ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

(Structural) VAR models with ignored changes in mean and volatility[☆]

Matei Demetrescu [a], Nazarii Salish [b,*]

[a] Department of Statistics, TU Dortmund University, Vogelpothsweg 78, D-44227 Dortmund, Germany
[b] Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain

ARTICLE INFO

Keywords:
Structural breaks
High-order autoregression
Nonstationarity
Identification
Forecasting practice

ABSTRACT

The paper discusses how standard forecasting tools in multivariate time series analysis are affected when ignoring possible changes in the mean and the (co)variance. We study the estimation, forecasts, and estimated impulse responses of so-called long vector autoregressions, for which the complexity of the model increases with the sample size. We prove that, in spite of structural change in the data generating process, coefficient estimates and out-of-sample forecasts based on such long vector autoregressions are consistent. The sampling behaviour of estimated impulse responses depends primarily on the residual covariance matrix, which converges to an "average" covariance matrix in the case of varying (co)variances. Localised estimators (also obtained by means of a suitable long vector autoregression) may be more suitable in this case. Monte Carlo simulations support our theoretical findings. The empirical relevance of the theory is illustrated in two applications: (i) the international dynamics of inflation, and (ii) uncertainty and economic activity.

1. Motivation

Whether they are used for impulse-response analysis, for providing error variance decompositions, or for forecasting (unconditional or conditional on different scenarios for shocks), structural vector autoregression (SVAR) models have become a standard tool in applied macroeconomic work. The empirical literature using SVAR models is vast, and we refer the reader to Kilian (2013) and Kilian and Lütkepohl (2017) for a review of the developments that have made SVAR models what they are today.

Such developments include for instance the exploitation of breaks in the covariance matrix of the reduced-form errors to identify structural parameters (Rigobon, 2003). This requires explicit modelling of changes in the error covariance matrix; see, among others, Lanne et al. (2010, 2017), Lewis (2021), Lütkepohl and Netšunajev (2017), and Milunovich and Yang (2013).[1] And indeed, shifts in the volatility are not uncommon for economic data; see e.g. Clark (2009) and Stock and Watson (2002) or, for financial data, Amado and Teräsvirta (2014) and Guidolin and Timmermann (2006). Changes in the mean on the other hand are less informative regarding structural forms. Although of concern when modelling—see

☆ The authors would like to thank Majid Al Sadoon, Benjamin Born, Jörg Breitung, Kai Carstensen, Manfred Deistler, Juan J. Dolado, Tobias Görbert, Christoph Hanck, Uwe Hassler, Helmut Lütkepohl, Lutz Kilian, Ingo Klein and Rolf Tschernig for helpful comments and suggestions. Financial support from the Deutsche Forschungsgemeinschaft (DFG), Germany through grants DE 1617/2-1 and DE 1617/4-1 is gratefully acknowledged by the first author. The second author also gratefully acknowledges the financial support from Juan de la Cierva Incorporación, Spain, IJC2019-041742-I.

* Corresponding author.
E-mail addresses: mdeme@stat-econ.uni-kiel.de (M. Demetrescu), nsalish@eco.uc3m.es (N. Salish).

1 Along the same lines, the Bayesian approach allows for a more tractable treatment of time-varying volatility (and, more generally, time-varying parameters); see for example, Bognanni (2018), Cogley and Sargent (2005), Primiceri (2005), Sims and Zha (2006), and Uhlig (1997).

https://doi.org/10.1016/j.ijforecast.2023.06.002

e.g. the local-level model proposed by Stock and Watson (2005)—detrending is often just a preliminary step of the analysis; see e.g. Carstensen and Salzmann (2017).

However, modelling changes in the mean or the error covariance matrix is challenging: such changes may be hard to detect, for instance due to low signal-to-noise ratios in macroeconomic data, even if methods for detecting structural change are well developed (e.g. Bai & Perron, 2003). As a consequence, changes often go undetected in practice or are not even considered as a possibility. For example, the influential work of Christiano et al. (2005) works with a time-invariant nine-dimensional vector autoregression (VAR) with a diagonal covariance matrix of the shocks. The model is estimated with data from 1965 to 1995, such that the Great Moderation, say, would not easily be detected in this dataset. Mertens and Ravn (2013) use data from 1950 to 2006 when estimating the dynamic effect of tax changes using a proxy SVAR, but still postulate models with constant coefficient matrices and a unity covariance matrix of the structural shocks.[2] In their discussion of the informative content of narrative sign restrictions, Antolín-Díaz and Rubio-Ramírez (2018) resort exclusively to VAR models with a constant intercept and white noise errors with a constant covariance matrix, even if their data extend to 2016 and thus even cover the end of the Great Moderation signalled by Clark (2009). Baumeister and Hamilton (2019) on the other hand do allow for possible structural change in a Bayesian framework, albeit not in the sample: their model is only allowed to change when taking earlier data into account to generate a more informative prior.

It is therefore of practical relevance to know the consequences of ignoring changes in the moments of the series to be modelled. For ordinary least squares (OLS) estimation with neglected changes in the (co)variance, the consistency of the autoregressive matrix estimators is not affected (see Phillips & Xu, 2006, for the univariate case), but a weighted LS estimation scheme may lead to more precise estimates.[3] On the other hand, changes—not accounted for—in the mean are known to bias autoregressive roots; see e.g. Ng and Vogelsang (2002). Demetrescu and Hassler (2016) and Wang et al. (2013) suggest that fitting an autoregression to a process with ignored changes in the mean is not *a priori* wrong: it is argued that (some of the effects of) the changes are implicitly mitigated when fitting autoregressive models that are sufficiently complex. Technically, this is achieved by letting the order of the autoregressive fit increase as the sample size grows to infinity. Such high-order (or "long") autoregressions have in fact often been employed in the analysis of stationary time series. Bhansali (1978) uses them to optimally forecast (in the mean-squared-error sense) univariate general linear processes with i.i.d. innovations and absolutely summable coefficients; Lewis and Reinsel (1985) provide the relevant multivariate extension. Berk (1974) shows spectral density estimators based on long autoregressions to be consistent and, more recently, Poskitt (2007) allows for long memory. Wang et al. (2013) and Demetrescu and Hassler (2016) study the univariate case with changes in the mean, while Gonçalves and Kilian (2007) allow for conditionally heteroskedastic innovations but maintain the stationarity assumption.

This paper discusses the estimation VAR models which are taken to be stationary, ignoring possible changes in both their means and their (co)variance. The structural changes in mean and (co)variance we consider here may be deterministic or stochastic in nature (subject to some smoothness restrictions), and we examine the consequences of not modelling them explicitly. Concretely, we derive the first-order asymptotic behaviour of the estimators of VAR coefficient matrices, of the resulting forecasts, and of the estimated impulse response sequences. Finally, as reported in Demetrescu and Hassler (2016) (see Section 4) changes in dynamic structure will not allow us to obtain any kind of convergence within the scope of a least-squares (LS) estimator/long VAR and hence will not yield a valid forecast function. Therefore, we concentrate our attention only on the relevant cases and do not consider changes in autoregressive coefficients.

In more detail, our findings are as follows. After describing the model in Section 2, we show in Section 3 that the VAR coefficient matrices and point forecasts are consistent for the true values when suitably letting the model order increase with the sample size; information criteria may be used to this end. The finding holds true irrespective of possible structural changes in mean or (co)variance. This contrasts with keeping the model order fixed, which leads to inconsistent estimators whenever the changes are in the mean. Changes in the (co)variances are negligible (up to losses in efficiency) as long as information from the residual covariance matrix is not used for identification. Not surprisingly, ignoring the changes in (co)variances when estimating the covariance matrix of the reduced-form shocks leads to a blend of variance regimes which affect identification; see e.g. Patilea and Raïssi (2020). Should this be the case, a *localised* estimator of the covariance matrix may be the better choice, if one is not willing, or not in a position, to model the changes in the (co)variances explicitly. In this context, we show how a long vector autoregression fitted to the cross-products of the residuals of the long VAR model could be used to achieve such local estimation. Finite-sample evidence is provided in Section 4, while Section 5 provides two empirical examples further quantifying the point that ignoring structural change in the mean is a viable strategy in practice, unless there is little uncertainty about the nature of the change. Section 6 summarises our findings

[2] Mertens and Ravn, op. cit., also use wild bootstrap-based confidence intervals for impulse responses. The wild bootstrap is actually reputed to be capable of dealing with unconditional heteroskedasticity as well; see, among many other successful applications, Gonçalves and Kilian (2004, 2007). This is not the case when nonlinear transformations of the bootstrap shocks are involved, however; see Brüggemann et al. (2016) and, specifically for the case of proxy VARs, Jentsch and Lunsford (2019).

[3] Weights would have to be estimated, however. Xu & Phillips, 2008 resort to a one-sided nonparametric smoother for this purpose; see Patilea & Raïssi, 2012 for the VAR case. So it is actually not uncommon to even knowingly neglect the changes in the (co)variance. Needless to say, neglect is not an option if modelling the dynamics of volatility is the key aspect of the analysis; see Lütkepohl and Schlaak (2018) for a recent discussion of selecting between volatility models.

and recommendations for applied work. All formal derivations as well as extensive Monte Carlo results have been gathered in Supplementary Appendix A and B.

In what concerns notation, $\|\cdot\|$ stands for both the Euclidean vector norm and the corresponding induced matrix norm. Boldface symbols stand for vectors and matrices. Stochastic orders of magnitude are denoted by the usual probabilistic Landau symbols, O_p and o_p. The lag operator is denoted by L, $Ly_t = y_{t-1}$, while $\mathbf{I}_K$ is the identity matrix of size $K\times K$, $\boldsymbol{\iota}_q$ denotes a q-vector of ones, and vech$(\cdot)$ is the half-vectorisation operator.

2. Time-varying VAR model

Consider a system of K series with the following component representation:

$$\boldsymbol{y}_t = \boldsymbol{m}_t + \boldsymbol{x}_t, \qquad t = 1,\dots,T, \tag{1}$$

where $\boldsymbol{m}_t$ is a $K\times 1$ vector of (possibly random) functions of time, and the zero-mean stochastic component $\boldsymbol{x}_t$ is given by the $K\times 1$ reduced-form vector autoregression

$$\boldsymbol{x}_t = \mathbf{A}_1\boldsymbol{x}_{t-1} + \cdots + \mathbf{A}_p\boldsymbol{x}_{t-p} + \boldsymbol{\varepsilon}_t \tag{2}$$

with $\mathbf{A}_j$ $K\times K$ matrices of autoregressive parameters. The autoregressive characteristic polynomial $\mathbf{A}(L) = \mathbf{I}_K - \sum_{j=1}^{p}\mathbf{A}_jL^j$ is assumed to have stable roots throughout this paper.

