INTERNATIONAL
ECONOMIC
REVIEW

May 2002
Vol. 43, No. 2

BOOTSTRAPPING SMOOTH FUNCTIONS OF SLOPE PARAMETERS AND INNOVATION VARIANCES IN VAR(∞) MODELS*

By Atsushi Inoue and Lutz Kilian[1]

North Carolina State University, U.S.A.;
University of Michigan, U.S.A., and CEPR, U.K.

It is common to conduct bootstrap inference in vector autoregressive (VAR) models based on the assumption that the underlying data-generating process is of finite-lag order. This assumption is implausible in practice. We establish the asymptotic validity of the residual-based bootstrap method for smooth functions of VAR slope parameters and innovation variances under the alternative assumption that a sequence of finite-lag order VAR models is fitted to data generated by a VAR process of possibly infinite order. This class of statistics includes measures of predictability and orthogonalized impulse responses and variance decompositions. Our approach provides an alternative to the use of the asymptotic normal approximation and can be used even in the absence of closed-form solutions for the variance of the estimator. We illustrate the practical relevance of our findings for applied work, including the evaluation of macroeconomic models.

1. INTRODUCTION

It is common in applied vector autoregressive (VAR) analysis to condition on the assumption that the lag order of the VAR data-generating process (DGP) is finite. The implausibility of finite-lag order VAR models has been pointed out by Braun and Mittnik (1993), among others, but the finite-lag order assumption continues to play a central role in econometric inference in practice.

The fact that the DGP is thought to be represented by a VAR(∞) process has important implications for VAR inference. For example, Lütkepohl and Poskitt (1991) show that, although the VAR impulse response estimator retains its asymptotic normal distribution in the infinite-lag order case, its asymptotic variance is a nondecreasing function of the forecast horizon. Unlike in the finite-lag order

* Manuscript received April 1999; revised April 2000.
[1] We thank Bob Barsky, Larry Christiano, Paul Fackler, Nicolai Gospodinov, Alastair Hall, Ken West, Mark Watson, Jonathan Wright, and two anonymous referees for helpful comments and suggestions. E-mail: *lkilian@umich.edu.*

case, the asymptotic variance does not converge to zero as the horizon gets large. The additional sampling uncertainty arises from the thought experiment that the lag order is allowed to grow to infinity with the sample size. Thus, the resulting delta-method intervals are quite different from traditional intervals for the finite-lag order model (see Lütkepohl, 1990, p. 122, for an illustrative example).

In this article, we explore an alternative approach to inference in VAR(∞) models based on the bootstrap method. Although the asymptotic validity of the bootstrap method for inference on standard statistics such as orthogonalized impulse responses or variance decompositions is well established for finite-lag order VAR models (e.g., Bose, 1988; Kilian, 1998), no corresponding theoretical results are available for VAR(∞) models. We will demonstrate both the theoretical validity and the practical feasibility of the bootstrap proposal. Our results cover orthogonalized impulse responses and variance decompositions but also extend to other smooth functions of slope parameters and innovation variances for which closed-form solutions for the asymptotic variance are not available in the VAR(∞) case, such as measures of predictability (see Granger and Newbold, 1986; Diebold and Kilian, 2000).

In related work, Paparoditis (1996) proves the asymptotic validity of bootstrapping the autoregressive coefficients of the VAR(∞) model by means of a sequence of finite-order autoregressive approximations. Paparoditis shows that if the autoregressive lag order increases at a suitable rate with the sample size, the bootstrap approximations of the distributions of these estimators are as sound asymptotically as conventional large sample Gaussian approximations. His results also extend to the implied moving-average coefficients (reduced-form impulse responses), but they do not cover nonlinear functions of *both* slope parameters and innovation variances such as orthogonalized impulse responses and variance decompositions or predictability measures.[2]

The remainder of the article is organized as follows: In Section 2, we motivate and describe the proposed bootstrap algorithm. Section 3 contains the main theoretical results. Details of the proofs are relegated to the Appendix. Section 4 contains some Monte Carlo evidence of the small-sample properties of the proposed bootstrap procedure. In Section 5, we illustrate the practical usefulness of our results for the econometric evaluation of macroeconomic models. We conclude in Section 6.

2. THE BOOTSTRAP ALGORITHM FOR VAR(∞) MODELS

VAR analysis plays an important role in empirical macroeconomics. It is well known that the reduced-form representation of dynamic general equilibrium (DGE) macroeconomic models will in general not have a finite-lag order VAR representation but often can be represented as a vector autoregressive moving-average (VARMA) process. It may therefore seem that in conducting macroeconometric

[2] Paparoditis (1996) focuses on autoregressive slope parameters, the number of which grows with the sample size. In contrast, the theoretical results in Bühlmann (1997) for the univariate autoregressive sieve model cover nonlinear functions of the data that depend on a fixed number of lags (such as autocorrelation coefficients). Bühlmann's results do not apply if one is interested in bootstrapping statistics such as impulse responses and variance decompositions.

analysis we should condition on the particular reduced-form VARMA structure implied by the theoretical economic model.[3]

It turns out that this proposal is neither theoretically appealing nor practical: First, we have little confidence in the VARMA specification implied by a given DGE model, because that specification depends on inherently atheoretical assumptions about the exogenous driving processes and on the absence of measurement error and aggregation of various forms (see Lütkepohl, 1993, p. 230, for further discussion). In fact, in some cases, there may not even exist a finite-lag order VARMA representation. Second, and more important, even if some finite-order VARMA structure provides a good approximation to the reduced form, VARMA models of the large dimensions of interest in most empirical work are notoriously difficult to estimate in practice. This may explain why there are few, if any, examples of VARMA models with more than two variables in applied macroeconometrics.

These considerations suggest that applied users consider an alternative class of reduced-form representations known as VAR(∞) processes (see Bühlmann, 1995). Note that all stable invertible VARMA models can be represented as VAR(∞) processes. Assuming an exponential rate of decay of the coefficients of the autoregressive representation of the VAR(∞) process, the VAR(∞) process may be approximated by a sequence of finite-lag order VAR models, where the order k of the approximating model increases at a suitable rate with the sample size. Unlike VARMA models, which have to be estimated by numerical methods, approximating VAR(k) models are easy to fit by standard least-squares (LS) techniques.

In this article, we consider bootstrap inference based on such models. Consider a vector time series $\{y_t\}_{t=1}^{T}$, generated from

$$y_t = \sum_{j=1}^{\infty} A_j y_{t-j} + \varepsilon_t$$

where $A(k) = (A_1, A_2, \ldots, A_k)$ and Σ is the covariance matrix of the independent and identically distributed (iid) innovations ε_t. All deterministic components are assumed to have been removed. Suppose that the statistic of interest is a smooth function of $A(k)$ and Σ. Then bootstrap approximations to the distribution of this statistic may be constructed as follows:

(1) Use the LS method to estimate the approximating VAR(k) model:

$$y_t = A_{1,k} y_{t-1} + \cdots + A_{k,k} y_{t-k} + \varepsilon_{k,t}$$

Denote the LS estimate of $A(k)$ by $\hat{A}(k) = (\hat{A}_{1,k}, \hat{A}_{2,k}, \ldots, \hat{A}_{k,k})$ and the LS estimate of Σ by $\hat{\Sigma}_k = \sum_{t=k+1}^{T} \hat{\varepsilon}_{k,t} \hat{\varepsilon}_{k,t}^{\mathrm{T}} / (T-k)$, where $\hat{\varepsilon}_{k,t} = \tilde{\varepsilon}_{k,t} - \sum_{t=k+1}^{T} \tilde{\varepsilon}_{k,t} / (T-k)$ and $\tilde{\varepsilon}_{k,t} = y_t - \hat{A}_{1,k} y_{t-1} - \cdots - \hat{A}_{k,k} y_{t-k}$.

(2) Generate $T-k$ bootstrap innovations ε_t^* by random sampling with replacement from the centered regression residuals $\hat{\varepsilon}_{k,t}$, $t = k+1, \ldots, T$.

(3) Generate a random draw for the vector of k initial observations $Y_0^* = (y_1^*, \ldots, y_k^*)$, as described by Berkowitz and Kilian (2000).

[3] Methods for constructing confidence intervals for many of the statistics of interest in VARMA models have been proposed, for example, by Mittnik and Zadrozny (1993).

(4) Given Y_0^*, $\hat{A}(k)$ and $\{\varepsilon_t^*\}_{t=k+1}^T$, generate a sequence of pseudodata of length T from the recursion

$$y_t^* = \hat{A}_{1,k} y_{t-1}^* + \cdots + \hat{A}_{k,k} y_{t-k}^* + \varepsilon_t^*$$

Fit a VAR(k) model to $\{y_t^*\}_{t=1}^T$ and calculate the bootstrap LS regression estimates $\hat{A}^*(k) = (\hat{A}_{1,k}^*, \hat{A}_{2,k}^*, \ldots, \hat{A}_{k,k}^*)$ and $\hat{\Sigma}_k^* = \sum_{t=k+1}^T \hat{\varepsilon}_{k,t}^* \hat{\varepsilon}_t^{*\mathrm{T}}/(T-k)$, where $\hat{\varepsilon}_{k,t}^* = \tilde{\varepsilon}_{k,t}^* - \sum_{t=k+1}^T \tilde{\varepsilon}_{k,t}^*/(T-k)$ and $\tilde{\varepsilon}_{k,t}^* = y_t^* - \hat{A}_{1,k}^* y_{t-1}^* - \cdots - \hat{A}_{k,k}^* y_{t-k}^*$. Use these LS estimates to compute the bootstrap analog of the statistic of interest.

(5) Repeat steps 2–4 until the empirical distribution of the statistic of interest is approximated to the desired degree of accuracy.

3. THEORETICAL RESULTS

Let $Y_{t,k} = (y_t^{\mathrm{T}}, y_{t-1}^{\mathrm{T}}, \ldots, y_{t-k+1}^{\mathrm{T}})^{\mathrm{T}}$, $\hat{\Gamma}_{1,k} = (T-k)^{-1}\sum_{t=k}^{T-1} Y_{t,k} y_{t+1}^{\mathrm{T}}$, and $\hat{\Gamma}_k = (T-k)^{-1} \sum_{t=k}^{T-1} Y_{t,k} Y_{t,k}^{\mathrm{T}}$. Let $a(k) = \mathrm{vec}(A(k))$.[4] The LS estimator of $a(k)$ based on the kth order approximating vector autoregression is denoted by $\hat{a}(k) = \mathrm{vec}(\hat{A}(k))$, where $\hat{A}(k) = \hat{\Gamma}_{1,k}^{\mathrm{T}} \hat{\Gamma}_k^{-1}$. Denote the corresponding bootstrap LS estimator by $\hat{A}^*(k) = \hat{\Gamma}_{1,k}^{*\mathrm{T}} \hat{\Gamma}_k^{*-1}$, where $Y_{t,k}^* = (y_t^{*\mathrm{T}}, y_{t-1}^{*\mathrm{T}}, \ldots, y_{t-k+1}^{*\mathrm{T}})^{\mathrm{T}}$, $\hat{\Gamma}_{1,k}^* = (T-k)^{-1}\sum_{t=k}^{T-1} Y_{t,k}^* y_{t+1}^{*\mathrm{T}}$ and $\hat{\Gamma}_k^* = (T-k)^{-1} \sum_{t=k}^{T-1} Y_{t,k}^* Y_{t,k}^{*\mathrm{T}}$. Then, $\hat{a}^*(k) = \mathrm{vec}(\hat{A}^*(k))$ and $\hat{\Sigma}_k^* = (T-k)^{-1}\sum_{t=k+1}^T \hat{\varepsilon}_{k,t}^* \hat{\varepsilon}_{k,t}^{*\mathrm{T}}$ are the bootstrap versions of $\hat{a}(k)$ and $\hat{\Sigma}_k$, respectively.

