Università Commerciale "Luigi Bocconi"
PhD School

PhD program in Economics and Finance
Cycle: 32
Disciplinary Field (code): SECS-P/05

Essays in Econometric Analysis

Advisor: Massimiliano Marcellino

PhD Thesis by
Yu Bai
ID number: 3029133

Academic Year: 2022

Contents

List of Figures

List of Tables

Acknowledgements

I am deeply indebted to my supervisor, Massimiliano Marcellino, for his invaluable guidance, support and encouragement, without whom I could probably not be able to reach the current stage in which I can hardly imagine a few years ago. I am also grateful to Andrea Carriero and George Kapetanios for their hospitality to arrange my visiting at Queen Mary, University of London and King's College London during the 2019/2020 academic year, and to Todd E. Clark for his kind and generous help in Chapter 1 of the thesis. I also thank professional services staff at Bocconi University, for their patient and outstanding administrative support.

Abstract

The thesis consists of three chapters on econometrics analysis, both theoretical and applied.

In the first chapter, which is coauthored with Andrea Carriero, Todd Clark and Massimiliano Marcellino, we propose a hierarchical shrinkage approach for multi-country VAR models. To make the approach operational, we consider three different scale mixtures of Normals priors — specifically, Horseshoe, Normal-Gamma, and Normal-Gamma-Gamma priors. We provide new theoretical results for the Normal-Gamma prior. Empirically, we use a quarterly data set for G7 economies to examine how model specifications and prior choices affect the forecasting performance for GDP growth, inflation, and a short-term interest rate. We find that hierarchical shrinkage, particularly as implemented with the Horseshoe prior, is very useful in forecasting inflation. It also has the best density forecast performance for output growth and the interest rate. Adding foreign information yields benefits, as multi-country models generally improve on the forecast accuracy of single-country models.

In the second chapter, which is coauthored with George Kapetanios and Massimiliano Marcellino, we consider kernel-based non-parametric estimation and inferential theory for large heterogeneous panel data models with stochastic time-varying coefficients. We propose mean group and pooled estimators, derive asymptotic distributions and show the uniform consistency and asymptotic normality of path coefficients. Then, we extend the procedures to the case with possibly endogenous regressors and propose a time-varying version of the Hausman exogeneity test. The finite sample performance of the proposed estimators is investigated through a Monte Carlo study and an empirical application on multi-country Phillips curve with time-varying parameters.

In the third chapter, I develop time-varying continuously updated GMM estimation and inferential theory for models whose parameters vary stochastically and smoothly over time. Then, I propose two structural stability tests in this context. After deriving the asymptotic properties of the estimators and test statistics, I assess their finite sample performance by an extensive Monte-Carlo study and illustrate their application by an empirical example on dynamic asset pricing models with stochastic discount factor (SDF) representation.

Chapter 1

Macroeconomic Forecasting in a Multi-country Context

[1]

1.1 Introduction

Since the seminal studies by Doan et al. [1984] and Litterman [1986], Bayesian vector autoregressions (VARs) have become workhorse models in macroeconomic forecasting. Reduced-form VARs are richly parameterized, which brings the risk of overfitting the data and large uncertainty for the future path projected by the model. It is well known that shrinkage generally improves forecasting performance, and Bayesian methods offer an effective way to shrink parameters by using prior information.

Due to increasing international trade and financial flows in recent decades, individual countries are more and more interlinked, which may make it helpful to use multi-country forecasting models and methods. The VAR literature includes three main approaches: (1) factor-augmented VAR models; (2) global vector autoregressive (GVAR) models; and (3) multi-country VARs. In factor-augmented VAR models, each country-specific VAR is augmented with "foreign variables," constructed by using principal components to extract common factors from all variables in foreign countries. In GVARs, used in studies such as Pesaran et al. [2009], Cuaresma et al. [2016], Huber [2016], and Dovern et al. [2016], each country-specific VAR is augmented with weakly exogenous "foreign variables," constructed by aggregating other countries' variables with international trade flows as weights. Then, country-specific models are combined to form a global model for the forecasting exercise. In multi-country VARs, used in studies such as Canova et al. [2007], Giannone et al. [2009], Korobilis [2016], Dées and Güntner [2017], and Koop and Korobilis [2019], variables for multiple countries are jointly modeled, with various degrees of cross-country interactions. Shrinkage is imposed to deal with the curse of dimensionality and is performed either by considering

[1] This is a joint work with Andrea Carriero, Todd Clark and Massimiliano Marcellino.

the panel dimension in the data or by simply treating the multi-country model as a large-scale BVAR and specifying priors on model coefficients.