Therefore, the process $\boldsymbol{y}_t$ has an MA representation,

$$\boldsymbol{y}_t = \boldsymbol{m}_t + \sum_{k\geq 0}\mathbf{C}_k\boldsymbol{\varepsilon}_{t-k},$$

where the MA coefficient matrices $\mathbf{C}_k$ are obtained by inversion of the VAR matrix lag polynomial $\mathbf{A}(L)$ and are known have exponential decay.

The zero-mean white noise innovations $\boldsymbol{\varepsilon}_t$ follow a stochastic volatility process given as

$$\boldsymbol{\varepsilon}_t = \mathbf{H}_t\,\boldsymbol{\epsilon}_t, \tag{3}$$

with $\boldsymbol{\epsilon}_t$ a standardised uncorrelated sequence and $\mathbf{H}_t$ a $K\times K$ scaling matrix, as specified below.

Assumption 1. Let $\boldsymbol{m}_t = \boldsymbol{\nu}(t/T)$, where $\boldsymbol{\nu}(s)$ is a $K\times 1$ vector process whose entries are bounded and almost surely pathwise piecewise Lipschitz continuous with a finite number of discontinuity points. Furthermore, assume without loss of generality (w.l.o.g.) that $\int_0^1\boldsymbol{\nu}(s)\,ds = \mathbf{0}$ and rank $\mathbf{M} = r$ almost surely, where $\mathbf{M} = \int_0^1\boldsymbol{\nu}(s)\,\boldsymbol{\nu}'(s)\,ds$ and $1\leq r\leq K$ is fixed.

Assumption 2.

(i) $\mathbf{H}_t = \mathbf{H}\left(t/T\right)$ is a $K\times K$ matrix of bounded functions on the interval $[-\infty, 1]$ and almost surely positive definite. Each entry of the matrix $\mathbf{H}(\cdot)$ is almost surely pathwise piecewise Lipschitz continuous with a finite number of discontinuity points.

(ii) The standardised innovations $\boldsymbol{\epsilon}_t$ are a martingale difference sequence with regard to the filtration $\mathcal{F}_{t-1} = \{\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2},\dots; \boldsymbol{m}_t, \boldsymbol{m}_{t-2},\dots; \mathbf{H}_{t-1}, \mathbf{H}_{t-2},\dots\}$, satisfying $\mathrm{Cov}(\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}) = \mathbf{I}_K$ and $\sup_{t\in\mathbb{Z}}\mathbb{E}\left(\|\boldsymbol{\epsilon}_t\|^4\right) < \infty$.

Assumption 1 encompasses a wide range of common setups for changes in the mean vector used in applied literature, including single/multiple step changes (breaks), slowly evolving trends, smooth transitions, or a random level model. These may be deterministic as well as stochastic in nature, and in fact even endogenously generated within the model, as long as the innovations remain unpredictable at time t given the past. This type of infill asymptotics for deterministic components is commonly used in the literature on structural change, and ensures that the gap between $\boldsymbol{m}_t$ and $\boldsymbol{m}_{t-1}$ goes to zero as $T\to\infty$ at all times with the exception of sudden breaks. To interpret the condition $\int_0^1\boldsymbol{\nu}(s)\,ds = \mathbf{0}$, consider a case where $\boldsymbol{m}_t$ is deterministic. Then the condition $\int_0^1\boldsymbol{\nu}(s)\,ds = \mathbf{0}$ essentially implies that the process $\boldsymbol{y}_t$ is demeaned. Hence, assuming that $\boldsymbol{m}_t$ has stochastic nature as described by Assumption 1 implies that the process $\boldsymbol{y}_t$ is demeaned "on average" and is used without loss of generality to simplify the discussion.

Further, $\mathbf{M}$ may then be interpreted as a second-moment matrix of *the path* of the mean component. We focus on the nontrivial case when there are changes in the mean (i.e. $\mathbf{M}\neq\mathbf{0}$, and accordingly $r > 0$) and they are ignored when estimating the long vector autoregression for $\boldsymbol{y}_t$.

Assumption 2 describes the heterogeneous nature of the shocks $\boldsymbol{\varepsilon}_t$. Similar to Assumption 1, the piecewise Lipschitz condition in Assumption 2(i) accounts for many empirically relevant setups with unknown changes in the conditional (co)variance dynamics, such as step changes or smooth transitions (trending or periodic), be they deterministic or stochastic in nature. In particular, volatility dynamics are described by the conditional covariance matrix $\boldsymbol{\Sigma}_t \equiv \mathbb{E}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}\right) = \mathbf{H}_t\mathbf{H}_t'$.

This covariance matrix is time varying in a (piecewise) smooth manner, and we may speak of an "average" covariance matrix, defined as

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\Sigma}_t \xrightarrow{p} \int_0^1\boldsymbol{\Sigma}_{[sT]}ds = \int_0^1\mathbf{H}(s)\mathbf{H}'(s)ds := \boldsymbol{\Omega}.$$

(The convergence as $T\to\infty$ to the integral representation is implied by the pathwise Lipschitz condition on $\mathbf{H}$.) Finally, we require $\mathbf{H}(\cdot)$ to be defined on the interval $[-\infty, 1]$, since the vector MA(∞) representation of $\boldsymbol{y}_t$ is used to establish some of the results.

The following section provides a guide of what to expect when ignoring time-varying components (as in Assumptions 1 and 2) in VAR models.

3. Main results

3.1. Long autoregression

We now introduce some notation to describe the long autoregression estimation procedure which will help us to deal with the time-varying nature of the mean and the volatility of $\boldsymbol{y}_t$. When estimating models (1)–(3) we simply ignore the possibility of changes, and fit instead a long vector autoregression of $\boldsymbol{y}_t$ estimated by the LS. That is, while p in the data generating Eq. (2) remains fixed, the number of lags, denoted as h_T, that we include in the estimation procedure is growing with T. Hence, the true matrix of autoregressive matrices that we are attempting to estimate takes a form as the $K \times (K h_T)$ matrix $\mathbf{A}_{(h_T)} = \left[\mathbf{A}_1; \ldots; \mathbf{A}_p; \mathbf{0}; \ldots; \mathbf{0}\right]$ for $h_T \geq p$. In what follows we also require the matrix of "average" second moments of the process $\boldsymbol{x}_t$. Hence, we define the $(K h_T) \times (K h_T)$ matrix $\boldsymbol{\Sigma}_{(h_T)}$ whose (i, j)-th block is the matrix $\bar{\mathbf{G}}_{j-i}$ defined as $\bar{\mathbf{G}}_h = \sum_{k\geq 0} \mathbf{C}_k \boldsymbol{\Omega} \mathbf{C}'_{k+h}$ for positive h and as $\bar{\mathbf{G}}'_{-h}$ for negative h, and $\boldsymbol{\Gamma}_{(h_T)}$ is the $K \times (K h_T)$ matrix with $\bar{\mathbf{G}}_i$ as the ith block, $1 \leq i \leq h_T$. Recall that $\{\mathbf{C}_k\}_{k\geq 0}$ denotes coefficient matrices of the MA representation of the process $\boldsymbol{y}_t$. The matrix $\boldsymbol{\Sigma}_{(h_T)}$ can be seen as the "average" covariance matrix of regressors in (2), i.e. of the stacked h_T lags $\boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_{t-h_T}$. Similarly, $\boldsymbol{\Gamma}_{(h_T)}$ forms the "average" covariance matrix of the left-hand-side vector $\boldsymbol{x}_t$ and its stacked h_T lags on the right-hand side. These matrices relate to the VAR coefficient matrices $\mathbf{A}_j$ for $j = 1, \ldots, h_T$, as commented in the following remark.

Remark 1. In the case of constant volatility, the matrix $\boldsymbol{\Gamma}_{(h_T)} \boldsymbol{\Sigma}^{-1}_{(h_T)}$ is immediately seen to contain the coefficient matrices $\mathbf{A}_j$ for $j = 1, \ldots, h_T$ (where of course, $\mathbf{A}_j = 0$ for $j > p$), i.e. $\mathbf{A}_{(h_T)} = \boldsymbol{\Gamma}_{(h_T)} \boldsymbol{\Sigma}^{-1}_{(h_T)}$. But this is in fact true as well even if the volatility of $\boldsymbol{\varepsilon}_t$ is time varying as defined in Assumption 2 (see e.g. Phillips & Xu, 2006 for the univariate deterministic case), and we make extensive use of this relation to establish the main results of the paper; see the Appendix for details.

When estimating VAR coefficient matrices, it is the time-varying mean component $\boldsymbol{m}_t$ that appears to have the more problematic consequences. Namely, time-varying volatility may lead to efficiency loss; but, if ignoring changes in the mean, the sample counterparts of $\boldsymbol{\Gamma}_{(q)}$ and $\boldsymbol{\Sigma}_{(q)}$ (on which OLS estimation is based) are asymptotically biased for any fixed model order q. See also Lemma A.3 in the Appendix. For *fixed* model orders, this translates into a lack of consistency of the LS estimators of the autoregressive coefficient matrices $\mathbf{A}_j$, as already pointed out by Ng and Vogelsang (2002). When letting the order of the fitted model go to infinity, however, the asymptotic behaviour may change; see Demetrescu and Hassler (2016) for a univariate analysis with constant volatility. Our interest therefore lies in the behaviour of the LS estimators of the AR coefficient matrices for $h_T \to \infty$ under time-varying volatility.