ASSUMPTION 1.

$$y_t = \sum_{j=1}^{\infty} A_j y_{t-j} + \varepsilon_t$$

(a) $\{\varepsilon_t\}$ *is a sequence of iid r-dimensional random vectors with* $E(\varepsilon_t) = 0_{r\times 1}$, $E(\varepsilon_t \varepsilon_t^{\mathrm{T}}) = \Sigma$ *and finite-dimensional moments up to the eighth order.*
(b) $\det(I_r - \sum_{j=1}^{\infty} A_j z^j) \neq 0$ *for all* $|z| \leq 1$.
(c) $\sum_{j=1}^{\infty} \|A_j\|(1+\eta)^j < \infty$ *for some* $\eta > 0$, *where* $\|\cdot\|$ *is the matrix norm defined by* $\|X\| = \{\mathrm{tr}(X^{\mathrm{T}}X)\}^{1/2}$.
(d) $k \to \infty$ *as* $T \to \infty$ *and* $k = o(T^{1/7})$.[5]

Except for the assumption of finite moments up to the eighth order, assumptions (a)–(d) are from Paparoditis (1996). Throughout this article, we establish the validity of the bootstrap by showing that, conditional on the sample, the bootstrap analog converges weakly to the limiting distribution of the original statistic *in probability*. The in-probability bootstrap asymptotics used in this article are described in detail in Giné and Zinn (1990) (also see Freedman, 1981; Kreiss, 1988; Paparoditis and Streitberg, 1992; Paparoditis, 1996).

[4] We use *vec* to denote the column stacking operator and *vech* to denote the operator that stacks only the elements on and below the diagonal.
[5] Assumption (c) ensures that we do not require a lower bound on k.

THEOREM 1. *Under Assumption* 1, *for all k*,

$$(1)\qquad \begin{pmatrix} (T-k)^{1/2} l(k)^{\mathrm{T}}(\hat{a}(k)-a(k)) \\ (T-k)^{1/2}\mathrm{vech}(\hat{\Sigma}_k-\Sigma) \end{pmatrix} \xrightarrow{d} N(0_{\{1+r(r+1)/2\}\times 1}, \Omega)$$

where Ω *is defined in the Appendix*, *and* $\{l(k)\}_{k\in N}$ *is a sequence of* $kr^2\times 1$ *vectors satisfying* $0 < M_1 \le \|l(k)\| \le M_2 < \infty$, *and*, *conditional on the sample*,

$$(2)\qquad \begin{pmatrix} (T-k)^{1/2} l(k)^{\mathrm{T}}(\hat{a}^*(k)-\hat{a}(k)) \\ (T-k)^{1/2}\mathrm{vech}(\hat{\Sigma}^*_k-\hat{\Sigma}_k) \end{pmatrix} \xrightarrow{d} N(0_{\{1+r(r+1)/2\}\times 1}, \Omega)$$

in probability.

Next, we consider several applications of Theorem 1. The first application is a sequence of smooth functions of $A(k)$ and Σ. Consider a sequence of functions $\{g_k\}_{k=1}^{\infty}$ such that $g_k:\Re^{kr^2+r(r+1)/2}\to\Re$. Let $Dg_k(x,y)$ denote the derivative of g_k with respect to $(x^{\mathrm{T}},y^{\mathrm{T}})^{\mathrm{T}}$, where $x\in\Re^{kr^2}$ and $y\in\Re^{r(r+1)/2}$.

ASSUMPTION 2.

(a) $0 < M_1 \le \|Dg_k(x,y)\| \le M_2 < \infty$ *for all* $x\in\Re^{kr^2}$ *and* $y\in\Re^{r(r+1)/2}$.
(b) Dg_k *satisfies a Lipschitz condition: There is* $M > 0$ *such that*

$$\|Dg_k(x',y') - Dg_k(x,y)\| \le M\|(x'^{\mathrm{T}},y'^{\mathrm{T}})^{\mathrm{T}} - (x^{\mathrm{T}},y^{\mathrm{T}})^{\mathrm{T}}\|$$

for all $x,x'\in\Re^{kr^2}$ *and* $y,y'\in\Re^{r(r+1)/2}$.

COROLLARY 1. *Under Assumptions* 1 *and* 2,

$$(3)\qquad (T-k)^{1/2}\{g_k(\mathrm{vec}(\hat{A}(k)),\mathrm{vech}(\hat{\Sigma}_k)) - g_k(\mathrm{vec}(A(k)),\mathrm{vech}(\Sigma))\} \xrightarrow{d} N(0,\Omega_g)$$

where Ω_g *is defined in Appendix* A.1, *and*, *conditional on the sample*,

$$(4)\qquad (T-k)^{1/2}\{g_k(\mathrm{vec}(\hat{A}^*(k)),\mathrm{vech}(\hat{\Sigma}^*_k)) - g_k(\mathrm{vec}(\hat{A}(k)),\mathrm{vech}(\hat{\Sigma}_k))\} \xrightarrow{d} N(0,\Omega_g)$$

in probability.

An example of the class of statistics covered by Corollary 1 is the half-life of a unit shock in an AR(p) process. Corollary 1 is a general result, but in many cases it is easier to prove the asymptotic validity of the bootstrap directly rather than verifying Assumption 2. The following three corollaries to Theorem 1 are of particular interest in applied work. The first two examples are orthogonalized impulse responses and variance decompositions. Let

$$\hat{B}_{j,k} = \sum_{i=1}^{j} \hat{B}_{j-i,k}\hat{A}_{i,k}$$

for $0 < j \le k$ with $\hat{B}_{0,k} = I_r$, and let $\hat{P}_k$ be the lower triangular Cholesky decomposition of $\hat{\Sigma}_k$ such that $\hat{P}_k\hat{P}_k^{\mathrm{T}} = \hat{\Sigma}_k$. Further, let $\hat{B}^*(h,k) = (\hat{B}^*_{1,k},\hat{B}^*_{2,k},\ldots,\hat{B}^*_{h,k})$ and $\hat{P}^*_k$ denote the corresponding bootstrap moving-average parameters and the Cholesky decomposition of $\hat{\Sigma}^*_k$, respectively, where h denotes the horizon. Define the orthogonalized impulse responses by $\Theta_i = B_i P$ and denote their estimates and bootstrap estimates by $\hat{\Theta}_{i,k}$ and $\hat{\Theta}^*_{i,k}$, respectively.

COROLLARY 2. *Under Assumption* 1, *for h fixed*,

$$\sqrt{T}\,\text{vec}([\hat{\Theta}_{0,k}, \hat{\Theta}_{1,k}, \ldots, \hat{\Theta}_{h,k}] - [\Theta_0, \Theta_1, \ldots, \Theta_h]) \xrightarrow{d} N(0_{(h+1)r^2\times 1}, \Omega_\theta) \tag{5}$$

where Ω_θ *is defined in Appendix* A.1, *and*, *conditional on the sample*,

$$\sqrt{T}\,\text{vec}\Big([\hat{\Theta}^*_{0,k}, \hat{\Theta}^*_{1,k}, \ldots, \hat{\Theta}^*_{h,k}] - [\hat{\Theta}_{0,k}, \hat{\Theta}_{1,k}, \ldots, \hat{\Theta}_{h,k}]\Big) \xrightarrow{d} N(0_{(h+1)r^2\times 1}, \Omega_\theta) \tag{6}$$

in probability.

We follow Lütkepohl and Poskitt (1991) in defining forecast error variance components

$$\omega_{mn,h} = \sum_{j=0}^{h-1} \frac{e_m^{\mathrm{T}} \Theta_j e_n}{\text{PMSE}_{m,h}}, \quad h = 1, 2, \ldots$$

where e_m is the mth column of I_r and $\text{PMSE}_{m,h}$ is the mth diagonal element of the forecast error covariance matrix $\sum_{j=0}^{h-1} \Theta_j \Theta_j^{\mathrm{T}}$.

COROLLARY 3. *Under Assumption* 1, *for h fixed*,

$$\sqrt{T}(\hat{\omega}_{mn,h}(k) - \omega_{mn,h}) \xrightarrow{d} N(0, \sigma^2_{mn,h}) \tag{7}$$

where $\sigma^2_{mn,h}$ *is defined in Appendix* A.1, *and*, *conditional on the sample*,

$$\sqrt{T}(\hat{\omega}^*_{mn,h}(k) - \hat{\omega}_{mn,h}(k)) \xrightarrow{d} N(0, \sigma^2_{mn,h}) \tag{8}$$

in probability.

The third example that is of special interest is the measure of predictability proposed by Diebold and Kilian (2000). We prove the asymptotic validity of bootstrapping this statistic under Assumptions 1 and 2 for the special case of a quadratic loss function. The predictability measure $P(m, n)$ for a given series $i = 1, \ldots, r$ is defined as

$$P^i(m, n) = 1 - \frac{\text{PMSE}^i(m)}{\text{PMSE}^i(n)}, \quad m \le n < \infty$$

where $\text{PMSE}^i(h)$ denotes the ith diagonal element of the PMSE matrix at horizon h. Under our assumptions, $P^i(m, n) \in [0, 1]$ for all $m \le n$ with larger values indicating higher predictability.

COROLLARY 4. *Under Assumption* 1, *for n fixed*, *we have*

$$\sqrt{T}(\hat{P}_k(m, n) - P(m, n)) \xrightarrow{d} N(0, \Sigma_p(m, n)) \tag{9}$$

where $\Sigma_p(m, n)$ *is defined in Appendix* A.1, *and*, *conditional on the sample*,

$$\sqrt{T}(\hat{P}_k^*(m,n) - \hat{P}_k(m,n)) \xrightarrow{d} N(0, \Sigma_p(m,n)) \tag{10}$$

in probability.

The next result establishes the asymptotic validity of bootstrapping the scalar measure of predictability proposed by Granger and Newbold (1986, p. 310). This measure of predictability is patterned after the familiar R^2 for univariate linear regressions:

$$P_{\mathrm{GN}} = 1 - \frac{\mathrm{var}(e_{t+1})}{\mathrm{var}(y_{t+1})}$$

where $e_{t+1} = y_{t+1} - \hat{y}_{t+1|t}$ denotes the one-step ahead forecast error. By construction, $P_{\mathrm{GN}} \in [0,1]$ with larger values indicating higher predictability.