While conventional Minnesota-type priors are shown to be useful in macroeconomic forecasting and are still widely used in the literature, other work suggests instead applying scale mixtures of Normals priors or other alternatives on single-country BVARs. These prior specifications have advantages with respect to the Minnesota-type prior, since they involve less hyperparameter tuning. They are also computationally more efficient than spike-and-slab priors while enjoying similarly nice theoretical properties at the same time. Huber and Feldkircher [2019] propose applying the Normal-Gamma prior, originally introduced by Griffin and Brown [2010], to BVARs and show that it is beneficial for macroeconomic forecasting. Follett and Yu [2019] use the Horseshoe prior, popular in the statistical literature (Carvalho et al. [2010]), and find that it improves forecast accuracy in a single-country context. Cadonna et al. [2020] propose a more general Normal-Gamma-Gamma prior, originated in Griffin and Brown [2017], for time-varying parameter BVARs and find that it delivers more sparse parameter estimates (but do not address forecast performance). Recently, there has been some limited interest in applying these types of priors in multi-country VARs. Korobilis [2016] proposes the stochastic search specification selection prior, which is a modification of the stochastic search variable selection prior proposed by George et al. [2008] applicable under some model restrictions. Korobilis finds that, to design priors for better forecasting performance, it is important to consider the panel structure in the data. However, the prior specifications considered there introduce dependence across equations, which makes efficient estimation, such as the approach of Carriero et al. [2019], difficult to apply.

In this paper, we propose and examine the use of hierarchical shrinkage approaches in multi-country VARs used for macroeconomic forecasting. To make the approach operational, we use three different scale mixtures of Normals priors that have been shown to be successful in single-country BVARs but have not been examined in multi-country models. These priors include the Horseshoe, Normal-Gamma, and Normal-Gamma-Gamma specifications. The hierarchical shrinkage is able to handle restrictions suggested in Canova and Ciccarelli [2013] for multi-country VARs. It is shown to be computationally more efficient than the existing stochastic search specification selection prior and also delivers better forecasting performance than the existing alternatives. We also provide some novel theoretical results for the Normal-Gamma prior.

Empirically, we work with a quarterly Group of Seven (G7) data set to examine the (point and density) forecasting ability of the new priors for three key macroeconomic variables: output growth, inflation, and a short-term interest rate. We also compare forecasting accuracy across various models with other specification choices. These models include: (1) country-specific VARs, either with Minnesota-type priors or hierarchical shrinkage proposed by Chan [2021]; (2) country-specific factor-augmented VARs; (3) GVARs; and (4) multi-country VARs in which shrinkage is performed either by imposing a particular hierarchical factor structure on the model parameters or by using priors. Because stochastic volatility (SV) has been found to be widely useful in macroeconomic forecasting with single-country models and also improves performance in our results, we include SV in all of our model specifications. In addition, since Cross et al. [2020] have raised some questions on the usefulness of scale mixtures of Normals priors in single-country macroeconomic forecasting, we consider alternative hierarchical shrinkage approaches for a robustness check (the appendix includes these results in the multi-country context).

Our results show that hierarchical shrinkage of multi-country VARs, particularly as implemented with the Horseshoe prior, improves macroeconomic forecast accuracy. It has outright advantages for inflation forecasting, and the Horseshoe specification of a multi-country VAR also performs best in density forecasts of output growth and the interest rate.[2] In point forecast accuracy, the Normal-Gamma prior performs best for output growth, whereas Canova and Ciccarelli [2009]'s factor shrinkage approach of multi-country models performs best for the interest rate. These results indicate that, although the Normal-Gamma-Gamma prior is more flexible than the Horseshoe prior and serves as a heavy-tailed extension of the Normal-Gamma prior, these advantages do not yield consistently better forecasting performance. We also find that modeling cross-country interactions achieves gains, as multi-country models generally outperform single-country models. As is common in the literature, we also find that models' forecasting performance varies over both countries and time. There are countries in which alternative models and priors do better than our hierarchical shrinkage of a multi-country VAR implemented with the Horseshoe prior, but there are no consistent patterns in which one of the alternatives is clearly better. In the interest of brevity, the paper's appendix provides results confirming that stochastic volatility is one important feature to improve forecast accuracy in the multi-country context.