We now examine the OLS estimation of a long autoregression for the model given by Eqs. (1)–(3) and Assumptions 1 and 2. Let $\hat{\mathbf{A}}_j$ denote the OLS estimation of the coefficient matrices $\mathbf{A}_j$ in the vector autoregression of order $h_T < T$,

$$\boldsymbol{y}_t = \hat{\mathbf{A}}_1 \boldsymbol{y}_{t-1} + \cdots + \hat{\mathbf{A}}_{h_T} \boldsymbol{y}_{t-h_T} + \hat{\boldsymbol{\varepsilon}}_t, \tag{4}$$

where h_T grows to infinity with T, but at a slower rate than the sample size T. Without loss of generality, an intercept is not included for the theoretical derivations, since the mean component is zero on average according to Assumption 1; the intercept would of course be used in practice. We have the following:

Proposition 1. *Let*

$$\tilde{\mathbf{A}}_{(h_T)} = \mathbf{A}_{(h_T)} + \left(\mathbf{I}_K - \mathbf{A}_{(h_T)} \left(\boldsymbol{\iota}_{h_T} \otimes \mathbf{I}_K\right)\right) \mathbf{B}^{-1}_{h_T} \left(\boldsymbol{\iota}'_{h_T} \otimes \mathbf{M}\right) \boldsymbol{\Sigma}^{-1}_{(h_T)}, \tag{5}$$

where $\mathbf{B}_{h_T} = \mathbf{I}_K + \left(\boldsymbol{\iota}'_{h_T} \otimes \mathbf{M}\right) \boldsymbol{\Sigma}^{-1}_{(h_T)} \left(\boldsymbol{\iota}_{h_T} \otimes \mathbf{I}_K\right)$. *Under Assumptions 1 and 2, it holds for* $h_T^{-1} + h_T T^{-1/4} \to 0$ *as* $T \to \infty$ *that*

(1) $\left\| \hat{\mathbf{A}}_{(h_T)} - \tilde{\mathbf{A}}_{(h_T)} \right\| = o_p\left(h_T^{-0.5}\right)$, *and*

(2) $\left\| \hat{\mathbf{A}}_{(h_T)} - \mathbf{A}_{(h_T)} \right\| = O_p\left(h_T^{-0.5}\right)$.

Proof. See the Appendix.

Proposition 1 shows us several important facts. First, it demonstrates why estimating VAR model with a fixed number of lags will result in inconsistent outcomes. The second term on the right-hand side (r.h.s) of Eq. (5) gives us a closed-form expression of the bias, which we denote for the simplicity of the following discussion as $Bias(\mathbf{M}, \boldsymbol{\Omega})$. That is, in the case of fixed p-lags, the OLS estimator will converge to $\hat{\mathbf{A}}_{(p)} = \mathbf{A}_{(p)} + Bias(\mathbf{M}, \boldsymbol{\Omega})$, where $Bias(\mathbf{M}, \boldsymbol{\Omega}) \neq \mathbf{0}$ as long as there are some changes in the mean, i.e. $\mathbf{M} \neq \mathbf{0}$.

Second, it shows that volatility dynamics can affect only the magnitude of the bias through $\boldsymbol{\Sigma}_{(h_T)}$, which is composed of the MA coefficient matrices $\{\mathbf{C}_k\}_{k\geq 0}$ and the matrix $\boldsymbol{\Omega} = \int_0^1 \mathbf{H}(s)\mathbf{H}'(s)ds$, depending on changes in volatility.

Third, if we allow the number of lags to increase with the sample size but not too fast (i.e. $h_T \to \infty$ and $h_T/T^{-1/4} \to 0$), then the proposition implies "pointwise" convergence of the long VAR coefficient matrix estimates, even when $\mathbf{M} \neq \mathbf{0}$. To be more precise, the difference to a fixed-order vector autoregression is that the bias derived in the proposition is asymptotically negligible for each coefficient matrix alone, i.e. $\hat{\mathbf{A}}_i \to \mathbf{A}_i$ for each $1 \leq i \leq p$ and $\hat{\mathbf{A}}_i \to 0$ for each $p < i \leq h_T$.[4] Consistency notwithstanding, since $\tilde{\mathbf{A}}_{(h_T)} \neq \mathbf{A}_{(h_T)}$ when $\mathbf{M} \neq \mathbf{0}$, the proposition shows that the long VAR estimators exhibit what may be interpreted as second-order bias, while fixed-order VAR estimation leads to asymptotic non-vanishing first-order bias.

Finally, Proposition 1 provides a generalisation of the results obtained by Demetrescu and Hassler (2016) to the multivariate, possibly nonstationary, case. In particular, it establishes the rates of the convergence of the long

[4] This fact follows from both items of the proposition and Eq. (5). That is, $\left\|\hat{\mathbf{A}}_i - \mathbf{A}_i\right\| \leq \left\|\hat{\mathbf{A}}_i - \tilde{\mathbf{A}}_i\right\| + \left\|\tilde{\mathbf{A}}_i - \mathbf{A}_i\right\|$, where by the first item of the proposition, we have $\left\|\hat{\mathbf{A}}_i - \tilde{\mathbf{A}}_i\right\| = o_p(1)$. Following the same arguments as for the proof of item 2 yields that $\left\|\tilde{\mathbf{A}}_i - \mathbf{A}_i\right\| = O(h_T^{-1/2})$. The intuition behind this result follows from Eq. (5). That is, $\left\|\tilde{\mathbf{A}}_i - \mathbf{A}_i\right\| = \left\|\left(\mathbf{I}_K - \mathbf{A}_{(h_T)} \left(\iota_{h_T} \otimes \mathbf{I}_K\right)\right) \mathbf{B}^{-1}_{h_T} \mathbf{W}_i\right\|$, where $\mathbf{W}_i$ is the corresponding $K \times K$ block of $\left(\iota'_{h_T} \otimes \mathbf{M}\right) \Sigma^{-1}_{(h_T)}$. Matrix $\left(\mathbf{I}_K - \mathbf{A}_{(h_T)} \left(\iota_{h_T} \otimes \mathbf{I}_K\right)\right)$ is bounded, whereas $\mathbf{B}^{-1}_{h_T} \mathbf{W}_i$ gives the rate $h_T^{-1/2}$. See the details in the proof of item 2.

VAR OLS estimators given in (4) to the true parameters, even under changes in the mean and the volatility. The rate restrictions on h_T are somewhat different though, since Demetrescu and Hassler (2016) do not allow for smoothly varying volatility, and the variation of the volatility adds a source of approximation errors that need to be controlled for by stricter conditions on the order h_T of the fitted model.

3.2. Lag selection

Let us briefly discuss the issue of selecting the model order and give a practical answer to the question, "what *is* long autoregression in the first place"?

The use of information criteria such as the Akaike information criterion (AIC) can help determine the appropriate order h_T, as shown in Demetrescu and Hassler (2016). If we ignore breaks in the mean, this approach leads to $h_T \to \infty$. The result can immediately be extended here, but we omit the details to save space. In essence, the information criterion, as a function of a hypothetical model order ℓ, can be proven to converge uniformly in $\ell = 1, \ldots, h_{\max}$ to a monotonically decreasing function, even when there are changes in the (co)variances. This ensures that larger model orders are selected as the sample size increases.

When employing an information criterion, it is necessary to specify an upper bound for the maximum number of lags, h_{max}. In econometric and forecasting literature, a common rule of thumb suggests setting $h_{max} = [4\,(T/100)^{0.25}]$. For example, we apply this rule to the simulated data in Section 4. However, practitioners should exercise caution and consider adjusting this bound based on additional economic characteristics of the data, such as data frequency or seasonality, which might require a larger h_{max}. For instance, in our empirical application concerning uncertainty and economic activity (see Section 5.2), we set the minimum number of lags to 12 for all models (as this choice adequately captures the dynamics of the given example) and $h_{max} = 24$ to allow for the long VAR aspect.

Finally, we caution against fixing a small model order without considering the implications. It is not uncommon in applied work to fix a small order, such as $h_T = p = 1$ or $h_T = p = 2$.[5] This practice is often employed to control the number of parameters to be estimated. However, it is only valid when the changes in the mean are negligible, as stated in Proposition 1. In Section 4, we provide illustrations that demonstrate the effectiveness of information criteria in this regard.

3.3. Forecasts

We now elaborate on the usefulness of the derived convergence rates and bias expression. Proposition 2, below, facilitates the understanding of how neglecting the time-varying mean and the unconditional (co)variances affects dynamics and forecasts based on these estimators. To this end, let $\boldsymbol{y}_t\,(H)$ denote a mean-squared-error (MSE)-optimal point forecast of $\boldsymbol{y}_{t+H}$ given time t information (i.e. the H-step-ahead forecast at time t), given as

$$\boldsymbol{y}_t\,(H) = \boldsymbol{m}_{t+H} + \boldsymbol{x}_t\,(H)$$

with $\boldsymbol{x}_t\,(H)$ as the (iterated) H-step-ahead MSE-optimal point forecasts of the latent component $\boldsymbol{x}_t$,

$$\boldsymbol{x}_t\,(H) = \sum_{j=1}^{H-1} \mathbf{A}_j \boldsymbol{x}_t\,(H-j) + \sum_{j=H}^{p} \mathbf{A}_j \boldsymbol{x}_{t+H-j}.$$

We use the convention that $\sum_a^b \cdot = 0$ of $b < a$, and the iteration is initialised at

$$\boldsymbol{x}_t\,(1) = \mathbf{A}_1 \boldsymbol{x}_t + \cdots + \mathbf{A}_p \boldsymbol{x}_{t-p+1}.$$

The estimated point forecast is based on the fitted autoregression,

$$\hat{\boldsymbol{y}}_t\,(H) = \sum_{j=1}^{H-1} \hat{\mathbf{A}}_j \hat{\boldsymbol{y}}_t\,(H-j) + \sum_{j=H}^{h_T} \hat{\mathbf{A}}_j \boldsymbol{y}_{t+H-j},$$

where the H-step-ahead forecast is computed iteratively, and the first iteration is

$$\hat{\boldsymbol{y}}_1\,(1) = \hat{\mathbf{A}}_1 \boldsymbol{y}_t + \cdots + \hat{\mathbf{A}}_{h_T} \boldsymbol{y}_{t-h_T+1}.$$

Proposition 2. *Under the assumptions of Proposition 1, it holds that*

1. $\sum_{j=1}^{h_T} \hat{\mathbf{A}}_j - \mathbf{R} \xrightarrow{p} \mathbf{0}$, *where the* $K \times K$ *matrix* $\mathbf{R}$ *has* r *unity eigenvalues almost surely;*
2. $\hat{\boldsymbol{y}}_t\,(H) - \boldsymbol{y}_t\,(H) \xrightarrow{p} \mathbf{0}$, *for each* $t = h_T + 1, \ldots, T$.

Proof. See the Appendix.

The first result of the proposition implies that the fitted process (4) appears to have unit roots. This is irrespective of the path of the mean process, as long as $r > 0$.

Remark 2. Changes in mean have long been known to mimic persistence, e.g. in the form of fractional unit roots. This insight can be traced back to Bhattacharya et al., 1983. But the mechanism is not so obvious here, since *individual* coefficient matrices can still be consistently estimated; in fact it is only *their sum* over all h_T coefficient matrix estimators that is asymptotically first-order biased under ignored changes in mean. This suggests that using long-run identification in the style of Blanchard and Quah (1989) may be particularly problematic, given that such identification builds the sum of the VAR coefficients.[6]

Remark 3. Depending on whether $r = K$ or $r < K$, the time series $\boldsymbol{y}_t$ may even appear to be cointegrated. It is this pseudo-cointegration behaviour (which is nothing else than a co-feature; see Engle & Kozicki, 1993)

[5] See, among others, Benati (2008), Cogley and Sargent (2005), and Gambetti et al. (2008).