THEOREM 2. *Under Assumption* 1 *of Theorem* 1 *with* (*d*) *replaced by* (d') $k \to \infty$ *as* $T \to \infty$ *and* $k = o(T^{1/8})$,*c*

$$\sqrt{T}(\hat{P}_{\mathrm{GN}}(k) - P_{\mathrm{GN}}) \xrightarrow{d} N(0, \Sigma_{\mathrm{GN}}) \tag{11}$$

where Σ_{GN} *is defined in Appendix* A.1, *and*, *conditional on the sample*,

$$\sqrt{T}(\hat{P}_{\mathrm{GN}}^*(k) - \hat{P}_{\mathrm{GN}}(k)) \xrightarrow{d} N(0, \Sigma_{\mathrm{GN}}) \tag{12}$$

in probability.

4. MONTE CARLO EVIDENCE

We conduct two Monte Carlo simulation studies to analyze the accuracy of the proposed bootstrap method. Details of the DGPs can be found in Appendix A.2. The first study examines a univariate measure of predictability, and the second application focuses on orthogonalized impulse responses in a vector autoregressive model. Measures of predictability play an important role in policy analysis as well as in the evaluation of macroeconomic models (see Barsky, 1987; Ball and Cecchetti, 1990; Rotemberg and Woodford, 1996; Diebold and Kilian, 2000). Here, we focus on the scalar measure of predictability, P_{GN}, proposed by Granger and Newbold (1986) and described in Section 3. Theorem 2 provides the basis for constructing bootstrap confidence intervals for P_{GN}. The bootstrap approach has several advantages over the delta method in this context. Not only are closed-form solutions for the asymptotic standard errors not available for P_{GN}, but the application of the delta method itself is questionable because the P_{GN} statistic is bounded between 0 and 1. It is well known that the delta-method interval is not range respecting and may produce confidence intervals that are logically invalid. In contrast, the bootstrap percentile interval by construction preserves these constraints (see Efron and Tibshirani, 1993).

Here, we present evidence that the bootstrap may be used to provide reliable measures of sampling uncertainty for the P_{GN} statistic. Our DGP is based on an ARMA(2,4) model for U.S. post-war inflation (chosen by the Akaike information criterion). The data are for monthly residential consumer prices excluding shelter

(DRI code: PRXHS) from 1960.1 to 1998.10. The parameters of the estimated model are frozen and treated as the parameters of the population model that underlies the simulation study. We then proceed by analyzing the coverage probabilities of the nominal 90 percent bootstrap percentile confidence intervals based on the approximating autoregressive model. We implement the bootstrap using the bias corrections described in Kilian (1998). These small-sample bias corrections have no effect on the asymptotic validity of the procedure but greatly enhance the small-sample performance (see Diebold and Kilian (2000) for further simulation evidence).

Table 1 shows effective coverage probabilities for the nominal 90 percent bootstrap confidence interval based on 1000 Monte Carlo trials each. The sample size in the Monte Carlo study is $T = 240$. Asymptotic theory provides no guidance as to the choice of the lag order k of the approximating model for a given sample size. We therefore display results for a number of alternative lag orders k.

We find that good approximations may require fairly high lag orders in practice. This point has also been illustrated by Berkowitz et al. (1999) for univariate ARMA(p, q) models and by Braun and Mittnik (1993) for VARMA(p, q) models. Table 1 shows that for $k = 3$, 4, and 5 the sieve approximation does not work well. Even for $k = 6$ and 7, the sieve approximation is inadequate, although the coverage accuracy steadily improves, as we add more lags. Table 1 shows that for $k = 8$, the reliability of the bootstrap method is excellent. For larger approximating models, the bootstrap coverage rates are conservative. This simulation study illustrates that valid bootstrap inference is possible based on the theoretical results in this article, provided care is taken to include a sufficient number of lags.

The second simulation study is based on a direct application of Corollary 2. The DGP is based on the trivariate example used by Braun and Mittnik (1993). Models of this dimension are not uncommon in applied work. Models of similar or smaller size have been analyzed, for example, by Rotemberg and Woodford (1996, 1997), Cogley and Nason (1995), Canova and Marrinan (1998), Galí (1999), and Leeper and Sims (1994). We follow Braun and Mittnik in estimating a VARMA(1,1) model for aggregate, quarterly postwar U.S. time series data on investment expenditures, the price of investment, and the 90-day commercial paper rate. The DRI codes are GIFQF, GDIF, and FYCP90. The sample period is 1971.I–1998.II. The investment and deflator series are specified in log differences. This amounts to assuming that the variables are $I(1)$ but not cointegrated. The interest rate is specified in levels. The estimated parameters are frozen and treated as the population DGP in the Monte Carlo study. Based on this DGP we compare the coverage rates and average length of the bootstrap percentile and delta-method intervals for the orthogonalized

TABLE 1

COVERAGE RATE OF NOMINAL 90 PERCENT CONFIDENCE INTERVAL FOR GRANGER–NEWBOLD PREDICTABILITY MEASURE OF INFLATION RATE (RESULTS FOR ALTERNATIVE AR(k) APPROXIMATING MODELS GIVEN $T = 240$)

k	3	4	5	6	7	8	9	10
Coverage	56.4	65.0	75.5	82.2	86.4	90.3	92.5	93.7

NOTES: Univariate ARMA(2,4) data-generating process for U.S. postwar monthly inflation rate. For details see text and Appendix A.2.

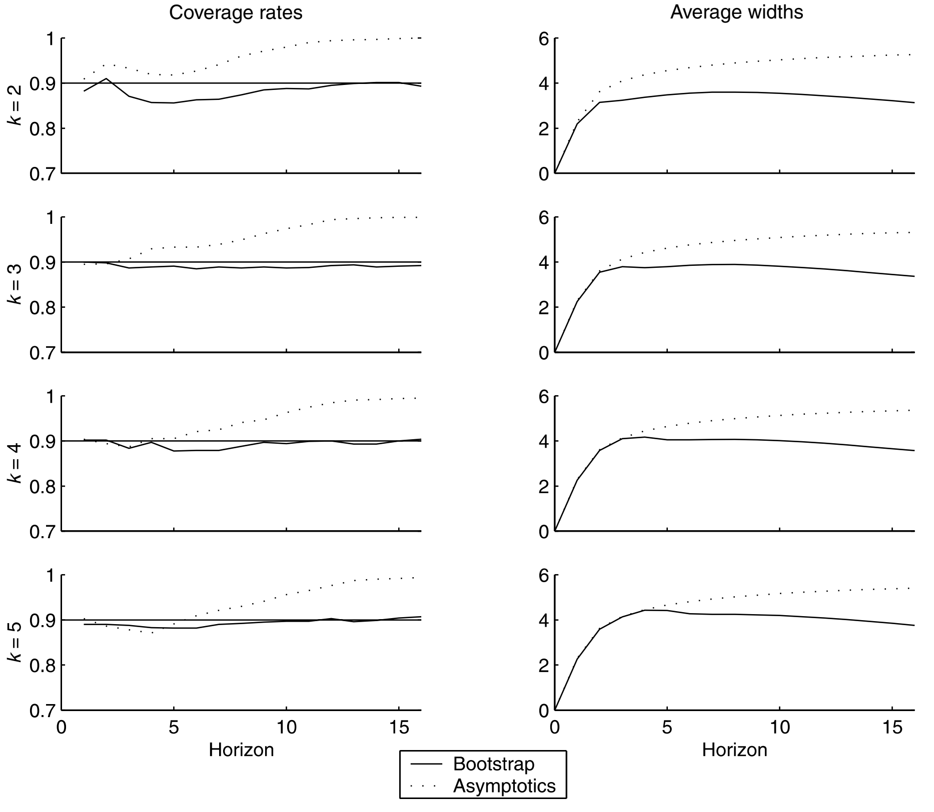

FIGURE 1

POINTWISE NOMINAL 90 PERCENT CONFIDENCE INTERVALS FOR RESPONSE OF INVESTMENT TO AN ORTHOGONALIZED INTEREST RATE INNOVATION (RESULTS FOR ALTERNATIVE VAR(k) APPROXIMATING MODELS GIVEN $T = 200$)

NOTES: Trivariate varma(1,1) data-generating process for quarterly investment expenditures, investment deflator, and discount rate. For details see text and Appendix A.2.

impulse responses. The delta-method intervals are calculated based on the closed-form solutions proposed by Lütkepohl and Poskitt (1991). We again implement the percentile interval using the bias corrections described by Kilian (1998).[6] All results are based on 1000 Monte Carlo trials.

For illustrative purposes, we focus on the response of investment to an orthogonalized innovation in the discount rate. The first column of Figure 1 shows the coverage rates for the pointwise impulse response confidence intervals based on alternative VAR(k) approximating models given $T = 200$.

The coverage rates of the bootstrap intervals are remarkably accurate overall except for $k = 2$. There is little difference between the results for $k = 3$, 4, and 5.

[6] We do not explore percentile-t intervals. See Kilian (1999) for a detailed analysis of the tradeoffs between alternative bootstrap confidence intervals for impulse responses and other nonlinear functions of slope parameters and innovation variances.

Unlike in Kilian's (1998) analysis of the finite-lag order VAR model, the coverage rates of the delta-method interval do not decline as the horizon increases. However, the percentile interval tends to be more accurate than the delta-method interval at longer horizons, nevertheless. The reason is that the coverage rates of the delta-method intervals tend to approach 100 percent coverage at long horizons, whereas those of the bootstrap interval remain closer to the nominal coverage probability of 90 percent even after 16 quarters.

Broadly similar results hold for most other impulse responses not shown here. The only exception to the good performance of the bootstrap is that the three own-impulse responses tend to have somewhat lower coverage accuracy on impact. This pattern is suggestive of poor bootstrap approximations of at least some elements of the innovation covariance matrix. However, the coverage accuracy tends to improve drastically after a few quarters in all cases.

The second column of Figure 1 reveals that the conservative coverage rates of the delta method come at a price. The delta-method interval tends to be much wider on average at longer horizons than the bootstrap interval. Although we know that, as T approaches infinity, the interval endpoints of Lütkepohl and Poskitt's delta-method interval and of the bootstrap percentile interval will coincide, for $h > k$ and fixed T the intervals can be quite different. We also note that, for $h > k$ and fixed T, the conventional asymptotic theory for bootstrapping finite-lag order models appears to provide a better approximation than the bootstrap asymptotic theory for VAR(∞) models. We conclude that the bootstrap approximation to the distribution of impulse response estimates in VAR(∞) processes provides a useful alternative to the use of the asymptotic normal approximation and may have important practical advantages both in terms of accuracy and in terms of width.