The paper is structured as follows. Section 2 briefly introduces multi-country VAR models, existing prior specifications, and their challenges. Section 3 provides our new hierarchical shrinkage approach for multi-country VARs, the prior specifications, and some new theoretical results. Section 4 gives a brief summary of estimation algorithms and highlights some computational comparisons. Section 5 describes the data, forecasting metrics, and design of our forecasting exercise. Section 6 presents the main empirical results. Section 7 concludes. Technical details, sampling algorithms for various models, and additional empirical results are provided in the appendix.

1.2 Multi-country VARs

1.2.1 The model

The multi-country VAR model we consider has the form

$$y_{i,t} = c_i + B_i(L)Y_{t-1} + u_{i,t}, \tag{1.1}$$

[2] Koop and Korobilis [2019] and Feldkircher et al. [2021] also find that multi-country VARs are more beneficial for inflation forecasting.

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$; $y_{i,t}$ is a $G \times 1$ vector of variables for each country i, and $Y_t = (y'_{1,t}, \ldots, y'_{N,t})'$; c_i is a $G \times 1$ vector of constant terms for each i, $B_i(L) = \sum_{\ell=1}^{p} B_{i,\ell} L^{\ell}$, where $B_{i,\ell}$ are $G \times NG$ coefficient matrices associated with lag ℓ, $\ell = 1, ..., p$; and $u_{i,t}$ is a $G \times 1$ vector of disturbances. The lag length is assumed to be p. Combining equations across countries, the VAR can be written in matrix form as:

$$Y_t = c + \sum_{\ell=1}^{p} B_{\ell} Y_{t-\ell} + u_t, \tag{1.2}$$

where $c = (c'_1, c'_2, \ldots, c'_N)'$, each B_ℓ has dimension $NG \times NG$, and $u_t = (u'_{1,t}, \ldots, u'_{N,t})'$. For the stochastic volatility (SV) specification, we assume that

$$u_t = A^{-1} H_t^{0.5} \epsilon_t, \epsilon_{i,t} \sim i.i.d.\ N(0, I_{NG}),$$

where A^{-1} is a lower triangular matrix with diagonal elements equal to 1, and H_t is diagonal with generic j-th element $h_{j,t}$ evolving as a random walk (RW):[3]

$$\ln h_{j,t} = \ln h_{j,t-1} + e_{j,t}, j = 1, \ldots, NG, \tag{1.3}$$

where $e_t = (e_{1,t}, e_{2,t}, \ldots, e_{NG,t})'$ and $e_t \sim N(0, \Phi)$ with a full covariance matrix Φ as in Primiceri [2005]. The reduced-form error covariance matrix is $\Sigma_t = A^{-1} H_t A^{-1\prime}$.

While some work has examined time-varying coefficient VAR models (see, e.g., Cogley and Sargent [2005]; Primiceri [2005]; Koop et al. [2009]; and D'Agostino et al. [2013]), we restrict attention to constant coefficient VAR models with stochastic volatility for two reasons. First, time-varying coefficient VAR models are rarely used with more than 4-5 variables. This is mainly due to computational complexity and makes recursive forecasting with MCMC methods computationally infeasible. Second, in a forecasting context, reaching parsimony (in terms of both controlling time variation and getting rid of irrelevant regressors) in large models with time-varying coefficients remains a challenging task.[4]

1.2.2 Existing priors

The specification in (1.2) can incorporate complex dynamic structures for each variable in different countries. However, it also suffers from the curse of dimensionality due to the high dimensionality of the

[3]The RW specification may come at the cost of generating excessively thick forecast densities. Alternatively, the SV process (1.3) could be specified as an $AR(1)$ process, and $e_{ij,t}$ could be assumed to be t-distributed to incorporate fat tails. However, Clark and Ravazzolo [2015] find that these alternative specifications fail to dominate a baseline RW specification.