[6] We are thankful to one of the anonymous referees for pointing out the practical implications of the bias of the sum of the autoregressive coefficient matrix estimators.

that ensures forecast consistency, since changes in the mean cancel out in the direction of the co-feature space, while they are differenced away in the complement of the co-feature space. See the proof of the second part of Proposition 2 for the technical details.

This pseudo-(co-)integrating behaviour is essential in establishing the second result, namely that point forecasts are still consistent even when ignoring changes in the mean. One follow-up question of interest is: How do path forecasts, interval forecasts, and fan charts react to ignored changes in the mean? The answer depends, naturally, on how these forecasts are constructed. We note for instance that path forecasts, consisting of a sequence of point forecasts, are not asymptotically affected. Interval forecasts and fan charts, which also reflect the scale of the error distribution and not just the central tendency, are typically affected by time-varying variance. We discuss the implications and ways to deal with such issues for the important case of impulse responses.

3.4. Impulse responses

Given consistency of the individual VAR coefficient matrices, we now analyse the effects of ignored changes on impulse responses and on cumulated impulse responses.[7]

The true quantities of interest are in fact the impulse response matrices of $\boldsymbol{x}_t$, which are all based on the coefficient matrices of the MA representation of the VAR model, $\boldsymbol{x}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \boldsymbol{\varepsilon}_{t-j}$. We estimate them building on the VAR estimates for $\boldsymbol{y}_t$ when ignoring the changes in mean.

We consider two different situations. First, the identification scheme of the structural shocks $\boldsymbol{e}_t$ is known such that MA coefficients are simply multiplied with a suitable, known matrix $\mathbf{V}$, leading to

$$\boldsymbol{\Xi}_j = \frac{\partial \boldsymbol{x}_t}{\partial \boldsymbol{e}_{t-j}} = \mathbf{C}_j \mathbf{V}$$

and correspondingly to

$$\hat{\boldsymbol{\Xi}}_j = \hat{\mathbf{C}}_j \mathbf{V}. \tag{6}$$

In this case, it suffices to analyse the behaviour of the estimated MA coefficient matrices. Second, the matrix $\mathbf{V}$ may have to be estimated based on the residuals from the long autoregression, e.g. for a Cholesky-type identification scheme (but note that short-run or long-run identification approaches rely on the covariance matrix of the shocks as well), such that the behaviour of

$$\hat{\boldsymbol{\Xi}}_j = \hat{\mathbf{C}}_j \hat{\mathbf{V}}$$

is now of interest, with $\hat{\mathbf{V}}$ a continuous function of the residual covariance matrix $\frac{1}{T}\sum_{t=h_T+1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'$ and $\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{y}_t - \hat{\mathbf{A}}_1 \boldsymbol{y}_{t-1} - \cdots - \hat{\mathbf{A}}_{h_T} \boldsymbol{y}_{t-h_T}$. In this case we need to examine the behaviour of the residual covariance matrix estimator as well.

In both cases, the MA coefficients have to be estimated. They are given recursively as a function of the AR coefficient matrices as

$$\mathbf{C}_j = \sum_{k=0}^{j-1} \mathbf{A}_{j-k} \mathbf{C}_k \quad \text{with} \quad \mathbf{C}_0 = \mathbf{I}_K. \tag{7}$$

Recall that the vector MA [VMA] coefficient matrices have exponential decay.

Following Eq. (7), the estimated VMA coefficient matrices are obtained by numerical inversion of the estimated VAR coefficient matrices, defined recursively by

$$\hat{\mathbf{C}}_0 = \mathbf{I}_K \quad \text{and} \quad \hat{\mathbf{C}}_j = \sum_{k=0}^{j-1} \hat{\mathbf{A}}_{j-k} \hat{\mathbf{C}}_k \ \forall j > 0.$$

Building on them, the behaviour of impulse responses and cumulated impulse responses is given for a known identification scheme $\mathbf{V}$ in the following Proposition 3:

Proposition 3. *Under the assumptions of Proposition 1 and* $\sum_{j=1}^{p} \left\| \mathbf{A}_j \right\| < 1$, *we have for* $\hat{\boldsymbol{\Xi}}_j$ *from* (6)

1. $\max_{0 \le j \le \ell_T} \left\| \hat{\boldsymbol{\Xi}}_j - \boldsymbol{\Xi}_j \right\| = O_p\left(\frac{\ell_T}{h_T^{1/2}}\right)$ *and*
2. $\max_{0 \le j \le \ell_T} \left\| \sum_{k=0}^{j} \hat{\boldsymbol{\Xi}}_k - \sum_{k=0}^{j} \boldsymbol{\Xi}_k \right\| = O_p\left(\frac{\ell_T^2}{h_T^{1/2}}\right)$

for $\ell_T \to \infty$, *such that* $\ell_T \le h_T$.

Proof. See the Appendix.

We note that the bounds for convergence rates are "less tight" compared to Proposition 1. The reason behind the reduced convergence rates is that, due to the recursive nature of the inversion, the estimation errors in $\hat{\mathbf{A}}_k$, $k \le j$, cumulate in $\hat{\mathbf{C}}_j$ as j increases, so we need some upper bound for j and an assumption about the propagation of the estimation errors when computing $\hat{\mathbf{C}}_j$. This is particularly relevant when uniform consistency is required (note that pointwise consistency of $\hat{\boldsymbol{\Xi}}_j$ for some j does not suffice to conclude about convergence behaviour of the cumulated impulse responses when $h_T \to \infty$), say. The additional condition $\sum_{j=1}^{p} \left\| \mathbf{A}_j \right\| < 1$ serves precisely to control how estimation errors cumulate, and it is a plausible restriction for stable finite-order vector autoregressions.

To consider the case involving an estimate of the error covariance matrix, we may build on the previous proposition and only need to analyse the behaviour of the residual covariance matrix. It turns out that changes in the volatility play the most important role in this setup, while those in the mean are not essential: the problem lies not with estimation error in the residuals, but rather in the fact that the residual covariance matrix consistently estimates an "average" covariance of the reduced-form shocks, so any identification scheme based on this estimate is working only on average.

Proposition 4. *Under the assumptions of Proposition 1, it holds with* $\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{y}_t - \sum_{j=1}^{h_T} \hat{\mathbf{A}}_j \boldsymbol{y}_{t-j}$ *that* $\frac{1}{T}\sum_{t=h_T+1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' - \boldsymbol{\Omega} \xrightarrow{p} \mathbf{0}$.

[7] While this is related to the issue of providing confidence intervals and bands for impulse responses (see e.g. Brüggemann et al., 2016; Lütkepohl et al., 2015, and Inoue & Kilian, 2016, for recent contributions), it actually precedes the confidence issue: if consistency is not given, confidence bands cannot be meaningfully interpreted anyway.

Proof. See the Appendix.

It is seen that the residual covariance matrix does converge in spite of the omitted variable bias in $\hat{\mathbf{A}}_j$. The limit, however, is the average pathwise volatility of the process, which is different from the actual volatility at any time—except of course for the case where there is no variation in the unconditional covariance matrix of the reduced-form shocks. The residual covariance matrix therefore lacks identification power in general.

Remark 4. This lack of consistency has implications for interval forecasts, say, whenever forecast intervals rely on the residual covariance matrix. Estimating the wrong scale leads to under- or over-coverage of forecast intervals and fan charts.

On the other hand, a *localised* variant of the estimator does converge to the time-specific covariance matrix. And indeed, under possible time-varying volatility, one should ask the question: At which time is identification actually needed? Say a projection in the near future is required. Then the relevant quantity is the conditional error covariance matrix at the end of the sample, $\mathbf{H}(1)\mathbf{H}'(1)$, rather than the average $\boldsymbol{\Omega}$, and a localised estimator of the covariance matrix is the suitable tool.[8] This can be obtained either by putting more weight on observations close to the relevant time of the identification, or by specifying a (parametric) model for the variation of the covariance matrix. The nonparametric approach is more in the spirit of this paper, and is for instance strongly advocated by Patilea and Raïssi (2020); state-of-the-art methods for nonparametric variance estimation are given e.g. in Casas and Gijbels (2012).

Alternatively, one may exploit the robustness of long autoregressions to changes in the mean as discussed in Section 3 to provide estimates of $\boldsymbol{\Sigma}_t$. The key observation is that, given that the shocks $\boldsymbol{\varepsilon}_t$ have zero mean, it holds that $\mathbb{E}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right) = \boldsymbol{\Sigma}_t$, such that the sequence

$$\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' - \boldsymbol{\Sigma}_t\right) + \boldsymbol{\Sigma}_t$$

can be seen to be a sequence of uncorrelated variables with smoothly varying mean, whose one-step-ahead forecast at time $t-1$ is precisely the desired local covariance matrix $\boldsymbol{\Sigma}_t$.

So let $\boldsymbol{\omega}_t := \operatorname{vech}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)$, $\boldsymbol{\zeta}_t := \operatorname{vech}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' - E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)\right)$ and $\boldsymbol{s}_t = \operatorname{vech}\left(\boldsymbol{\Sigma}_t\right)$, and note that the process

$$\boldsymbol{\omega}_t := \boldsymbol{s}_t + \boldsymbol{\zeta}_t$$

is a particular case of the time-varying vector autoregressive models (1)–(2), where $p = 0$. Therefore, under the additional restriction that $\sup_{t\in\mathbb{Z}} \mathbb{E}\left(\|\boldsymbol{\epsilon}_t\|^8\right) < \infty$, Propositions 1 and 2 apply, and a long autoregression will deliver consistent approximations of the one-step-ahead forecast of $\boldsymbol{\omega}_t$ at time $t-1$, which is nothing else than $\boldsymbol{s}_t$. Then, a consistent estimate of $\boldsymbol{\Sigma}_t$ is easily recovered from $\hat{\boldsymbol{s}}_t$.[9]

This motivates the use of a long autoregression on $\hat{\boldsymbol{\omega}}_t = \operatorname{vech}(\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t')$ to obtain consistent estimates of $\boldsymbol{\Sigma}_t$. We note that, like in the discussion on model selection following Proposition 1, an information criterion would ensure that the fitted vector autoregression does have an order growing to infinity. We also note that if there is no time-varying volatility, an information criterion like the Bayesian information criterion (BIC) would pick the correct order $p = 0$ of this "autoregression in squares" in the limit, in which case the VAR-based forecast is nothing else than the constant mean of $\boldsymbol{\omega}_t$ as estimated by the half-vectorised residual sample covariance matrix.