5. APPLICATION: EVALUATING THE FIT OF MACROECONOMIC MODELS

It is common in applied work to compare the spectra, impulse responses, autocorrelations, measures of predictability, and other statistics implied by a theoretical macroeconomic model to the corresponding statistics calculated from reduced-form VAR representations (e.g., Cogley and Nason, 1995; King and Watson, 1996; Rotemberg and Woodford, 1996, 1997). Nonparametric approximations to the reduced form may be used to construct confidence intervals for the statistic of interest that do not depend on a correctly specified theoretical model (e.g., Diebold et al., 1998; Schmitt-Grohé, 1998; Diebold and Kilian, 2000). A model is said not to conform to the data if the statistic generated by the theoretical model is not contained within the confidence bands estimated from the reduced form of the data.

For example, Canova and Marrinan (1998, p. 139) suggest identifying pseudo-structural shocks from the actual data using arbitrary restrictions and comparing the resulting impulse responses of the model with those obtained from data simulated from different specifications of the macroeconomic model where shocks are identified using the same arbitrary restrictions. In other words, the impulse responses are used as a "window" through which we measure the quality of the model approximation of the data. This approach avoids some of the pitfalls in the econometric evaluation of macroeconomic models recently discussed by Cooley and

Dwyer (1998). It is important, however, that the theoretical model has enough shocks to avoid singularities of the covariance matrix of the data. For example, it is not valid to fit an approximating VAR model to data generated from theoretical models with just one exogenous driving process. Alternative "windows" that do not depend on the number of shocks in the theoretical model economy have been suggested by Watson (1993), Rotemberg and Woodford (1996), Diebold and Kilian (2000), and Diebold et al. (1998), among others. Here, we focus on an application based on the predictability measure proposed by Diebold and Kilian (2000). The asymptotic validity of bootstrapping this statistic has been established in Corollary 4.

For expository purposes, consider the following cash-in-advance model economy: The representative household chooses labor input, h_t, and next period's capital stock, k_{t+1}, to maximize expected lifetime utility,

$$E_0 \sum_{t=0}^{\infty} \beta^t U(c_t, h_t)$$

subject to a sequence of cash-in-advance constraints, wealth constraints, and time endowments for each period t (see Cooley and Hansen, 1995, p. 195, for details). The utility function of the representative household is $U(c_t, h_t) = (c_t^{\psi} + (1-h_t)^{1-\psi})^{1-\rho}/(1-\rho) - 1$ and $0 < \beta < 1$. Firms choose capital and labor services each period to maximize profits. Since firms are competitive, they treat all prices as given. Production is Cobb–Douglas of the form $x_t k_t^{\theta} h_t^{1-\theta}$, where $0 < \theta < 1$ is the capital share. x_t is an exogenous productivity factor that follows a log-linear process $x_t = (1-\gamma) + \gamma x_{t-1} + e_t^x$ with $e_t^x \sim \text{NID}(0, \sigma_x^2)$. Capital is accumulated according to $k_{t+1} = (1-\delta)k_t + i_t$, with depreciation rate $0 \leq \delta \leq 1$. Money supply growth is exogenous and follows the log-linear process $g_t = (1-\mu) + \mu g_{t-1} + e_t^g$ with $e_t^g \sim \text{NID}(0, \sigma_g^2)$. Agents treat k_0, x_0, and g_0 as given.

We follow the existing literature in parameterizing the model as follows: $\psi = 1/3$, $\beta = 0.98$, $\rho = 2$, $\delta = 0.19$, $\theta = 0.4$, $\gamma = 0.95$, $\mu = 0.586$, $\sigma_g = 0.0097$, $\sigma_x = 0.007$. The money supply process is calibrated to U.S. M1 money growth for 1959.I–1998.III using the standard procedure described by Cooley and Hansen (1995). The choice of the other parameters is conventional. After solving for the approximate linear decision rules for h_t and k_{t+1} in terms of the current period states k_t, x_t, and g_t, the predictability of the endogenous model variables can be calculated numerically by fitting an approximating autoregressive model to 5000 observations generated by the model economy. We compute the predictability of the actual U.S. data for 1959.IV–1998.III based on univariate sieve approximations with $k = 8$ after linearly detrending the data. Very similar results are obtained with $k = 4$ and 6. The computation of the predictability measure of the model data is not subject to sampling error, while the predictability estimates for the U.S. data are; thus, we compute confidence intervals only for the latter. As noted in Section 2, it is essential to allow for possibly infinite-lag orders in evaluating macroeconomic models. This application illustrates the practical relevance of our bootstrap theory.

Figure 2a shows the predictability measures implied by the macroeconomic model together with nominal 90 percent bootstrap confidence intervals based on the sieve approximation of the DGP. We set $n = 40$ and plot $P(m, n)$ for $m \leq 20$.

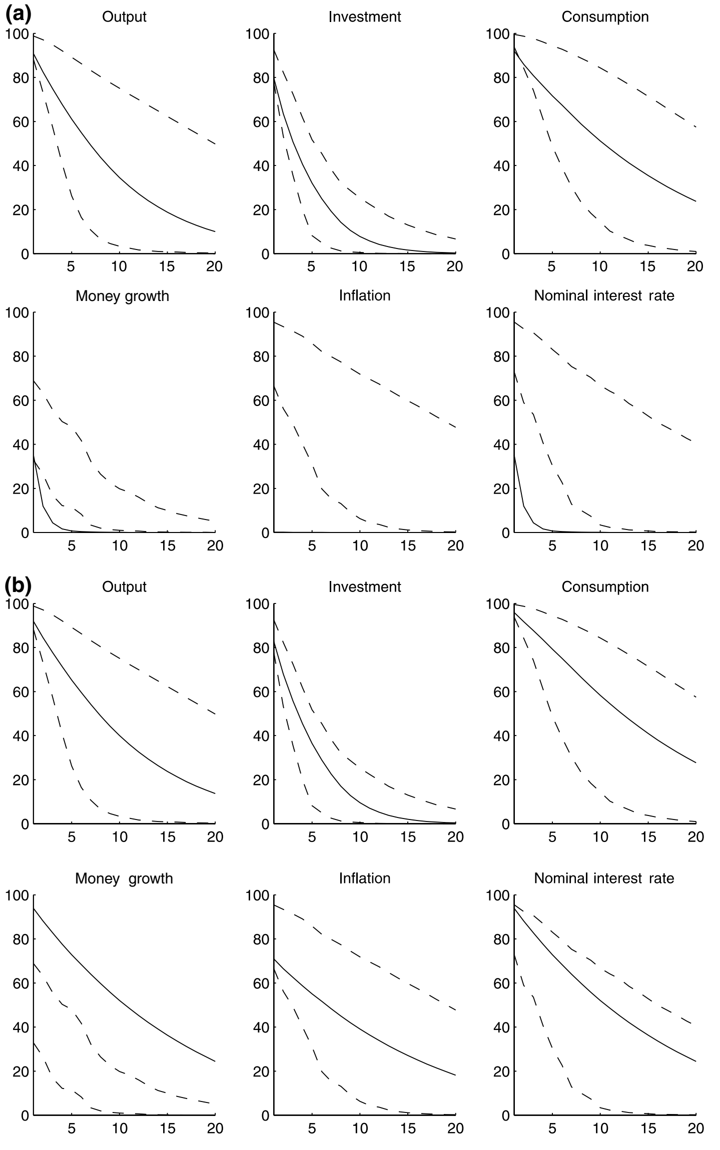

FIGURE 2

(a) PERFORMANCE OF CASH-IN-ADVANCE MODEL IN TERMS OF PREDICTABILITY: $\mu = 0.586$.
(b) PERFORMANCE OF CASH-IN-ADVANCE MODEL IN TERMS OF PREDICTABILITY: $\mu = 0.97$.

NOTES: For a detailed description of the model and its parameterization see Section 5. The dashed lines are the pointwise 90 percent confidence intervals for the U.S. estimates. All data are from DRI. Output is computed as real GDP minus real government purchases. Investment is real gross fixed private sector investment. Consumption refers to real private consumption of nondurables and services. Money growth is growth of M1. Inflation is based on the implicit deflator of private sector output. The interest rate is the three-month T-bill rate. The sample period is 1959.I–1998.III. All data are linearly detrended.

How consistent are the data generated by this model with the U.S. data? We find that the predictability of private sector output, fixed investment, and consumption of nondurables and services in the model economy is generally consistent with the interval estimates based on the reduced form. The predictability of model output and investment is within the bootstrap confidence intervals at all horizons. Consumption is slightly less predictable in the model than in the data at short horizons, but its predictability in the model is generally within the bootstrap confidence intervals. In sharp contrast, both the inflation rate and the nominal interest rate are far less predictable in the model than in the data at all horizons, indicating that the model is completely unable to explain the predictability of nominal variables. Notably, inflation is virtually unpredictable at all horizons. Its predictability is so low that the plotted line in Figure 2a cannot be distinguished from the horizontal axis. Given that the model-based predictability is completely outside the bootstrap confidence intervals for these series, we can reject the hypothesis that the model is consistent with the data.

Unlike summary measures of goodness of fit, our approach is constructive in that it helps to pinpoint the weaknesses of the model. We illustrate this point for the model at hand. One of the problems with the baseline model in Figure 2a is the standard procedure for calibrating exogenous money supply growth by fitting an AR(1) model to U.S. M1 data (see Cooley and Hansen, 1995). An AR(1) model is unlikely to be an adequate representation if M1 growth data possibly follow an infinite-order process. This inadequacy of the AR(1) model is reflected in the large deviations between the model measure and the data measure of money growth predictability in Figure 2a.

A second and more fundamental objection to the standard approach of calibrating the exogenous money supply process is that the U.S. M1 data used to calibrate the exogenous money supply process in the model clearly contain an important endogenous component. This fact suggests that we treat μ as a free parameter instead and calibrate this parameter to produce predictability profiles consistent with the interval estimates. Figure 2b shows that for $\mu = 0.97$ the performance of the model improves dramatically.

Now the predictability of both the inflation rate and the nominal interest rate is consistent with the data. There is little change in the predictability of output and investment, but model consumption now is contained within the bootstrap intervals at all horizons. This exercise suggests that a cash-in-advance model is capable in principle of accounting for several of the stylized facts of the U.S. economy, provided that the exogenous money supply growth process is highly persistent and much more predictable than U.S. M1 data and other monetary aggregates. Thus, a key question to be addressed is the plausibility of such highly persistent money supply growth processes. Insights of the type developed here are useful for understanding the propagation mechanism of the theoretical model and for determining which features of the model are essential and which are not.

6. CONCLUDING REMARKS

We have established the asymptotic validity of the residual-based autoregressive bootstrap method for smooth functions of slope parameters and innovation variances under the assumption that a sequence of finite-lag order VAR models is fitted to data generated by a VAR process of possibly infinite order. Our theoretical results cover a wide range of statistics currently in use in macroeconometrics. The existence of finite-lag order VAR models is highly implausible in practice and often inconsistent with the assumptions of the macroeconomic model underlying the empirical analysis. The proposed bootstrap approach for VAR(∞) models provides an alternative to the use of the asymptotic normal approximation and can be used even in the absence of closed-form solutions for the variance of the estimator. Our results are of interest both for structural VAR modeling and for the econometric evaluation of dynamic economic models. In a Monte Carlo study, the bootstrap approach compared favorably with results based on the asymptotic normal approximation for orthogonalized impulse responses in VAR(∞) models. We also illustrated the practical usefulness of our theoretical results for the evaluation of dynamic general equilibrium models.