[4]To have parameter time variation in large VARs, Koop and Korobilis [2019] introduce forgetting factors, and Kapetanios et al. [2019b] use non-parametric methods combined with stochastic coefficient constraints. Yet, both methods become computationally infeasible if combined with the commonly used stochastic volatility specification. Gefang et al. [2019] develop variational Bayes methods (which utilize approximations of conventional posteriors) that permit large models with time-varying parameters and volatilities. Empirically, since forecasts are computed recursively, we implicitly consider the potential time variation of parameters in the model.

parameter space. For instance, in the forecasting exercises we use data on 3 dependent variables ($G = 3$) for the G7 countries ($N = 7$) and four lags ($p = 4$). A multi-country VAR with such choices would have 1,785 VAR coefficients. Thus, shrinkage is desirable.

In a single-country framework, the macro VAR literature generally relies on Bayesian shrinkage by imposing a Minnesota-type prior (Litterman [1986]) on the VAR coefficients. Applied analogously in our multi-country setup, the prior for B is $\text{vec}(B) \sim N(\text{vec}(\underline{\mu}_B), \underline{\Omega}_B)$, and $\underline{\Omega}_B$ is set to

$$\text{Var}(B_\ell^{(ii)}) = \frac{\lambda_1}{\ell^{\lambda_3}}, \quad \ell = 1, \ldots, p \tag{1.4}$$

$$\text{Var}(B_\ell^{(ij)}) = \frac{\lambda_2}{\ell^{\lambda_3}} \frac{\sigma_i^2}{\sigma_j^2}, \quad \forall i \neq j, \ell = 1, \ldots, p, \tag{1.5}$$

where $B_\ell^{(ij)}$ denotes the element in row i and column j of the matrix B_ℓ, $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$ is the collection of prior hyperparameters, and σ_i^2, σ_j^2 are local scale parameters. For each element i of the intercept vector c, it is common to specify an uninformative prior by setting the prior variance equal to $100 \times \sigma_i^2$.

In view of the fact that the usual Minnesota-type prior ignores the panel structure in the data, Angelini et al. [2019] recently proposed a modified Minnesota-type shrinkage prior to carefully deal with the panel structure. In particular, a different hyperparameter λ_4 is introduced in (1.5) on coefficients related to other countries' variables. Angelini et al. [2019] apply this approach in a forecasting exercise with a Euro area data set and find that it provides some gains. However, this still belongs to the class of Minnesota-type priors. This may come with costs due to parameter uncertainty, since the hyperparameters $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and local scale parameters σ_i^2, σ_j^2 (commonly obtained from AR(1) estimates) are fixed.

Since there are a variety of restrictions of interest in multi-country VARs, another strand of literature suggests designing shrinkage priors to explore these restrictions. Consider the coefficient matrix B_l defined in (1.2):

$$B_\ell = \begin{bmatrix} B_{11,\ell} & \cdots & B_{1N,\ell} \\ \vdots & \ddots & \vdots \\ B_{N1,\ell} & \cdots & B_{NN,\ell} \end{bmatrix}, \ell = 1, \ldots, p, \tag{1.6}$$

where each block $B_{ij,\ell}$, $i, j = 1, \ldots, N$, has dimension $G \times G$. According to Canova and Ciccarelli [2013], it is interesting to check whether certain restrictions exist and what their implications are. For example, cross-sectional heterogeneity (CSH) exists when $\exists\ i, j, i \neq j$, such that $B_{ii,l} \neq B_{jj,l}$ for some l and $c_i \neq c_j$. Dynamic interdependencies (DI) occur when at least one block $B_{ij,l} \neq 0$ for a given i, l and $i \neq j$.[5]

In a special case with one lag ($p = 1$) and no SV, Koop and Korobilis [2016] develop the stochastic

[5]There is one other class of possibly important restrictions: static interdependencies (SI), which occur when the covariance matrix Σ_t is not block diagonal. However, it is not easy to implement these restrictions when stochastic volatility is allowed.