Remark 5. There are some particular cases where such care is unnecessary, at least for impulse responses. Should the time variation be proportional for all elements of $\boldsymbol{x}_t$—i.e. $\mathbf{H}(s) = h(s)\mathbf{H}$, with $h(\cdot)$ a scalar function and $\mathbf{H}$ a constant full-rank matrix—$\boldsymbol{\Omega}$ is proportional to the conditional covariance matrix $\boldsymbol{\Sigma}_t$, so there is no need for action if one is interested in responses to standardised shocks. In other words, if all series are affected in the same way by changes in the volatility, there is no problem with identification in spite of heterogeneity, reflecting the fact that the error correlations (and thus the structure the SVAR is capturing) do not change, even if their variance does. This does not apply to interval forecasts, however.

4. Monte Carlo experiments

4.1. Design

We now study the effects of the ignored changes in mean and volatility in finite samples. The aim of this section is to illustrate the relevance of our limiting results in Section 3 from several different perspectives. In what follows we focus our attention on deterministic mean and volatility. First, we study how one-, six-, and 12-step-ahead forecast errors behave when different nonstationarity patterns are considered, illustrating Proposition 1. Second, we show how a break in the mean and the variance affects the impulse-response analysis, relating to our results of Proposition 3. We investigate the advantages and disadvantages of ignoring structural changes in the mean in comparison to explicitly modelling breaks in the VAR intercept. We use the popular break detection procedure of Bai and Perron (2003) (henceforth BP), which is often used for detecting and modelling structural changes in empirical applications.[10] Finally, we supplement our comparative analysis with random walk forecasts (RW, hereafter) as the standard benchmark commonly employed in time series analysis.

In all setups data are generated according to the bivariate VAR(1):

$$\boldsymbol{y}_t = \boldsymbol{m}_t + \boldsymbol{x}_t, \tag{8}$$

$$\boldsymbol{x}_t = \mathbf{A}\boldsymbol{x}_{t-1} + \boldsymbol{\varepsilon}_t \tag{9}$$

where an autoregressive coefficient matrix $\mathbf{A}$ is set to [0.4, 0.1; 0.2, 0.3]. We consider four cases:

[8] The same arguments apply for variance decompositions.

[9] Moreover, the long autoregression is potentially useful when the VAR shocks follow an ARCH-type model such that $\boldsymbol{\omega}_t = \operatorname{vech}\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)$ has a finite-order VAR representation. Allowing for ARCH effects requires additional technicalities in the proof of Proposition 1, however, and we leave this for further research.

[10] As of July 26, 2020, Google Scholar reports 4912 citations of Bai and Perron (2003).

1. **Break in the mean.** Components of the mean vector $\boldsymbol{m}_t = [m_{1,t}, m_{2,t}]$ are modelled with one break as

$$m_{i,t} = \begin{cases} -(1-\tau_i)\cdot Br_i \text{ for } t \le \tau_i T \\ \tau_i \cdot Br_i \text{ for } t > \tau_i T, \end{cases} \tag{10}$$

for $i = 1, 2$, with break fractions $\tau_1 = 0.55$ for the first equation and $\tau_2 = 0.45$ for the second. The magnitudes of the breaks $[Br_1, Br_2]$ are either $[0.2; 0.3]$ ("small") or $[2.4; 2.6]$ ("large"). The error term $\boldsymbol{\varepsilon}_t$ is generated as $i.i.d.\, N(0, \mathbf{I}_2)$.

2. **Smooth changes in the mean.** Assumption 1 allows for more than just sharp breaks, and we also consider smooth changes in the mean component given as

$$m_{1,t} = \gamma_1 \sin\left(\frac{2\pi t}{T}\right) \text{ and } m_{2,t} = -\gamma_2 \sin\left(\frac{2\pi t}{T}\right), \tag{11}$$

where parameters γ_1 and γ_2 control the magnitude of changes in the mean. We consider the two cases when $[\gamma_1, \gamma_2]$ takes values $[0.2; 0.3]$ and $[2.4; 2.6]$. The error term $\boldsymbol{\varepsilon}_t$ is generated as $i.i.d.\, N(0, \mathbf{I}_2)$.

3. **Break in the mean and the variance.** The vector $\boldsymbol{m}_t$ has the same structure as in (10), and $\boldsymbol{y}_t$ exhibits a jump in the variance in the middle of the sample, $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \mathbf{I}_2)$ for $t < T/2$ and $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, 2\mathbf{I}_2)$ for $t \ge T/2$.

4. **Smooth changes in the mean and a break in the variance.** The vector $\boldsymbol{m}_t$ has the same structure as in (11), and $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \mathbf{I}_2)$ for $t < T/2$ and $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, 2\mathbf{I}_2)$ for $t \ge T/2$.

We note that cases 3 and 4 exhibit breaks in the variance but not in the correlations, so impulse responses based on the residual covariance matrix should not be affected by the breaks; see Remark 5.

Fig. 1 illustrates a typical series generated by data generating processes (DGPs) (8)–(9) with a sharp break and smooth changes in the mean component generated by (10) and (11), respectively. For clarity of illustration, only $y_{1,t}$ is presented; $y_{2,t}$ exhibited the same (apparent) behaviour. While "large" changes in the mean, characterised by $[Br_1, Br_2] = [2.4; 2.6]$ or $[\gamma_1, \gamma_2] = [2.4; 2.6]$, are clearly visible, cases with "small" changes (i.e. $[Br_1, Br_2] = [0.2; 0.3]$ and $[\gamma_1, \gamma_2] = [0.2; 0.3]$) hardly differ from the constant mean scenario (based on visual inspection only). The later cases aim to mimic empirical scenarios when changes in the mean are often overlooked and ignored.

We analyse the performance of the long VAR (LVAR) with the number of lags set to $h_{max} = [4\,(T/100)^{0.25}]$, the long VAR with the lag length chosen by Akaike's information criterion (LVARa) with maximum order h_{max}, and the long VAR with the lag length chosen by the Bayesian information criterion (LVARb). For comparison, our analysis is supplemented with benchmark forecasts and impulse responses obtained from a VAR model accommodating for breaks in the mean detected by the BP procedure (BP-VAR). To be more precise, we use the sequential detection scheme of BP at a significance level of 5% to test for the presence and estimate the number of breaks in the mean component as well as their locations.[11] After that, the estimated structural changes in the mean (if detected) are incorporated into a VAR model. Both the AIC and BIC with maximum order h_{max} are used to select the number of lags in the VAR model. When applying Bai and Perron's methodology, we allow for different moment matrices of the regressors across segments, as well as heterogeneity and autocorrelation in the errors. The covariance matrix of the errors is allowed to be different across segments. Further, the maximum number of structural changes across simulations is set to two, with trimming parameter $\varepsilon = 0.15$.

In what follows we consider samples of size $T = \{100, 200, 300, 500, 1000\}$, where $T_0 = 100$ pre-sample observations were generated additionally and discarded. The results for each setup are based on 2000 Monte Carlo replications.

4.2. Multi-step-ahead forecasts

The forecast errors of the one-, six-, and 12-step-ahead forecasts obtained from the models described above are examined in the first set of simulations. We report detailed outcomes in Supplementary Material parts B.1 and B.3. Here, we summarise our main findings as follows:

1. The AIC and BIC lead to virtually identical results (for both LVAR and BP-VAR), so we only report the results for the AIC.
2. The main difference between the scenarios with and without a break in the variance is that the overall variance of the forecast errors is higher in the case with a break in variance. Both methods (for both LVAR and BP-VAR) outperform RW forecasts. Otherwise, there is no qualitative difference in the ranking of the LVAR and BP-VAR methods.
3. In all scenarios, when changes in the mean are "small" (see Suppl.Figs. 9, 11, 13, and 15), the performance of the considered methods is roughly the same. In other words, if a one-step-ahead forecast is the focus of interest, there is little, if any, advantage of the BP-VAR procedure over the long VAR model with ignored changes in the mean.
4. There is a different forecast outcome when two scenarios with "large" sharp breaks (Suppl. Figs. 10 and 14) and "large" smooth changes in the mean are compared (see Suppl. Figs. 12 and 16). That is, the BP-VAR procedure performance is marginally better when the DGP has a sharp break, but is outperformed by a wider margin when the break is smooth. This observation becomes more prominent as we move from one-step-ahead forecast to six or 12 steps ahead. Such an outcome is expected, since the BP methodology is developed precisely to account for sharp breaks in the mean.

[11] The Matlab code for the detection procedure used in our simulations was downloaded from Pierre Perron's website and developed by Yohei Yamamoto; see http://people.bu.edu/perron/. We thank them for making the code available.

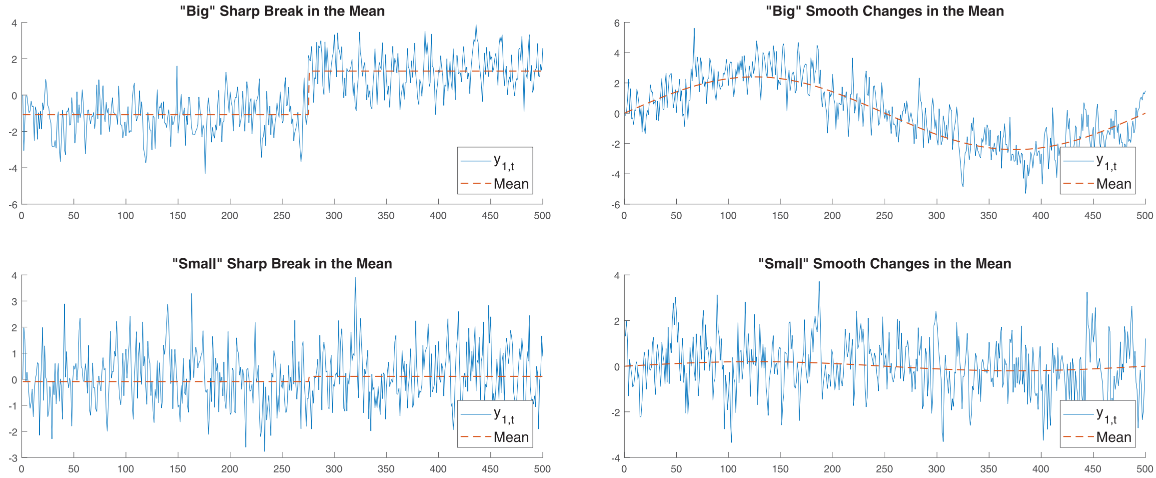

Fig. 1. Typical output from data generated processes (8)–(9) with $T = 500$ and different scenarios for the changes in the mean component. The solid line depicts the first series $y_{1,t}$ of the DGP, and the dotted line plots the corresponding mean component.