APPENDIX: PROOFS AND SIMULATION DESIGN

A.1. *Proofs*

Notation

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \text{plim}_{T\to\infty}\{l(k)^{\mathrm{T}}(\hat{\Gamma}_k^{-1}\otimes\hat{\Sigma}_k)l(k)\} & \lim_{T\to\infty} E[l(k)^{\mathrm{T}}\operatorname{vec}(\varepsilon_t Y_{t-1,k}^{\mathrm{T}}\Gamma_k^{-1})\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}})^{\mathrm{T}}] \\ \lim_{T\to\infty} E[\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}})\operatorname{vec}(\Gamma_k^{-1}Y_{t-1,k}\varepsilon_t^{\mathrm{T}})l(k)] & E[\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}}-\Sigma)\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}}-\Sigma)^{\mathrm{T}}] \end{pmatrix}$$

$$\Omega_g = \lim_{T\to\infty} Dg_k^{\mathrm{T}} \begin{pmatrix} \Gamma_k^{-1}\otimes\Sigma & E[\operatorname{vec}(\varepsilon_t Y_{t-1,k}^{\mathrm{T}}\Gamma_k^{-1})\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}})^{\mathrm{T}}] \\ E[\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}})\operatorname{vec}(\varepsilon_t Y_{t-1,k}^{\mathrm{T}}\Gamma_k^{-1})^{\mathrm{T}}] & E[\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}}-\Sigma)\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}}-\Sigma)^{\mathrm{T}}] \end{pmatrix} Dg_k$$

$$\begin{aligned} \Omega_\theta = {} & (I_{h+1}\otimes P^{\mathrm{T}}\otimes I_r)\lim_{T\to\infty}\{\Psi_{h,k}(\Gamma_k^{-1}\otimes\Sigma)\Psi_{h,k}^{\mathrm{T}}\}(I_{h+1}\otimes P\otimes I_r) \\ & + (I_{h+1}\otimes P^{\mathrm{T}}\otimes I_r)\lim_{T\to\infty}\Psi_{h,k}E]\operatorname{vec}(\varepsilon_t Y_{t-1,k}^{\mathrm{T}}\Gamma_k^{-1})\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}})^{\mathrm{T}}]H^{\mathrm{T}}F^{\mathrm{T}}(I_{r(h+1)}\otimes[I_r, B_1,\ldots,B_h])^{\mathrm{T}} \\ & + (I_{r(h+1)}\otimes[I_r, B_1,\ldots,B_h])FH\lim_{T\to\infty}E[\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}})\operatorname{vec}(\varepsilon_t Y_{t-1,k}^{\mathrm{T}}\Gamma_k^{-1})^{\mathrm{T}}\Psi_{h,k}^{\mathrm{T}}(I_{h+1}\otimes P\otimes I_r) \\ & + (I_{r(h+1)}\otimes[I_r, B_1,\ldots,B_h])FHE[\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}}-\Sigma)\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}}-\Sigma)^{\mathrm{T}}]H^{\mathrm{T}}F^{\mathrm{T}}(I_{r(h+1)} \\ & \otimes[I_r, B_1,\ldots,B_h])^{\mathrm{T}} \end{aligned}$$

$$\Psi_{h,k} = \begin{pmatrix} J_k \otimes B_j \\ \sum_{j=0}^{1} J_k \Pi_k^{\mathrm{T}1-j} \otimes B_j \\ \vdots \\ \sum_{j=0}^{h-1} J_k \Pi_k^{\mathrm{T}h-j-1} \otimes B_j \end{pmatrix}$$

$$J_k = (I_r, 0_{r\times(k-1)r})$$

$$\Pi_k = \begin{pmatrix} A_1 & A_2 & \cdots & \cdots & A_k \\ I_r & 0_{r\times r} & \cdots & \cdot & 0_{r\times r} \\ 0_{r\times r} & I_r & \cdots & \cdot & 0_{r\times r} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{r\times r} & 0_{r\times r} & \cdots & I_r & 0_{r\times r} \end{pmatrix}$$

$$\Gamma_{1,k} = E[Y_{t,k} y_{t+1}], \quad \Gamma_k = E[Y_{t,k} Y_{t,k}^{\mathrm{T}}]$$

$$F = \begin{pmatrix} G^{\mathrm{T}} & 0_{r^2\times(h+1)r} & G^{\mathrm{T}} & 0_{r^2\times(h+1)r} & \cdots & 0_{r^2\times(h+1)r} & G^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}}$$

$$G = \begin{pmatrix} \begin{matrix} I_r \\ 0_{r(r-1)\times r} \end{matrix} & 0_{r^2\times hr} & \begin{matrix} 0_{r\times r} \\ I_r \\ 0_{r(r-2)\times r} \end{matrix} & 0_{r^2\times hr} & \cdots & \begin{matrix} 0_{(r-1)\times r} \\ I_r \end{matrix} \end{pmatrix}^{\mathrm{T}}$$

and H is defined as in Lütkepohl and Poskitt (1991, Equation 5):

$$H = L_r^{\mathrm{T}}[L_r\{(I_r \otimes P)K_{rr} + (P \otimes I_r)\}L_r^{\mathrm{T}}]^{-1}$$

L_r is the $(r(r+1)/2 \times r^2)$ elimination matrix and K_{rr} is the $(r^2 \times r^2)$ commutation matrix.

$$\sigma^2_{mn,h} = [g_{mn,h}(0), \ldots, g_{mn,h}(h-1)]\Omega_\theta[g_{mn,h}(0), \ldots, g_{mn,h}(h-1)]^{\mathrm{T}}$$

$$\begin{aligned} g_{mn,h}(j) &= \frac{\partial \omega_{mn,h}}{\partial\, \mathrm{vec}(\Theta_j)^{\mathrm{T}}} \\ &= 2\,\frac{\left[(e_n^{\mathrm{T}} \otimes e_m^{\mathrm{T}})(e_m^{\mathrm{T}}\Theta_j e_n)\mathrm{PMSE}_{m,h} - (e_m^{\mathrm{T}}\Theta_j \otimes e_m^{\mathrm{T}})\sum_{l=0}^{h-1}(e_m^{\mathrm{T}}\Theta_l e_n)^2\right]}{\mathrm{PMSE}^2_{m,h}} \end{aligned}$$

$$\begin{aligned} \Sigma_P(m,n) &= g_P(m,n)^{\mathrm{T}} \begin{pmatrix} \Omega_{P11} & \Omega_{P12} \\ \Omega_{P21} & \Omega_{22} \end{pmatrix} g_P(m,n) \\ g_P(m,n) &= \begin{pmatrix} \frac{\partial P}{\partial(\mathrm{vec}(B_1)^{\mathrm{T}},\, \mathrm{vec}(B_2)^{\mathrm{T}}, \ldots,\, \mathrm{vec}(B_n)^{\mathrm{T}})^{\mathrm{T}}} \\ \frac{\partial P}{\partial\, \mathrm{vech}(\Sigma)^{\mathrm{T}}} \end{pmatrix} \\ \Omega_{P11} &= \lim_{T\to\infty} l(n,k)^{\mathrm{T}}\Psi_{h,k}(\Gamma_k^{-1} \otimes \Sigma)\Psi_{h,k}^{\mathrm{T}} l(h,k) \\ \Omega_{P12} &= \lim_{T\to\infty} l(n,k)^{\mathrm{T}}\Psi_{h,k}E[\mathrm{vec}(\varepsilon_t Y_{t-1}^{\mathrm{T}}\Gamma_k^{-1})\,\mathrm{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}})^{\mathrm{T}}]H^{\mathrm{T}} \\ \Omega_{P21} &= \Omega_{P12}^{\mathrm{T}} \end{aligned}$$

$$\begin{aligned} \Sigma_{\mathrm{GN}} &= \lim_{T\to\infty} g_{\mathrm{GN}}(k)^{\mathrm{T}}\Omega(k)g_{\mathrm{GN}}(k) \\ g_{\mathrm{GN}}(k) &= \begin{pmatrix} \frac{2b_0}{\left(\sum_{j=0}^{k} b_j^2\right)^2} & \frac{2b_1}{\left(\sum_{j=0}^{k} b_j^2\right)^2} & \frac{2b_2}{\left(\sum_{j=0}^{k} b_j^2\right)^2} & \cdots & \frac{2b_k}{\left(\sum_{j=0}^{k} b_j^2\right)^2} \end{pmatrix}^{\mathrm{T}} \end{aligned}$$

Proof of Theorem 1. Let

$$\Delta_k = (T-k)^{-1}\sum_{t=k}^{T-1}\varepsilon_{t+1}Y_{t,k}^{\mathrm{T}}$$

$$\Delta_k^* = (T-k)^{-1}\sum_{t=k}^{T-1}\varepsilon_{t+1}^*Y_{t,k}^{*\mathrm{T}}$$

$$\Gamma_k^* = (\Gamma^*(m-n))_{m,n=1,2,\dots,k}$$

$$\Gamma^*(m-n) = E^*(y_{t+n}^*y_{t+m}^{*\mathrm{T}})$$

and $\|A\|_1 = \sup_{x\neq 0}\{\|Ax\|/\|x\|\}$. Paparoditis (1996) shows that

$$\|\Delta_k^*\| = O_p(k^{1/2}T^{-1/2}) \tag{A.1}$$

$$\|\hat{\Gamma}_k^{*-1} - \Gamma_k^{*-1}\|_1 = O_p(k^3T^{-1/2}) \tag{A.2}$$

$$\|\Gamma_k^{*-1} - \Gamma_k^{-1}\|_1 = O_p(k^3T^{-1/2}) \tag{A.3}$$

in the proofs of his Theorems 3.1 and 3.2 (p. 289, ll.10–11; p. 290, l.14; and p. 288, l.20). We have

$$\begin{pmatrix}(T-k)^{1/2}l(k)^{\mathrm{T}}(\hat{a}(k)^* - \hat{a}(k))\\ (T-k)^{1/2}\mathrm{vech}(\hat{\Sigma}_k^* - \hat{\Sigma}_k)\end{pmatrix}$$

$$= \begin{pmatrix}(T-k)^{1/2}l(k)^{\mathrm{T}}\mathrm{vec}\left(\Delta_k^*\Gamma_k^{*-1}\right)\\ (T-k)^{1/2}\mathrm{vech}\left\{(T-k)^{-1}\sum\limits_{t=k+1}^{T}\varepsilon_t^*\varepsilon_t^{*\mathrm{T}} - (T-k)^{-1}\sum\limits_{s=k+1}^{T-1}\hat{\varepsilon}_{k,s}\hat{\varepsilon}_{k,s}^{\mathrm{T}}\right\}\end{pmatrix} + o_p(1)$$

$$= \begin{pmatrix}(T-k)^{1/2}l(k)^{\mathrm{T}}\mathrm{vec}(\Delta_k^*\Gamma_k^{-1})\\ (T-k)^{1/2}\mathrm{vech}\left\{(T-k)^{-1}\sum\limits_{t=k+1}^{T}\left(\varepsilon_t^*\varepsilon_t^{*\mathrm{T}} - (T-k)^{-1}\sum\limits_{s=k+1}^{T}\hat{\varepsilon}_{k,s}\hat{\varepsilon}_{k,s}^{\mathrm{T}}\right)\right\}\end{pmatrix} + o_p(1)$$

where the first equality follows from Theorem 3.1 of Paparoditis (1996), (A.1) and (A.2), and the second follows from (A.1) and (A.3). Thus, it suffices to show the joint convergence of

$$\begin{pmatrix}(T-k)^{1/2}l(k)^{\mathrm{T}}\,\mathrm{vec}(\Delta_k^*\Gamma_k^{-1})\\ (T-k)^{1/2}\,\mathrm{vech}\left\{(T-k)^{-1}\sum\limits_{t=k+1}^{T}\varepsilon_t^*\varepsilon_t^{*\mathrm{T}} - (T-k)^{-1}\sum\limits_{s=k+1}^{T}\hat{\varepsilon}_{k,s}\hat{\varepsilon}_{k,s}^{\mathrm{T}}\right\}\end{pmatrix} \tag{A.4}$$

to the normal distribution. We will apply a vector version of Theorem 24.3 of Davidson (1994) to (A.4), which can be paraphrased as follows:

