yields the formula for the PTV, with $\widetilde{\underline{\sigma}}$ replacing $\underline{\sigma}_\zeta$. The formula for the innovation variance follows from the general treatment of Section 3.1.

Now we demonstrate that this critical point really is a minimizer of KL, adapting the argument from Proposition 2. From (10), and some algebra, we obtain

$$\Omega(\xi) = \Omega(\widetilde{\xi}) + \Sigma_{\underline{\phi}-\widetilde{\underline{\phi}}} + \underline{\tau}\,(\underline{\phi}-\widetilde{\underline{\phi}})' + (\underline{\phi}-\widetilde{\underline{\phi}})\,\tau' \tag{A.8}$$

for any $\underline{\phi}$, where $\tau = \widetilde{\underline{\phi}} R_{1:p,1:p} - R_{1,2:p+1}$. Of course $\tau = 0$ whenever the Yule-Walker equations are satisfied, but this need not be the case for constrained VAR PTVs. Instead, the PTVs satisfy

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

which follows from algebraic manipulation of the formula for $\widetilde{\psi}$. Multiplying (A.8) by $\widetilde{\sigma}^{-1}$ and taking the trace, we obtain

$$\operatorname{tr}\left(\widetilde{\sigma}^{-1}\, \Omega(\widetilde{\xi})\right) = \operatorname{tr}\left(\widetilde{\sigma}^{-1}\, \Omega(\xi)\right) + \operatorname{tr}\left(\widetilde{\sigma}^{-1}\, \Sigma_{\underline{\phi}-\widetilde{\underline{\phi}}}\right).$$

This is true because by (13)

$$\operatorname{tr}\left(\widetilde{\sigma}^{-1}\, \tau\, (\underline{\phi} - \widetilde{\underline{\phi}})'\right) = \left[\operatorname{vec}(\underline{\phi} - \widetilde{\underline{\phi}})\right]' \operatorname{vec}\left[\widetilde{\sigma}^{-1}\tau\right] = (\xi - \widetilde{\xi})' J' \operatorname{vec}\left[\widetilde{\sigma}^{-1}\tau\right],$$

which is identically zero for any ξ. Now because $\widetilde{\sigma}$ is positive definite, we see that $\operatorname{tr}\left(\widetilde{\sigma}^{-1}\, \Sigma_{\underline{\phi}-\widetilde{\underline{\phi}}}\right)$ is non-negative, and is nonzero by the full rank assumption unless $\psi = \widetilde{\psi}$. Hence $\operatorname{tr}\left(\widetilde{\sigma}^{-1}\, \Omega(\widetilde{\xi})\right) > \operatorname{tr}\left(\widetilde{\sigma}^{-1}\, \Omega(\xi)\right)$, which establishes optimality. □

References

[1] Akaike, H. (1969) Fitting autoregresive models for prediction. *Annals of the Institute of Statistical Mathematics* **21**, 243–247.

[2] Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723.

[3] Akaike, H. and Kitagawa, G. (1999) *The Practice of Time Series Analysis.* New York: Springer.

[4] Brockwell, P. and Davis, R. (1991) *Time Series: Theory and Methods, 2nd Ed.* New York: Springer.

[5] Golub, G. and Van Loan, C. (1996) *Matrix Computations.* Baltimore: Johns Hopkins University Press.

[6] Jacobs, J. and Van Norden, S. (2011) Modeling data revisions: measurement error and dynamics of "true" values. *Journal of Econometrics* **161**, 101–109.

[7] Lütkepohl, H. (2006) *New Introduction to Multiple Time Series Analysis.* Springer-Verlag, Berlin.

[8] Magnus, J. and Neudecker, H. (1999) *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, New York.

[9] Maïnassara, B. and Francq, C. (2011) Estimating structural VARMA models with uncorrelated but non-independent error terms. *Journal of Multivariate Analysis* **102**, 496–505.

[10] McElroy, T., and McCracken, M. (2012) Multi-Step Ahead Forecasting of Vector Time Series. *Federal Reserve Bank of St. Louis, Working Papers*, 2012-060A.

[11] Pinheiro, J. and Bates, D. (1996) Unconstrained parameterizations for variance-covariance matrices. *Statistics and Computing* **6**, 289–296.

[12] Reinsel, G. (1997) *Elements of Multivariate Time Series Analysis.* 2nd Ed. Springer-Verlag, New York.

[13] Schorfheide, F. (2005) VAR forecasting under misspecification. *Journal of Econometrics* **128**, 99–136.

[14] Taniguchi, M. and Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series.* Springer-Verlag, New York.

[15] Whittle, P. (1953) The analysis of multiple stationary time series. *Journal of the Royal Statistical Society, Series B* **15**, 125–139.