Summing up, the use of long VAR models in empirical applications can be justified to obtain one-step-ahead forecasts—theoretically and based on simulations. Using long VARs is particularly recommended when the nature of the changes in the mean is not known in advance. The BP-VAR strategy is only competitive when changes in the mean are abrupt and relatively easy to spot. Concretely, LVARs do not lose much, if anything, compared to BP-VAR when break detection does better (expectedly, this is the case when the structural changes in the mean are abrupt). At the same time, LVARs dominate the BP-VAR by a noticeable margin in the case where the changes are smooth.

4.3. Impulse responses

In the second set of simulations we investigate how impulse response sequences are affected by the ignored changes in the mean and the (co)variances. DGPs (8)–(9) are employed with the three scenarios described in Section 4.1 and the sample size is set to $T = 500$. The impulse-response analysis is based on the structural shock $\boldsymbol{\varepsilon}_t$: $\varepsilon_{1,1}$ is of size one standard error and $\varepsilon_{1,t} = 0$ for $t > 1$. Detailed results are presented in Supplementary Material part B.2. Our main findings are the following:

1. In all scenarios where the changes in the mean are small, there are neither quantitative nor qualitative differences in the performance of the models.
2. When there are breaks in the variance, the sampling variability of the estimated impulse responses increases, but their location is essentially the same as in the cases without breaks in the variance. This parallels the findings for one-step-ahead forecasts.
3. The LVAR model tends to capture the correct impulse response for the first few periods and provides biased results in the long-run if the changes in the mean are large. For small changes, LVARs perform well at longer horizons as well.
4. The BP procedure provides the best approximation of the true impulse response function when a break in the mean is sharp and large. Somewhat expectedly, however, it gives a biased output for the scenario with smooth changes in the mean.

5. Two illustrations

5.1. International dynamics of inflation

We first consider short-run inflation dynamics in the G7 countries and the euro area over the period from March 1973 to December 2018. Monthly CPI inflation data series are obtained from the OECD database.[12] Inflation rates are constructed as the first difference of the logarithm of the CPI multiplied by 100. Suppl. Fig. 25 plots the series; one notices changes in the mean as well as the volatility.

The presence of structural breaks in the inflation rate has been reported in recent literature (see e.g. Bataa et al., 2013 and Bataa et al., 2014). We confirm such findings by using the BP test procedure with up to five breaks and trimming parameter $\varepsilon = 0.15$. We allow for serial correlation in the errors as well as different variances of the residuals across segments. The sequential procedure of Bai and Perron (at the 5% significance level) selects at least one break in each series. Hence, we have a clear presence of the structural changes in the mean. The precise number of breaks found, as well as their location, are indicated in Fig. 2 by means of dotted lines. It is not apparent (from a visual examination of the series) whether the changes are sharp breaks or smoother in nature).

After establishing the presence of breaks in the mean component, we compare the forecasting performance of the various procedures. As in the Monte Carlo section, we consider two versions of a long VAR: the LVAR with the number of lags set to $h_{max} = [4\,(T/100)^{0.25}]$, and the

12 Downloaded from http://www.oecd.org on 05/02/2019.

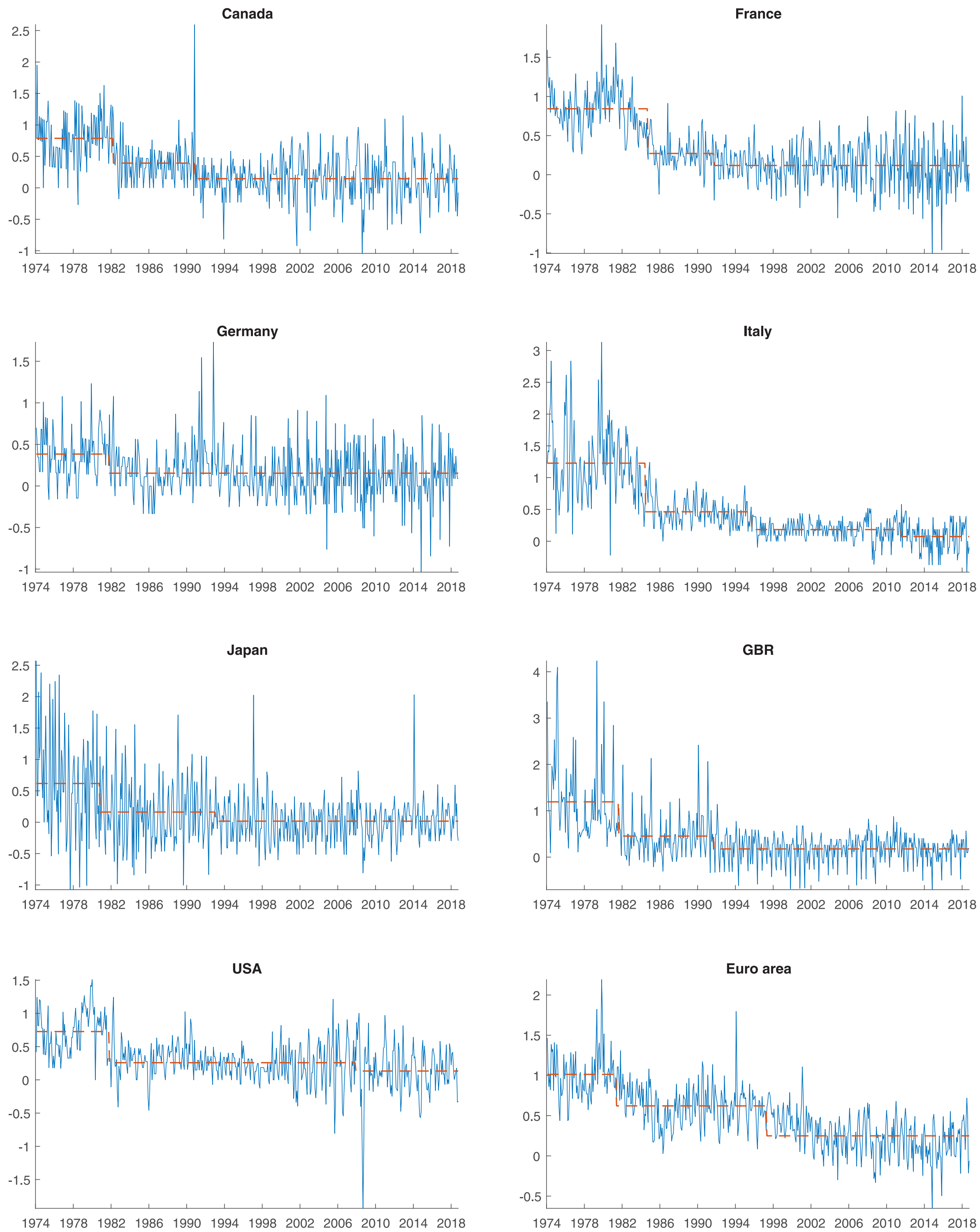

Fig. 2. Inflation series in the G7 and euro area. The figure shows observed inflation (solid line) and changes in the mean component (dotted line) obtained via the sequential BP procedure at the 5% significance level.

LVAR with the AIC (all ignoring the changes in the mean). To explicitly account for structural changes, we employ BP's methodology for each series. The sample is split into two subsamples where the first one is used for estimation and the second one is reserved for a comparison of the one-step-ahead forecasts. Two cases are studied: first, the subsample reserved for estimation ends in January 2000 right before the dot-com bubble, and, second, the

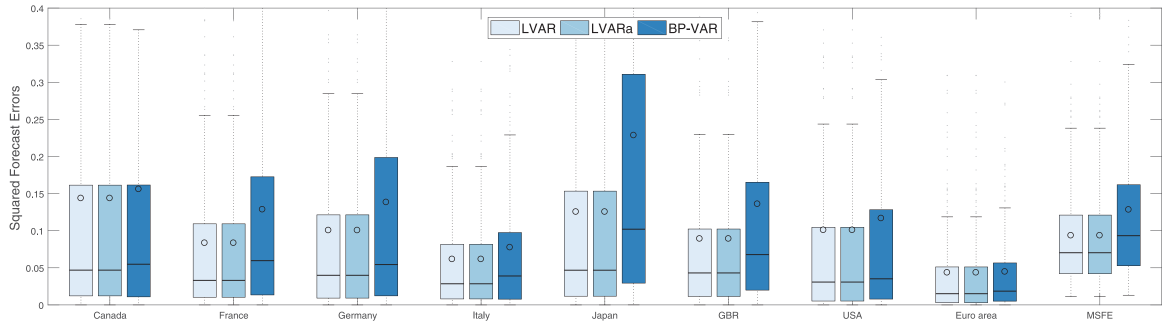

Fig. 3. Forecast performance of different methods where data until Jan 2000 are used for estimation, and after for forecast comparison. The presented boxplots report the squared forecast errors of LVAR, LVAR with AIC, truncated LVAR, and the BP procedure in each considered country, the euro area, and the joint mean square forecast.

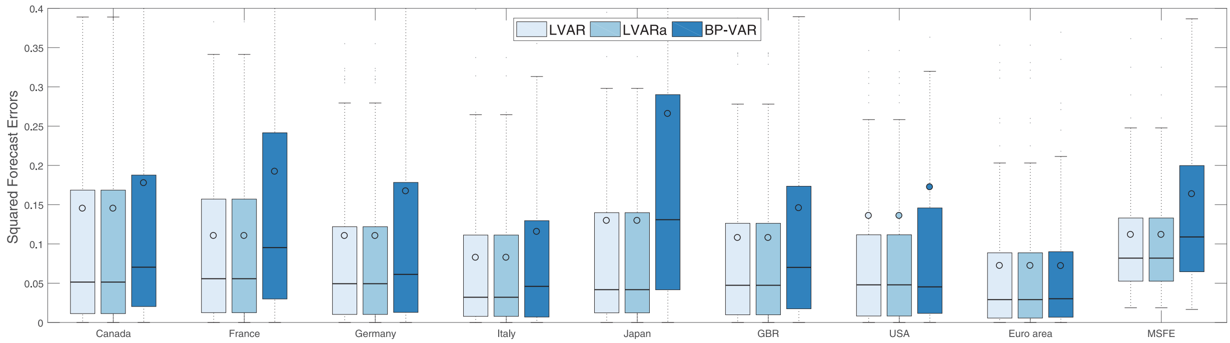

Fig. 4. Forecast performance of different methods where data until Jan 2007 are used for estimation, and after for forecast comparison. The presented boxplots report the squared forecast errors of LVAR, LVAR with AIC, truncated LVAR, and the BP procedure in each considered country, the euro area, and the joint mean square forecast.

estimation subsample ends in January 2007 before the global financial crisis.