Let $\{X_{Tt}, \mathcal{F}_{Tt}\}$ be a martingale-difference array, where $\{X_{Tt}\}$ is a triangular array of n-dimensional random vectors and $\mathcal{F}_{Tt}$ is the σ-algebra generated by $X_{T1}, X_{T2}, \dots, X_{Tt}$. If

(a) $\mathrm{plim}_{T\to\infty}\sum_{t=1}^{T}E[X_{Tt}X_{Tt}]^{\mathrm{T}} = V$, where V is an $n\times n$ positive definite matrix, and
(b) $\mathrm{plim}_{T\to\infty}\max_{1\le t\le T}|X_{Tt}| = 0$,

then $S_T = \sum_{t=1}^{T}X_{Tt} \xrightarrow{d} N(0, V)$.

First, we will show that (A.4) is a martingale array. Second, we will show that conditions (a) and (b) of the above theorem are satisfied.

Let $\mathcal{F}_t^* = \sigma(\varepsilon_t^*, \varepsilon_{t-1}^*, \ldots, \varepsilon_1^*)$ denote the σ-algebra generated by $\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_t^*$. Then, conditional on the sample,

$$E^*[\varepsilon_{t+1}^* Y_{t,k}^{*\mathrm{T}} \Gamma_k^{-1} | \mathcal{F}_t^*] = E^*[\varepsilon_{t+1}^* | \mathcal{F}_t^*] Y_{t,k}^{*\mathrm{T}} \Gamma_k^{-1} = \frac{1}{T-k} \sum_{s=k+1}^{T} \hat{\varepsilon}_{k,s} Y_{t,k}^{*\mathrm{T}} \Gamma_k^{-1} = 0 \tag{A.5}$$

and

$$\begin{aligned} E^*\left[\varepsilon_s^* \varepsilon_s^{*\mathrm{T}} - \frac{1}{T-k} \sum_{s=k+1}^{T} \hat{\varepsilon}_{k,s} \hat{\varepsilon}_{k,s}^{\mathrm{T}} \,|\, \mathcal{F}_t^*\right] &= E^*[\varepsilon_s^* \varepsilon_s^{*\mathrm{T}} \mid \mathcal{F}_t^*] - \frac{1}{T-k} \sum_{s=k+1}^{T} \hat{\varepsilon}_{k,s} \hat{\varepsilon}_{k,s}^{\mathrm{T}} \\ &= \frac{1}{T-k} \sum_{s=k+1}^{T} \hat{\varepsilon}_{k,s} \hat{\varepsilon}_{k,s}^{\mathrm{T}} - \frac{1}{T-k} \sum_{s=k+1}^{T} \hat{\varepsilon}_{k,s} \hat{\varepsilon}_{k,s}^{\mathrm{T}} \\ &= 0 \end{aligned} \tag{A.6}$$

By the law of iterated expectations, it follows from (A.5) and (A.6) that (A.4) with $\mathcal{F}_t^*$ is a martingale array under the bootstrap probability measure.

Next, we will show that condition (a) of the above theorem is satisfied.

First, the (1, 1) element in the asymptotic covariance matrix is positive by Theorem 3.3 of Paparoditis (1996). For the lower-right $\{r(r+1)/2\} \times \{r(r+1)/2\}$ matrix in the asymptotic covariance matrix, it suffices to show

$$\frac{1}{T-k} \sum_{t=k+1}^{T} \left\{ \varepsilon_{ti_1}^* \varepsilon_{ti_2}^* - \frac{1}{T-k} \sum_{s=k+1}^{T} \hat{\varepsilon}_{k,si_1} \hat{\varepsilon}_{k,si_2} \right\} \left\{ \varepsilon_{ti_3}^* \varepsilon_{ti_4}^* - \frac{1}{T-k} \sum_{s=k+1}^{T} \hat{\varepsilon}_{k,si_3} \hat{\varepsilon}_{k,si_4} \right\}$$

converges in probability to

$$E(\varepsilon_{ti_1} \varepsilon_{ti_2} \varepsilon_{ti_3} \varepsilon_{ti_4}) - E(\varepsilon_{ti_1} \varepsilon_{ti_2}) E(\varepsilon_{ti_3} \varepsilon_{ti_4})$$

for $\forall i_1, i_2, i_3, i_4 = 1, 2, \ldots, r$. Since

$$(T-k)^{-1} \sum_{t=k+1}^{T} \varepsilon_{ti_1}^* \varepsilon_{ti_2}^* E(\varepsilon_{ti_1} \varepsilon_{ti_2}) + o_p(1)$$

by the inequality of Paparoditis (1996, p. 290, ll.12–13) and

$$(T-k)^{-1} \sum_{t=k+1}^{T} \hat{\varepsilon}_{k,ti_1} \hat{\varepsilon}_{k,ti_2} = E(\varepsilon_{ti_1} \varepsilon_{ti_2}) + o_p(1)$$

by Theorem 2.4 of Paparoditis (1996), it suffices to show that

$$\left| (T-k)^{-1} \sum_{t=k+1}^{T} \left\{ \varepsilon_{ti_1}^* \varepsilon_{ti_2}^* \varepsilon_{ti_3}^* \varepsilon_{ti_4}^* - E(\varepsilon_{ti_1} \varepsilon_{ti_2} \varepsilon_{ti_3} \varepsilon_{ti_4}) \right\} \right| = o_p(1)$$

$$\left|(T-k)^{-1}\sum_{t=k+1}^{T}\left\{\varepsilon^*_{ti_1}\varepsilon^*_{ti_2}\varepsilon^*_{ti_3}\varepsilon^*_{ti_4}-E(\varepsilon_{ti_1}\varepsilon_{ti_2}\varepsilon_{ti_3}\varepsilon_{ti_4})\right\}\right|$$
$$\leq\left|(T-k)^{-1}\sum_{t=k+1}^{T}\left\{\varepsilon^*_{ti_1}\varepsilon^*_{ti_2}\varepsilon^*_{ti_3}\varepsilon^*_{ti_4}-(T-k)^{-1}\sum_{s=k+1}^{T}\hat{\varepsilon}_{k,si_1}\hat{\varepsilon}_{k,si_2}\hat{\varepsilon}_{k,si_3}\hat{\varepsilon}_{k,si_4}\right\}\right|$$
$$+\left|(T-k)^{-1}\sum_{t=k+1}^{T}\{\hat{\varepsilon}_{k,ti_1}\hat{\varepsilon}_{k,ti_2}\hat{\varepsilon}_{k,ti_3}\hat{\varepsilon}_{k,ti_4}-E(\varepsilon_{ti_1}\varepsilon_{ti_2}\varepsilon_{ti_3}\varepsilon_{ti_4})\}\right|$$
$$=\left|\frac{1}{T-k}\sum_{t=k+1}^{T}\left\{\varepsilon^*_{ti_1}\varepsilon^*_{ti_2}\varepsilon^*_{ti_3}\varepsilon^*_{ti_4}-\frac{1}{T-k}\sum_{t=k+1}^{T}\hat{\varepsilon}_{k,ti_1}\hat{\varepsilon}_{k,ti_2}\hat{\varepsilon}_{k,ti_3}\hat{\varepsilon}_{k,ti_4}\right\}\right|+O_p(k^{1/2}T^{-1/2})$$

where the equality follows from arguments analogous to the inequality in Paparoditis (1996, p. 288, ll.9–12) and the proof of his Theorem 2.3. Using arguments similar to the proof that (A.4) is a martingale array, one can show that

$$\left\{\varepsilon^*_{ti_1}\varepsilon^*_{ti_2}\varepsilon^*_{ti_3}\varepsilon^*_{ti_4}-(T-k)^{-1}\sum_{t=k+1}^{T}\hat{\varepsilon}_{k,ti_1}\hat{\varepsilon}_{k,ti_2}\hat{\varepsilon}_{k,ti_3}\hat{\varepsilon}_{k,ti_4}\right\}$$

is a sequence of square-integrable martingale-difference arrays. Hence, we have

$$(T-k)^{-1}\sum_{t=k+1}^{T}\varepsilon^*_{ti_1}\varepsilon^*_{ti_2}\varepsilon^*_{ti_3}\varepsilon^*_{ti_4}-(T-k)^{-1}\sum_{t=k+1}^{T}\hat{\varepsilon}_{k,ti_1}\hat{\varepsilon}_{k,ti_2}\hat{\varepsilon}_{k,ti_3}\hat{\varepsilon}_{k,ti_4}=o_p(1)$$

by the McLeish inequality (Gallant and White, 1988, Theorem 3.11) and the Markov inequality.