[16] Whittle, P. (1963) *Prediction and Regulation.* English Universities Press.

[17] Wilks, S. (1932) Certain generalizations in the analysis of variance. *Biometrika* **24**, 471–494.

Parameter Estimates						
Parameters	Models					
n=100	qmle Model A	qmle Model B	qmle Model C	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.808	.652	.734	.817	.658	.742
Min $\|\zeta\|$	.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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.821	.662	.748	.826	.665	.752
Min $\|\zeta\|$	.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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.826	.667	.753	.828	.668	.755
Min $\|\zeta\|$	.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 $\Sigma = 1_2$. 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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.642	.648	.646	.649	.654	.665
Min $\|\zeta\|$	.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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.643	.657	.668	.646	.661	.674
Min $\|\zeta\|$	.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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.646	.661	.679	.647	.663	.678
Min $\|\zeta\|$	.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 $\Sigma = 1_2$. 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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.913	.922	.926	.923	.932	.946
Min $\|\zeta\|$	.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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.932	.936	.946	.937	.941	.954
Min $\|\zeta\|$	.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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.942	.943	.957	.944	.946	.958
Min $\|\zeta\|$	.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 $\Sigma = 1_2$. 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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.707	1.199	.808	.714	1.243	.687
Min $\|\zeta\|$	.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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.722	1.224	.819	.726	1.247	.671
Min $\|\zeta\|$	.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	ols Model A	ols Model B	ols 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 $\|\zeta\|$	.737	1.237	.825	.739	1.249	.659
Min $\|\zeta\|$	.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 $\Sigma = 1_2$. 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		
NE	qmle Model I	qmle Model II	qmle Model III
Φ	.025 (.132) .152 (.192) .023 (.082) .422 (.120)	0 .153 (.192) .023 (.082) .422 (.120)	0 .153 (.192) 0 .423 (.120)
Σ	.335 -.005 -.005 .131	.335 -.005 -.005 .131	.335 -.005 -.005 .131
$\log\det\Sigma$	-3.1292	-3.1285	-3.1271
MW	qmle Model I	qmle Model II	qmle Model III
Φ	.145 (.135) -.096 (.213) .049 (.084) .263 (.132)	.165 (.131) -.088 (.213) 0 .243 (.128)	.179 (.126) 0 0 .230 (.125)
Σ	.328 -.051 -.051 .126	.328 -.051 -.051 .127	.329 -.051 -.051 .127
$\log\det\Sigma$	-3.2484	-3.2423	-3.2393
S	qmle Model I	qmle Model II	qmle Model III
Φ	.070 (.132) -.075 (.198) -.036 (.086) .222 (.129)	.074 (.132) -.075 (.198) 0 .219 (.129)	0 0 0 .222 (.129)
Σ	.770 .036 .036 .327	.770 .036 .036 .328	.776 .034 .034 .328
$\log\det\Sigma$	-1.3838	-1.3809	-1.3732
W	qmle Model I	qmle Model II	qmle Model III
Φ	.172 (.131) .172 (.286) .013 (.061) .062 (.133)	.173 (.131) .155 (.284) .017 (.061) 0	.169 (.130) .155 (.283) 0 0
Σ	.432 .026 .026 .094	.432 .026 .026 .094	.432 .026 .026 .094
$\log\det\Sigma$	-3.2210	-3.2172	-3.2158

Table 5: Fitted constrained VAR(1) models to revision vintage data for NE, MW, S, and W housing starts. In each case, Model I is fully unconstrained, whereas Model II and Model III represent different constrained models, wherein some entries of $\underline{\phi}_1$ are constrained to be zero. In each case, Model III cannot be rejected according to the GLR statistic.

Forecasts			
Dates	Vintages		
NE	Preliminary	First Revision	Second Revision
Dec 2012	**5.8**	8.0 (5.853)	7.9 (5.876)
Nov 2012	**5.7**	**5.4**	5.1 (5.445)
Oct 2012	**6.6**	**7.0**	**7.1**
MW	Preliminary	First Revision	Second Revision
Dec 2012	**11.6**	11.2 (11.690)	11.0 (11.747)
Nov 2012	**12.5**	**12.4**	12.1 (12.484)
Oct 2012	**15.3**	**14.8**	**15.0**
S	Preliminary	First Revision	Second Revision
Dec 2012	**31.1**	30.4 (31.360)	30.6 (31.467)
Nov 2012	**34.0**	**33.3**	33.0 (33.242)
Oct 2012	**36.9**	**38.1**	**37.3**
W	Preliminary	First Revision	Second Revision
Dec 2012	**12.9**	12.9 (12.813)	13.6 (12.882)
Nov 2012	**12.3**	**12.2**	11.9 (12.268)
Oct 2012	**19.1**	**17.3**	**17.7**

Table 6: Data and forecasts of the vintage triangle, using Model III of Table 5 for revision vintage data of NE, MW, S, and W housing starts. Data available as of Jan 2013 publication date is in bold, while unknown values are in normal font, with their forecasts in parentheses.