The outcomes of the forecasting comparison are reported in the form of boxplots of the squared forecast errors in Fig. 3 and Fig. 4. For all countries except Canada and the euro area, the VAR model with the BP procedure provides clearly less precise one-step-ahead forecasts. Furthermore, the overall performance of long VAR models (presented in the last column of Fig. 3 and Fig. 4 in terms of mean squared forecast errors) is better. Although there are structural changes in the mean component for the examined series, ignoring these changes outperforms modelling them when it comes to one-step-ahead forecasts using long VARs.

In the final part of this analysis, we study how ignored structural breaks in inflation series affect impulse-response analysis. As above, our benchmark model—the standard recursively identified VAR (different ordering was considered but provided similar final outcomes) with the BP procedure to account for changes in the mean—is compared with different versions of LVAR. We use all available data for the estimation purpose.

Fig. 5 shows impulse responses of the euro area inflation rate series to one standard deviation innovation in inflation series of different countries (i.e. Canada, France, Germany, Italy, Japan, GBR, the USA, and the euro area). Note that the LVAR with AIC provides identical results to the LVAR, since the selected number of lags for both models is equal. As can be seen from the figure, in all cases LVARs and VAR with the BP procedure provide very similar results qualitatively and quantitatively, except in the case of Italy. This outcome for Italy is expected, since its inflation series contains the largest break (mid-1980s) of all series. Consequently, its magnitude has a stronger influence on the LVAR results. If we remove this particular break, IRFs obtained from both approaches become similar (see Fig. 6 for these additional results) illustrating the point that obvious breaks could be accounted for.

The case of Italy provides a nice illustration on how the methodology developed here should be used in practice, and corroborates our findings from Monte Carlo simulations. If the data contain structural changes in the mean that are clearly visible (or even known) without inspecting the data with statistical procedures, it is advisable to account for them. If the data set contains structural changes that have “small” or “medium” magnitude breaks, or if the breaks are of an unknown nature (i.e. whenever they are subject to uncertainty and require additional statistical tools for analysis) those breaks can be ignored when forecasting, as well as when conducting impulse-response analysis.

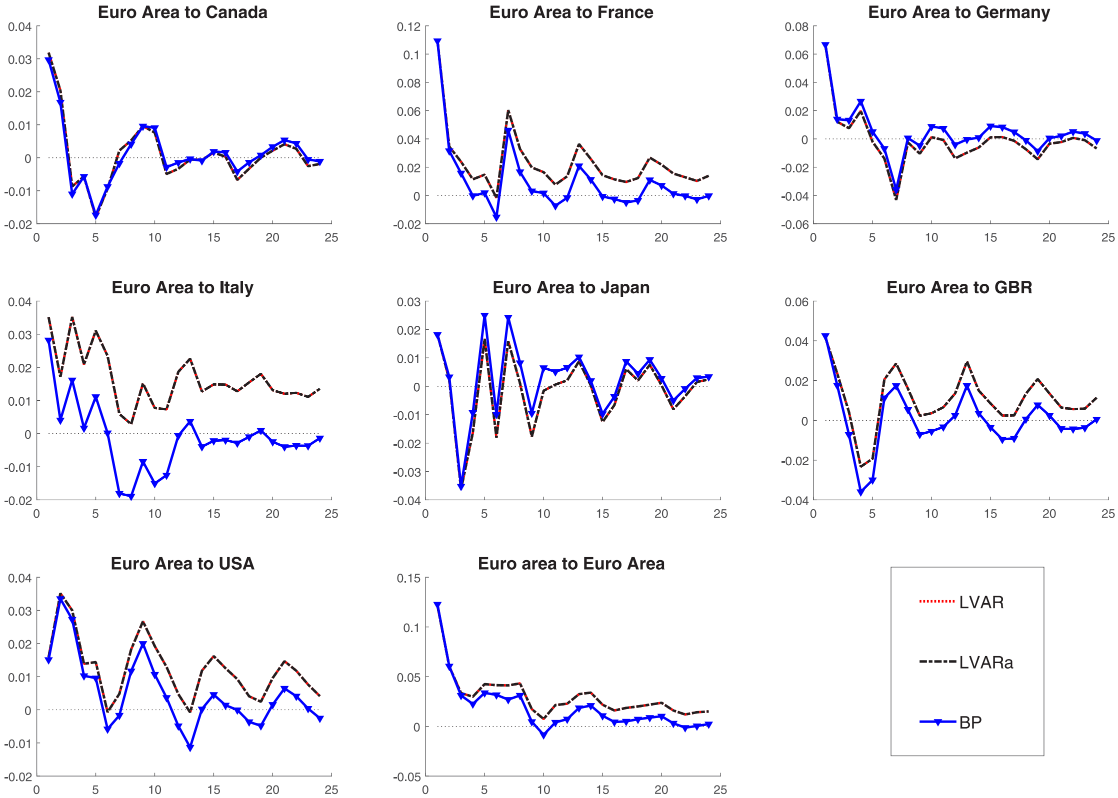

Fig. 5. Impulse responses of the euro area inflation rate to innovations in the inflation rates of different countries: Canada, France, Germany, Italy Japan, GBR, and the USA. The sample period is from 3/1973–12/2018.

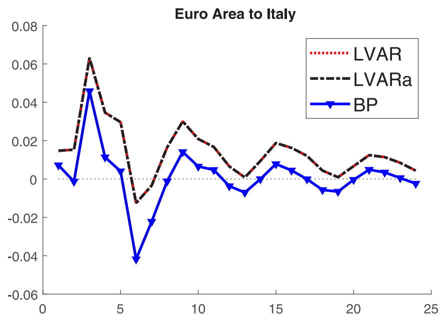

Fig. 6. Impulse responses of the euro area inflation rate to innovation in the inflation rate of Italy. The sample period is from 1/1990–12/2018.

5.2. Uncertainty and economic activity

We now revisit the impulse-response analysis presented in the seminal work of Bachmann et al. (2013) on uncertainty and economic activity. Among other questions, Bachmann et al. (2013) analysed the impulse responses of different measures of economic activity to uncertainty measures in Germany and the U.S. The case of Germany is of particular interest, since a natural structural change occurred after Germany's reunification, for which the authors accounted in their analysis. Hence, in what follows we investigate how ignoring this change with the use of LVARs or using the BP procedure will affect impulse-response outcomes.

The original data from (Bachmann et al., 2013) are used, where the sample has monthly frequency and runs from January 1980 to December 2010.[13] Three different measures of economic activity are considered (all in logs): manufacturing production, manufacturing employment, and average hours worked in manufacturing. As the uncertainty proxy variable, we use the survey-based measure FDISP, developed using the German IFO Business Climate Survey in Bachmann et al. (2013). See Bachmann et al. (2013) for more details.

The benchmark impulse-response functions (IRFs) are obtained from bivariate VAR models featuring the measure of uncertainty FDISP and one of the selected measures of economic activity. The uncertainty series is ordered first in a recursive identification (similar results were obtained with alternative ordering). All bivariate models in the benchmark case include an exogenous dummy variable to account for Germany's reunification in October 1990. We compare IRFs from the benchmark case with results obtained from LVAR and LVAR with AIC, where Germany's reunification is ignored. For completeness, we also perform the BP procedure within the VAR framework to estimate potential structural breaks instead of imposing them directly. Finally, as in Bachmann et al.

13 The data and Matlab code were downloaded from https://www.aeaweb.org/articles?id=10.1257/mac.5.2.217.

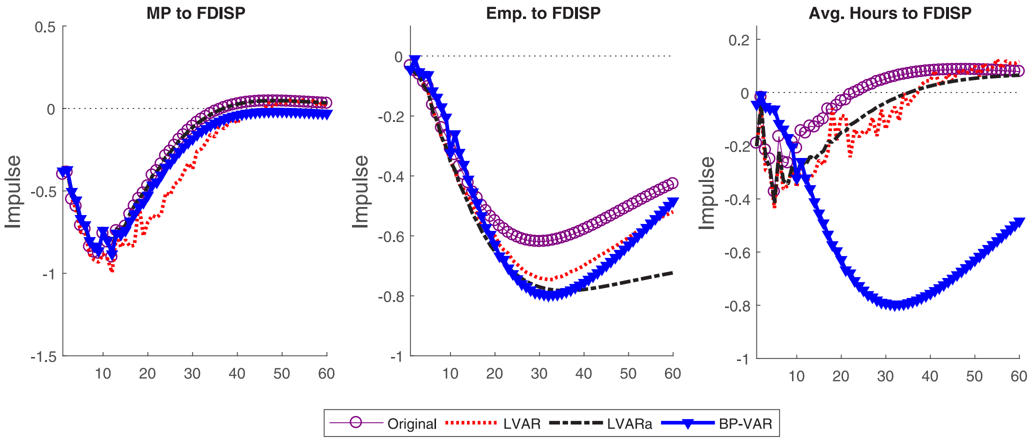

Fig. 7. Impulse responses of manufacturing production (MP), manufacturing employment (Emp.), and manufacturing average hours worked (Avg. Hours) to innovations in uncertainty series. The results are obtained from separately estimating bivariate LVAR models. LVARa denotes an outcome of the long VAR with the AIC. The sample period is common for all models and runs from 1/1980–12/2010.