Lastly, we consider the lower-left $\{r(r+1)/2\}\times 1$ matrix (or the upper-right $1\times\{r(r+1)/2\}$ matrix) in the asymptotic covariance matrix. Since $M_1\leq\|\ell(k)\|\leq M_2$, it suffices to show that

$$\text{(A.7)}\qquad \frac{1}{T-k}\sum_{t=k+1}^{T}\left(\varepsilon^*_{t,i_1}\varepsilon^*_{t,i_2}\varepsilon^*_{t,i_3}\ell_k^{\mathrm{T}}\mathrm{vec}(Y^*_{k,t-1}\Gamma_k^{-1})-\hat{\varepsilon}_{k,ti_1}\hat{\varepsilon}_{k,ti_2}\hat{\varepsilon}_{k,ti_3}\ell_k^{\mathrm{T}}\mathrm{vec}(Y_{k,t-1}\Gamma_k^{-1})\right)=o_p(1)$$

for $\forall i_1,i_2,i_3,i_4=1,2,\ldots,r$, where ℓ_k is an arbitrary $rk\times 1$ vector such that $M_1\leq\|\ell_k\|\leq M_2$. By the triangle inequality, it follows that

$$\left|(T-k)^{-1}\sum_{t=k+1}^{T}\left(\varepsilon^*_{ti_1}\varepsilon^*_{ti_2}\varepsilon^*_{ti_3}\ell_k^{\mathrm{T}}\mathrm{vec}(Y^*_{k,t-1}\Gamma_k^{-1})-\hat{\varepsilon}_{ti_1}\hat{\varepsilon}_{ti_2}\hat{\varepsilon}_{ti_3}\ell_k^{\mathrm{T}}\mathrm{vec}(Y_{kt-1}\Gamma_k^{-1})\right)\right|$$
$$\text{(A.8)}\qquad\leq\left|(T-k)^{-1}\sum_{t=k+1}^{T}\left(\varepsilon^*_{ti_1}\varepsilon^*_{ti_2}\varepsilon^*_{ti_3}-\hat{\varepsilon}_{ti_1}\hat{\varepsilon}_{ti_2}\hat{\varepsilon}_{ti_3}\right)\ell_k^{\mathrm{T}}\mathrm{vec}(Y^*_{k,t-1}\Gamma_k^{-1}))\right|$$
$$\text{(A.9)}\qquad+\left|(T-k)^{-1}\sum_{t=k+1}^{T}\hat{\varepsilon}_{ti_1}\hat{\varepsilon}_{ti_2}\hat{\varepsilon}_{ti_3}\left(\ell_k^{\mathrm{T}}\mathrm{vec}((Y^*_{k,t-1}-Y_{k,t-1})\Gamma_k^{-1})\right)\right|$$

One can show that

$$\left\{\left(\varepsilon^*_{ti_1}\varepsilon^*_{ti_2}\varepsilon^*_{ti_3} - \frac{1}{T-k}\sum_{t=k+1}^{T}\hat{\varepsilon}_{ti_1}\hat{\varepsilon}_{ti_2}\hat{\varepsilon}_{ti_3}\right)\frac{\ell_k^{\mathrm{T}}\operatorname{vec}(Y^*_{k,t-1}\Gamma_k^{-1})}{k}\right\}$$

with $\mathcal{F}^*_t$ is a sequence of square-integrable martingale-difference arrays along the same lines as the proof that (A.4) is a martingale array. Then, it follows from the McLeish inequality (Gallant and White, 1988, Theorem 3.11), Example 1 of Hall and Heyde (1980, p. 19), and the Markov inequality that (A.8) is $o_p(1)$. Let $\|x\|_2 = (E(x^2))^{1/2}$. Equation (A.9) is bounded by

$$\begin{aligned}
&\left\|(T-k)^{-1}\sum_{t=k}^{T-1}\hat{\varepsilon}_{ti_1}\hat{\varepsilon}_{ti_2}\hat{\varepsilon}_{ti_3}\ell_k^{\mathrm{T}}\operatorname{vec}((Y^*_{k,t-1}-\tilde{Y}_{k,t-1})\Gamma_k^{-1})\right\|_2 \\
&+\left\|(T-k)^{-1}\sum_{t=k}^{T-1}\hat{\varepsilon}_{ti_1}\hat{\varepsilon}_{ti_2}\hat{\varepsilon}_{ti_3}\ell_k^{\mathrm{T}}\operatorname{vec}((\tilde{Y}_{k,t-1}-Y_{k,t-1})\Gamma_k^{-1})\right\|_2 \\
&\leq \left\|(T-k)^{-1}\sum_{t=k}^{T-1}(\hat{\varepsilon}_{ti_1}\hat{\varepsilon}_{ti_2}\hat{\varepsilon}_{ti_3})^2\right\|_2\left\|(T-k)^{-1}\sum_{t=k}^{T-1}(\ell_k^{\mathrm{T}}\operatorname{vec}((Y^*_{k,t-1}-\tilde{Y}_{k,t-1})\Gamma_k^{-1}))^2\right\|_2 \\
&+\left\|(T-k)^{-1}\sum_{t=k}^{T-1}(\hat{\varepsilon}_{ti_1}\hat{\varepsilon}_{ti_2}\hat{\varepsilon}_{ti_3})^2\right\|_2\left\|(T-k)^{-1}\sum_{t=k}^{T-1}(\ell_k^{\mathrm{T}}\operatorname{vec}((\tilde{Y}_{k,t-1}-Y_{k,t-1})\Gamma_k^{-1}))^2\right\|_2 \\
&= O_p(k^3T^{-1/2}) + O_p(k^{5/2}T^{-1/2})
\end{aligned}$$

where $\tilde{Y}_{k,t}$ is defined in Paparoditis (1996, p. 291) and the last equality follows from the inequalities in the proof of Theorem 3.2 of Paparoditis (1996, p. 292, l.9 and l.18). By the Markov inequality, it follows that (A.9) is $o_p(1)$, which completes the proof of (A.7).

Because each element of (A.4) is the sum of $O_p(T^{-1/2})$, condition (b) is satisfied. Therefore, we obtain the desired results. A similar argument can be used to establish the first part of Theorem 1. ■

PROOF OF COROLLARY 1.

$$\begin{aligned}
&(T-k)^{1/2}\{g_k(\operatorname{vec}(\hat{A}(k)),\operatorname{vech}(\hat{\Sigma}_k)) - g_k(\operatorname{vec}(A(k)),\operatorname{vech}(\Sigma)\} \\
&= (T-k)^{1/2}Dg_k(\operatorname{vec}(\bar{A}(k)),\ \operatorname{vech}(\bar{\Sigma}_k))\{(\operatorname{vec}(\hat{A}(k))^{\mathrm{T}},\ \operatorname{vech}(\hat{\Sigma}_k)^{\mathrm{T}}) - (\operatorname{vec}(A(k))^{\mathrm{T}},\ \operatorname{vech}(\Sigma)^{\mathrm{T}})\}^{\mathrm{T}} \\
&= (T-k)^{1/2}Dg_k(\operatorname{vec}(A(k)),\ \operatorname{vech}(\Sigma_k))\{(\operatorname{vec}(\hat{A}(k))^{\mathrm{T}},\ \operatorname{vech}(\hat{\Sigma}_k)^{\mathrm{T}}) - (\operatorname{vec}(A(k))^{\mathrm{T}},\ \operatorname{vech}(\Sigma)^{\mathrm{T}})\}^{\mathrm{T}} \\
&\quad + O_p(kT^{-1/2}) \\
&\xrightarrow{d} N(0,\Omega_g)
\end{aligned}$$

where $(\operatorname{vec}(\bar{A}(k))^{\mathrm{T}},\operatorname{vech}(\bar{\Sigma}_k)^{\mathrm{T}})^{\mathrm{T}}$ is in the line segment between $(\operatorname{vec}(\hat{A}(k))^{\mathrm{T}}, \operatorname{vech}(\hat{\Sigma}_k)^{\mathrm{T}})^{\mathrm{T}}$ and $(\operatorname{vec}(A(k))^{\mathrm{T}},\operatorname{vech}(\Sigma)^{\mathrm{T}})^{\mathrm{T}}$. The first equality follows from the mean-value theorem, the second from Theorem 2.1 of Paparoditis (1996), Assumptions 2(b) and 1(d), and the last convergence from Assumption 2(a) and Assumption 1(d). Therefore, the first result of Corollary 1 follows. The proof of the second result is analogous to that of the first one with the use of Theorem 2.1 of Paparoditis (1996) and $O_p(kT^{-1/2})$ by the use of Theorem 1 and $O_p(k^2T^{-1/2})$, respectively. Thus, we omit the proof of the second result of Corollary 1. ■

Proof of Corollary 2. By the second-to-last equation in the proof of Theorem 3.4 in Paparoditis (1996, p. 295):

$$\sqrt{T}l(h,k)^{\mathrm{T}}\operatorname{vec}(\hat{B}^{*}(h,k)-\hat{B}(h,k))=\sqrt{T}l(h,k)^{\mathrm{T}}\Psi_{k,k}(\hat{a}^{*}(k)-\hat{a}(k))+o_p(1)$$

where $\{l(h,k)\}$ is a sequence of $kr^2\times 1$ vectors with jth element equal to zero for all $j>h$, by the fact that $\partial\operatorname{vec}(P)/\partial\operatorname{vech}(\Sigma)^{\mathrm{T}}=H$ (see Lütkepohl and Poskitt, 1991, p. 495), and by the Cramér–Wold device, it follows that

$$\text{(A.10)}\qquad \begin{pmatrix}\sqrt{T}\operatorname{vec}(\hat{B}^{*}(h,k)-\hat{B}(h,k))\\ \sqrt{T}\operatorname{vec}(\hat{P}_k^{*}-\hat{P}_k)\end{pmatrix}\xrightarrow{d}N\left(0,\begin{pmatrix}\Omega_{bb} & \Omega_{bp}\\ \Omega_{pb} & \Omega_{pp}\end{pmatrix}\right)$$

where

$$\begin{aligned}
\Omega_{bb} &= \lim_{T\to\infty} l(h,k)^{\mathrm{T}}\Psi_{h,k}(\Gamma_k^{-1}\otimes\Sigma)\Psi_{h,k}^{\mathrm{T}}l(h,k)\\
\Omega_{bp} &= \lim_{T\to\infty} l(h,k)^{\mathrm{T}}\Psi_{h,k}E[\operatorname{vec}(\varepsilon_t Y_{t-1}^{\mathrm{T}}\Gamma_k^{-1})\operatorname{vech}(\varepsilon_t\varepsilon_t^{\mathrm{T}})^{\mathrm{T}}]H^{\mathrm{T}}\\
\Omega_{pb} &= \Omega_{bp}^{\mathrm{T}}\\
\Omega_{pp} &= H\Omega_{22}H^{\mathrm{T}}
\end{aligned}$$

By applying the delta method to (A.10) and Equation (10) in Lütkepohl and Poskitt (1991), we obtain the desired result. ■

Proof of Corollary 3. The proof is analogous to the proof of Corollary 2 in Lütkepohl and Poskitt (1991) and thus is omitted. ■

Proof of Corollary 4. The proof of Corollary 4 is analogous to the proof of Corollary 2 and thus is omitted. ■

Proof of Theorem 2. In population, the Granger–Newbold predictability measure is

$$\text{(A.11)}\qquad 1-\frac{E(y_t-E(y_t\mid y_{t-1},y_{t-2},\dots))^2}{E(y_t^2)}=\frac{\sum_{j=1}^{\infty}b_j^2\sigma^2}{\sum_{j=0}^{\infty}b_j^2\sigma^2}=\frac{\sum_{j=1}^{\infty}b_j^2}{\sum_{j=0}^{\infty}b_j^2}$$

where b_j is the jth coefficient of the moving-average representation. By Theorem 3.4 of Paparoditis (1996), a linear combination of the first k moving-average coefficients is asymptotically normally distributed. The proof consists of two parts: First, we shall show that the moving-average version of Assumption 2 is satisfied for

$$g_k(b_1,b_2,\dots,b_k)=\sum_{j=1}^{k}b_j^2\Big/\sum_{j=0}^{k}b_j^2$$

Next, we shall show that $\{g_k\}$ approximates the predictive measure.