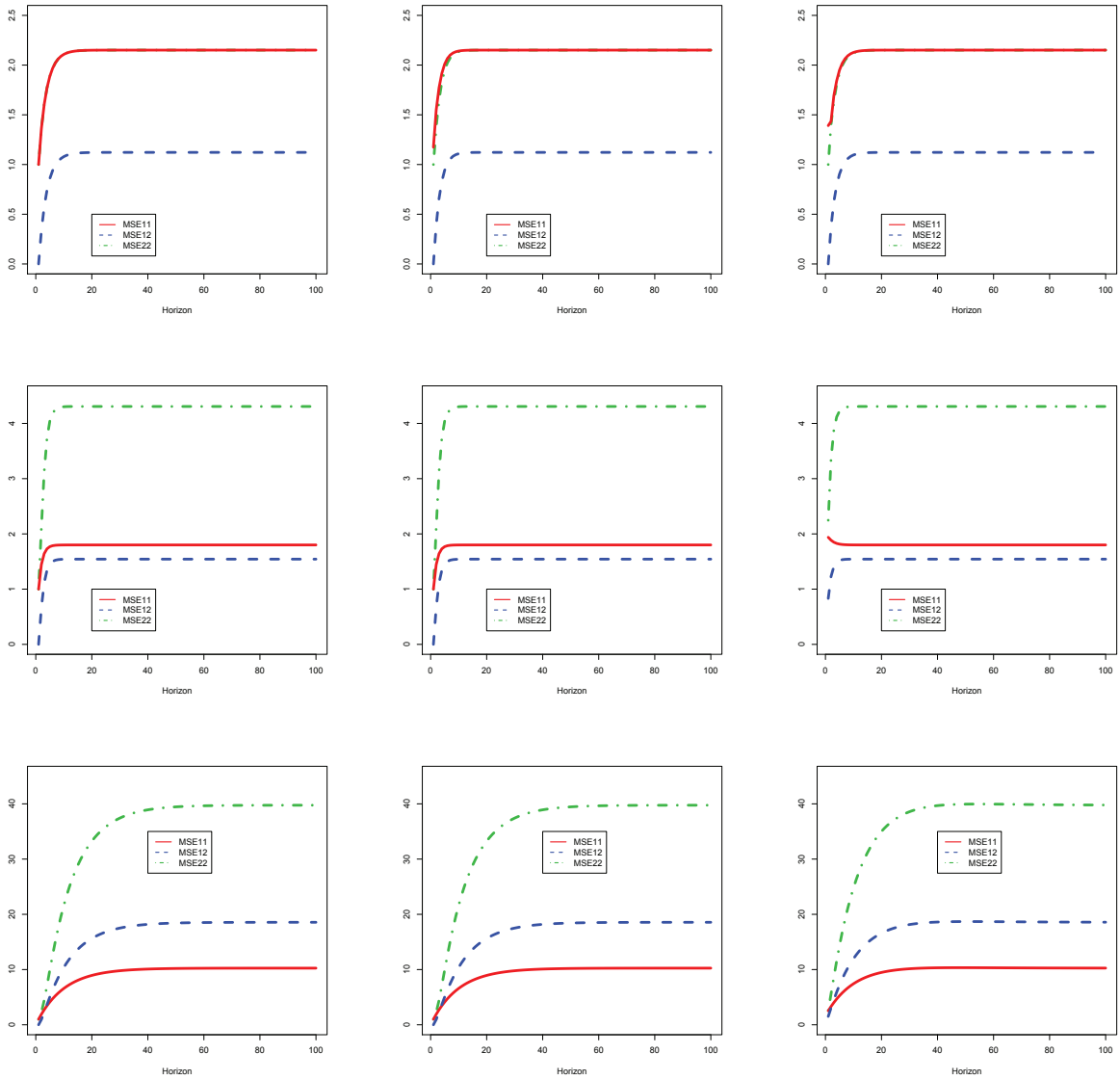

Figure 1: Asymptotic forecast MSE as a function of forecast horizon. In each panel, the entries of the FEV matrix are plotted, with the first diagonal entry in Red (Solid), the second diagonal entry in Green (Dotted-Dashed), and the off-diagonal in Blue (Dashed). The first row of panels corresponds to Process 1 of Section 4.1, 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.