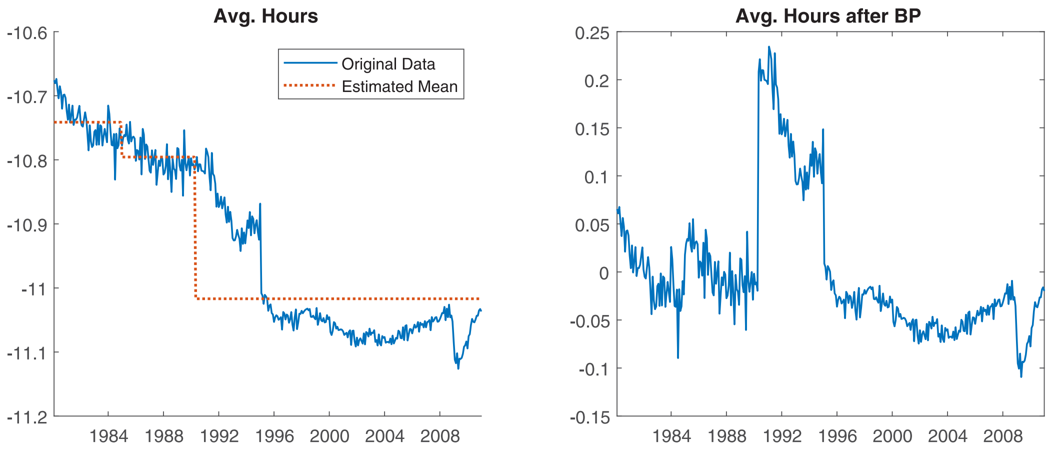

Fig. 8. Illustration of the BP procedure outcomes for average hours worked in manufacturing. The left panel shows the original data (solid line) together with the estimated changes in the mean component (dotted line) obtained via the BP procedure (at the 5% significance level). The right panel illustrates average hours series after the estimated mean component (together with changes) is extracted from the data.

(2013), we set the minimum number of lags to 12 in all models, as this appear to be the minimum necessary to adequately describe dynamics in such bivariate systems with monthly data. In the case of LVAR models and the BP procedure, we allow the number of lags to run from 12 to 24, chosen by the AIC. Results using the BIC provided no quantitative difference and are omitted.

As can be seen from Fig. 7, the selected economic activity reacts quantitatively and qualitatively in a similar way to innovations in uncertainty series if reunification in Germany is ignored and instead LVAR models are used in all three cases. Hence, the main message from the impulse-response analysis in Bachmann et al. (2013) is also confirmed when a known break (Germany's reunification) is ignored. In the case of using the BP procedure, similar outcomes are observed if manufacturing production and manufacturing employment are chosen to describe an economic activity. However, a large deviation from the original IRFs is observed in the case of average worked hours. The reason for such a departure is presented in Fig. 8. It is evident that in this case the BP procedure should not be used, since it is tailored to capture sharp breaks in the mean component. As can be seen from Fig. 8, the break associated with Germany's reunification in average worked hours series is not an especially sharp one. It starts in 1990 (which the BP procedure picks up) and ends in 1995. As the outcome, the series obtained after the BP procedure contains two sharp breaks (see the right panel of Fig. 8) that distort the final IRFs.

6. Summary and recommendations

We studied the estimation of autoregressive parameters, point forecasts, and impulse responses in (structural) VAR models with ignored changes in the mean and the (co)variance. Based on theoretical results, numerical simulations, and applications to inflation dynamics, as well as to uncertainty and economic activity, we find the following:

(i) Long vector autoregression provides a robust and competitive estimation method for autoregressive parameter matrices if changes in mean and (co)variance are ignored. This reinforces practical approaches using lag lengths that are high enough in applications.

(ii) When it comes to H-step-ahead forecasts, benefits from explicitly modelling possible breaks are only visible if one is certain of the location and abrupt nature of the breaks. Otherwise, long VARs are competitive in terms of the forecast MSE.

(iii) Regarding the impulse-response analysis, the long VAR approach dominates an explicit modelling of possible breaks whenever changes are not abrupt (both for forecasting and for impulse response estimation) or not certain. Should the changes in the mean be known to be abrupt, a break detection procedure followed by explicit modelling of the break may sometimes, but not always, outperform the long VAR approach.

(iv) One should treat outcomes of the impulse-response analysis with care if there is any evidence of heteroskedasticity (irrespective of whether changes in the mean are ignored or modelled explicitly). In such cases, sample information on the error covariance matrix should not be used for the identification of structural shocks, with the exception of the particular case where changes in the covariance matrix do not affect correlations and are the same for all series. Forecast intervals are affected in a similar manner by time-varying volatility. To mitigate the effects of time-varying volatility, one should either employ a model allowing for time-varying volatility (such as a time-varying parameter VAR) or resort to a local covariance matrix estimator. Localisation may be achieved via kernel-based estimation or by means of a long vector autoregression in the cross-products of the VAR residuals.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2023.06.002.

References

Amado, C., & Teräsvirta, T. (2014). Modelling changes in the unconditional variance of long stock return series. *Journal of Empirical Finance*, *25*(1), 15–35.

Antolín-Díaz, J., & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs. *American Economic Review*, *108*(10), 2802–2829.

Bachmann, R., Elstner, S., & Sims, E. R. (2013). Uncertainty and economic activity: Evidence from business survey data. *American Economic Journal: Macroeconomics*, *5*(2).

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*(1), 1–22.

Bataa, E., Osborn, D. R., Sensier, M., & van Dijk, D. (2013). Structural breaks in the international dynamics of inflation. *The Review of Economics and Statistics*, *95*(2), 646–659.

Bataa, E., Osborn, D. R., Sensier, M., & van Dijk, D. (2014). Identifying changes in mean, seasonality, persistence and volatility for G7 and Euro Area inflation. *Oxford Bulletin of Economics and Statistics*, *76*(3), 360–388.

Baumeister, C., & Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review*, *109*(5), 1873–1910.

Benati, L. (2008). The "great moderation" in the United Kingdom. *Journal of Money, Credit and Banking*, *40*(1), 121–147.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics*, *2*(3), 489–502.

Bhansali, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain. *The Annals of Statistics*, *6*(1), 224–231.

Bhattacharya, R. N., Gupta, V. K., & Waymire, E. (1983). The Hurst effect under trends. *Journal of Applied Probability*, *20*(3), 649–662.

Blanchard, O. J., & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, *79*(4), 655–673.

Bognanni, M. (2018). *A class of time-varying parameter structural VARs for inference under exact or set identification*: *Working paper*, Federal Reserve Bank of Cleveland.

Brüggemann, R., Jentsch, C., & Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, *191*(1), 69–85.

Carstensen, K., & Salzmann, L. (2017). The G7 business cycle in a globalized world. *Journal of International Money and Finance*, *73*, 134–161.

Casas, I., & Gijbels, I. (2012). Unstable volatility: The break-preserving local linear estimator. *Journal of Nonparametric Statistics*, *24*(4), 883–904.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, *113*(1), 1–45.

Clark, T. E. (2009). Is the Great Moderation over? An empirical analysis. *Economic Review*, *4*, 5–42.

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*(2), 262–302.

Demetrescu, M., & Hassler, U. (2016). (When) do long autoregressions account for neglected changes in parameters? *Economic Theory*, *32*(6), 1317–1348.

Engle, R. F., & Kozicki, S. (1993). Testing for common features. *Journal of Business & Economic Statistics*, *11*(4), 369–380.

Gambetti, L., Pappa, E., & Canova, F. (2008). The structural dynamics of US output and inflation: What explains the changes? *Journal of Money, Credit and Banking*, *40*(2–3), 369–388.

Gonçalves, S., & Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, *123*(1), 89–120.

Gonçalves, S., & Kilian, L. (2007). Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. *Econometric Reviews*, *26*(6), 609–641.

Guidolin, M., & Timmermann, A. (2006). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics*, *21*(1), 1–22.

Inoue, A., & Kilian, L. (2016). Joint confidence sets for structural impulse responses. *Journal of Econometrics*, *192*(2), 421–432.

Jentsch, C., & Lunsford, K. G. (2019). The dynamic effects of personal and corporate income tax changes in the United States: Comment. *American Economic Review*, *109*(7), 2655–2678.

Kilian, L. (2013). Structural vector autoregressions. In N. Hashimzade, & M. A. Thornton (Eds.), *Handbook of research methods and applications in empirical macroeconomics* (pp. 515–554).

Kilian, L., & Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.
Lanne, M., Lütkepohl, H., & Maciejowska, K. (2010). Structural vector autoregressions with Markov switching. *Journal of Economic Dynamics & Control*, *34*(2), 121–131.
Lanne, M., Meitz, M., & Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics*, *196*(2), 288–304.
Lewis, D. J. (2021). Identifying Shocks via Time-Varying Volatility. *Review of Economic Studies*, *88*(6), 3086–3124.
Lewis, R., & Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis*, *16*(3), 393–411.
Lütkepohl, H., & Netšunajev, A. (2017). Structural vector autoregressions with smooth transition in variances. *Journal of Economic Dynamics & Control*, *84*, 43–57.
Lütkepohl, H., & Schlaak, T. (2018). Choosing between different time-varying volatility models for structural vector autoregressive analysis. *Oxford Bulletin of Economics and Statistics*, *80*(4), 715–735.
Lütkepohl, H., Staszewska-Bystrova, A., & Winker, P. (2015). Comparison of methods for constructing joint confidence bands for impulse response functions. *International Journal of Forecasting*, *31*(3), 782–798.
Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review*, *103*(4), 1212–1247.
Milunovich, G., & Yang, M. (2013). On identifying structural VAR models via ARCH effects. *Journal of Time Series Econometrics*, *5*(2), 117–131.
Ng, S., & Vogelsang, T. J. (2002). Analysis of vector autoregressions in the presence of shifts in mean. *Econometric Reviews*, *21*(3), 353–381.
Patilea, V., & Raïssi, H. (2012). Adaptive estimation of vector autoregressive models with time-varying variance: Application to testing linear causality in mean. *Journal of Statistical Planning and Inference*, *142*(11), 2891–2912.
Patilea, V., & Raïssi, H. (2020). Orthogonal impulse response analysis in presence of time-varying covariance. ArXiv preprint arXiv:2003.11188.
Phillips, P. C. B., & Xu, K. L. (2006). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis*, *27*(2), 289–308.
Poskitt, D. S. (2007). Autoregressive approximation in nonstandard situations: The fractionally integrated and non-invertible cases. *Annals of the Institute of Statistical Mathematics*, *59*(4), 697–725.
Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*(3), 821–852.
Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics*, *85*(4), 777–792.
Sims, C. A., & Zha, T. (2006). Were there regime switches in U.S. monetary policy? *American Economic Review*, *96*(1), 54–81.
Stock, J. H., & Watson, M. W. (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual*, *17*(1), 159–218.
Stock, J. H., & Watson, M. W. (2005). Understanding changes in international business cycle dynamics. *Journal of the European Economic Association*, *3*(5), 968–1006.
Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica*, *65*(1), 59–73.
Wang, C. S.-H., Bauwens, L., & Hsiao, C. (2013). Forecasting a long memory process subject to structural breaks. *Journal of Econometrics*, *177*(2), 171–184.
Xu, K.-L., & Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics*, *142*(1), 265–280.