Since the gradient vector of g_k is given by

$$Dg_k=\begin{pmatrix}\dfrac{2b_0}{\left(\sum_{j=0}^{k}b_j^2\right)^2} & \dfrac{2b_1}{\left(\sum_{j=0}^{k}b_j^2\right)^2} & \dfrac{2b_2}{\left(\sum_{j=0}^{k}b_j^2\right)^2} & \cdots & \dfrac{2b_k}{\left(\sum_{j=0}^{k}b_j^2\right)^2}\end{pmatrix}^{\mathrm{T}}$$

Assumption 2(a) is satisfied with $M_1 = M_2 = 2/(\sum_{j=0}^{\infty} b_j^2)^{3/2}$. Because the gradient vector is differentiable, Assumption 2(b) is trivially satisfied. Thus, the bootstrap works for a sequence of functions $\{g_k\}$.

Next, we will show that $g_k = (\sum_{j=1}^{k} \hat{b}_{j,k}^{*2})(\sum_{j=0}^{k} \hat{b}_{j,k}^{*2})$ approximates the bootstrap predictive measure $1 - \hat{\sigma}_k^{*2}/\{(1/T)\sum_{t=1}^{T} y_t^{*2}\}$. It is sufficient to show

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\hat{b}_{j,k}^{*2} - \sum_{j=0}^{k}\hat{b}_{j,k}^{*2} = o_p(1) \tag{A.12}$$

which is in turn equivalent to

$$\frac{1}{T}\sum_{t=1}^{k}\sum_{j=t-1}^{k}\hat{b}_{j,k}^{*2} = \frac{1}{T}\sum_{j=1}^{k} j\hat{b}_{j,k}^{*2} = o_p(1) \tag{A.13}$$

and

$$\frac{1}{T}\sum_{t=k+2}^{T}\sum_{j=k+1}^{t-1}\hat{b}_{j,k}^{*2} = \frac{1}{T}\sum_{j=k+2}^{T}(T-j)\hat{b}_{j,k}^{*2} = o_p(1) \tag{A.14}$$

By Theorem 4.1 of Paparoditis (1996), his Equation (10) and the inequality following his Equation (10), it follows that

$$\begin{aligned}\frac{1}{T}\sum_{j=1}^{k} j\hat{b}_{j,k}^{*2} &\leq \frac{2}{T}\sum_{j=1}^{k} j(\hat{b}_{j,k}^{*} - \hat{b}_{j,k})^2 + \frac{2}{T}\sum_{j=1}^{k} j\hat{b}_{j,k}^{2}\\ &= O_p(k^2T^{-2}) + O_p(k^4T^{-1})\\ &= o_p(1)\end{aligned}$$

which proves (A.13). Similarly, by Theorem 4.1 and Equation (10) of Paparoditis (1996),

$$\begin{aligned}\frac{1}{T}\sum_{j=k+2}^{T}(T-j)\hat{b}_{j,k}^{*2} &\leq \frac{2}{T}\sum_{j=k+2}^{T}(T-j)(\hat{b}_{j,k}^{*} - \hat{b}_{j,k})^2 + \frac{2}{T}\sum_{j=k+2}^{T}(T-j)\hat{b}_{j,k}^{2}\\ &= O_p(k^5T^{-1}) + O_p(k^3T^{-1})\\ &= o_p(1)\end{aligned}$$

which establishes Equation (A.14). This completes the proof of the validity of the bootstrap. The derivation of the asymptotic distribution of the P_{GN} statistic is analogous to the above proof and thus is omitted. ■

A.2. *Simulation Design.* The ARMA(2, 4)-DGP for the inflation rate is of the form

$$y_t = 1.794y_{t-1} - 0.8030y_{t-2} + \varepsilon_t - 1.5207\varepsilon_{t-1} + 0.5297\varepsilon_{t-1} - 0.0890\varepsilon_{t-2} + 0.1387\varepsilon_{t-4}$$

where $\varepsilon_t \sim \mathrm{NID}(0, 8.7679)$.

The VARMA(1, 1)-DGP for investment expenditures, the corresponding deflator, and the commercial paper rate is of the form

$$y_t = A_1 y_{t-1} + \varepsilon_t + M_1 \varepsilon_{t-1}$$

where $\varepsilon_t \sim \text{NID}(0, \Sigma)$. The coefficient matrices are

$$A_1 = \begin{bmatrix} 0.5417 & -0.1971 & -0.9395 \\ 0.0400 & 0.9677 & 0.0323 \\ -0.0015 & 0.0829 & 0.8080 \end{bmatrix}, \quad M_1 = \begin{bmatrix} -0.1428 & -1.5133 & -0.7053 \\ -0.0202 & 0.0309 & 0.1561 \\ 0.0227 & 0.1178 & -0.0153 \end{bmatrix}$$

Let P be the lower triangular Cholesky decomposition defined by $PP^{\mathrm{T}} = \Sigma$. Then

$$P = \begin{bmatrix} 9.2352 & 0 & 0 \\ -1.4343 & 3.6070 & 0 \\ -0.7756 & 1.2296 & 2.7555 \end{bmatrix}$$

REFERENCES

Ball, L., and S. G. Cecchetti, "Inflation and Uncertainty at Short and Long Time Horizons," *Brookings Papers on Economic Activity* (1990), 215–45.

Barsky, R. B., "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics* 19 (1987), 3–24.

Berkowitz, J., I. Birgean, and L. Kilian, "On the Finite-Sample Accuracy of Nonparametric Resampling Algorithms for Economic Time Series," in T. B. Fomby and R. C. Hill, eds., *Advances in Econometrics: Applying Kernel and Nonparametric Estimation to Economic Topics* 14 (Greenwich, CT: JAI Press, 1999), 77–107.

——— and L. Kilian, "Recent Developments in Bootstrapping Time Series," *Econometric Reviews* 19 (2000), 1–48.

Bose, A., "Edgeworth Correction by Bootstrap in Autoregression," *Annals of Statistics* 16 (1988), 1709–22.

Braun, P. A., and S. Mittnik, "Misspecifications in Vector Autoregressions and Their Effects on Impulse Responses and Variance Decompositions," *Journal of Econometrics* 59 (1993), 319–41.

Bühlmann, P., "Moving-Average Representations of Autoregressive Approximations," *Stochastic Processes and Their Applications* 60 (1995), 331–42.

———, "Sieve Bootstrap for Time Series," *Bernoulli* 3 (1997), 123–48.

Canova, F., and J. Marrinan, "Sources and Propagation of International Output Cycles: Common Shocks or Transmission?," *Journal of International Economics* 46 (1998), 133–66.

Cogley, T., and J. M. Nason, "Output Dynamics in Real-Business Cycle Models," *American Economic Review* 85 (1995), 492–511.

Cooley, T. F., and M. Dwyer, "Business Cycle Analysis without Much Theory: A Look at Structural VARs," *Journal of Econometrics* 83 (1998), 57–88.

———, and G. D. Hansen, "Money and the Business Cycle," in T. F. Cooley, ed., *Frontiers of Business Cycle Research* (Princeton, NJ: Princeton University Press, 1995), 175–216.

Davidson, J., *Stochastic Limit Theory: An Introduction for Econometricians* (New York: Oxford University Press, 1994).

Diebold, F. X., and L. Kilian, "Measuring Predictability: Theory and Macroeconomic Applications," CEPR Discussion Paper No. 2424, April 2000.

———, L. E. Ohanian, and J. Berkowitz, "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *Review of Economic Studies* 65 (1998), 433–51.

Efron, B., and R. J. Tibshirani, *An Introduction to the Bootstrap* (New York: Chapman & Hall, 1993).

Freedman, D. A., "Bootstrapping Regression Models," *Annals of Statistics* 9 (1981), 1218–28.

Galí, J., "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review* 89 (1999), 249–71.

GALLANT, A. R., AND H. WHITE, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models* (New York: Basil Blackwell, 1988).

GINÉ, E., AND J. ZINN, "Bootstrapping General Empirical Measures," *Annals of Probability* 18 (1990), 851–69.

GRANGER, C. W. J., AND P. NEWBOLD, *Forecasting Economic Time Series*, 2nd ed. (Orlando: Academic Press, 1986).

HALL, P., AND C. C. HEYDE, *Martingale Limit Theory and its Application* (San Diego: Academic Press, 1980).

KILIAN, L., "Small-Sample Confidence Intervals for Impulse Response Functions," *Review of Economics and Statistics* 80 (1998), 218–30.

———, "Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Responses," *Review of Economics and Statistics* 81 (1999), 652–60.

KING, R. G., AND M. W. WATSON, "Money, Prices, Interest Rates and the Business Cycle," *Review of Economics and Statistics* 30 (1996), 35–53.

KREISS, J. P., "Asymptotic Statistical Inference for a Class of Stochastic Processes," Habilitationsschrift, Universität Hamburg, 1988.

LEEPER, E. M., AND C. A. SIMS, "Toward a Modern Macroeconomic Model Usable for Policy Analysis," in S. Fischer and J. J. Rotemberg, eds., *NBER Macroeconomics Annual 1994* (Cambridge, MA: MIT Press, 1994), 81–118.

LÜTKEPOHL, H., "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics* 72 (1990), 116–25.

———, *Introduction to Multiple Time Series Analysis*, 2nd ed. (New York: Springer-Verlag, 1993).

———, AND D. S. POSKITT, "Estimating Orthogonal Impulse Responses via Vector Autoregressive Models," *Econometric Theory* 7 (1991), 487–96.

MITTNIK, S., AND P.-A. ZADROZNY, "Asymptotic Distributions of Impulse Responses, Step Responses, and Variance Decompositions of Estimated Linear Dynamic Models, *Econometrica* 61 (1993), 857–70.

PAPARODITIS, E., "Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes," *Journal of Multivariate Analysis* 57 (1996), 277–96.

———, AND B. STREITBERG, "Order Identification Statistics in Stationary Autoregressive Moving-Average Models: Vector Autocorrelations and the Bootstrap," *Journal of Time Series Analysis* 13 (1992), 415–34.

ROTEMBERG, J., AND M. WOODFORD, "Real-Business Cycle Models and the Forecastable Movements in Output, Hours, and Consumption," *American Economic Review* 86 (1996), 71–89.

———, AND ———, "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in B. S. Bernanke and J. J. Rotemberg, eds., *NBER Macroeconomics Annual 1997* (Cambridge, MA: MIT Press, 1997), 297–346.

SCHMITT-GROHÉ, S., "The International Transmission of Economic Fluctuations: Effects of U.S. Business Cycles on the Canadian Economy," *Journal of International Economics* 44 (1998), 257–87.

WATSON, M. W., "Measures of Fit for Calibrated Models," *Journal of Political Economy* 101 (1993), 1011–41.