model specification search (SSSS) prior:

$$vec(B_{ij,1}^{\mathrm{DI}}) \sim (1-\gamma_{ij}^{\mathrm{DI}})N(0,\tau_{ij}^2 \times \underline{c}^{\mathrm{DI}} \times I_G) + \gamma_{ij}^{\mathrm{DI}}N(0,\tau_{ij}^2 \times I_G), \quad i,j=1,\ldots,N, i \neq j \tag{1.7}$$

$$vec(B_{ii,1}^{\mathrm{CSH}}) \sim (1-\gamma_{ii}^{\mathrm{CSH}})N(vec(B_{jj,1}),\xi_{ij}^2 \times \underline{c}^{\mathrm{CSH}} \times I_G) + \gamma_{ii}^{\mathrm{CSH}}N(vec(B_{jj,1}),\xi_{ij}^2 \times I_G), \quad i,j=1,\ldots,N, \tag{1.8}$$

where $\gamma^{\mathrm{DI}}, \gamma^{\mathrm{CSH}}$ are indicators, τ_{ij}^2, ξ_{ij}^2 are prior variance parameters, and $\underline{c}^{\mathrm{DI}}, \underline{c}^{\mathrm{CSH}}$ are small constants to make prior variances smaller in the spike components. The priors in (1.7)-(1.8) provide an extension of the stochastic search variable selection (SSVS) prior in George et al. [2008] to multi-country VARs. It is the first attempt to examine the existence (or absence) of certain dependencies and homogeneities for coefficients in multi-country VARs. If $\gamma_{ij}^{\mathrm{DI}} = 0$, the coefficients on the lags of all country j variables for country i are set to very small values near zero. If $\gamma_{ij}^{\mathrm{CSH}} = 0$, the coefficients on the lags of all country i variables for itself are to be concentrated at coefficients related to country j. Korobilis [2016] uses this prior in a multi-country forecasting exercise for bond yields of Eurozone countries and finds that it performs comparably with alternative shrinkage priors.

In terms of MCMC estimation, application of the DI restrictions is relatively straightforward, while application of the CSH restrictions is non-trivial, since we seek to use priors to push the model toward equality of matrices $B_{ii,\ell} = B_{jj,\ell}$ for $i \neq j$ and $\ell = 1,\ldots,p$. Koop and Korobilis [2016] provide a novel solution to the problem, but it still introduces prior dependence across equations, which also appears in the conditional posteriors, making it difficult to apply efficient algorithms as in Carriero et al. [2019] to estimate the model equation by equation. In addition, the priors of (1.7)-(1.8) involve tuning many hyper-parameters. For example, if we specify hyper-priors to infer $\gamma_{ij}^{\mathrm{DI}} \sim Bernoulli(\pi_{ij}), \gamma_{ii}^{\mathrm{CSH}} \sim Bernoulli(\pi_{ii})$, and $\tau_{ij}^2 \sim Ga(a_1, a_{ij}), \xi_{ij}^2 \sim Ga(b_1, b_{ij})$, then we need to specify many more hyperparameters related to those hyper-priors. Moreover, in contrast to the case of the scale mixtures of Normals priors we use, the theoretical properties are not known.

The next section introduces hierarchical shrinkage priors for multi-country VARs. Our priors make efficient MCMC algorithms easy to apply, without the need to tune many hyperparameters, and with the ability to push the model toward both CSH and DI restrictions.

1.3 Hierarchical shrinkage in multi-country VARs

The hierarchical shrinkage we consider is inspired by recent advances in the literature on scale mixtures of Normals priors and their successful applications in single-country Bayesian VARs. Since the seminal work by Carvalho et al. [2010], a variety of scale mixtures of Normals priors have been proposed in the literature. These priors include the Horseshoe prior (Carvalho et al. [2010]), the Normal-Gamma prior (Griffin and Brown [2010]), the Normal-Gamma-Gamma prior (Griffin and Brown [2017]), and several other alternatives. Compared to a conventional Normal prior, these prior distributions are spiked at the origin to provide severe shrinkage towards zero for the parameters of interest, while at the same time they also have heavy-tails to allow little shrinkage of, say, intercept terms in (1.2). In addition, these priors have computational advantages compared to the spike-and-slab prior (Carvalho et al. [2009]). Applications to single-country Bayesian VARs mostly focus on the Normal-Gamma prior; see, for instance, Huber and Feldkircher [2019] and Korobilis and Pettenuzzo [2019]. Follett and Yu [2019] introduce the Horseshoe prior. These papers find that scale mixtures of Normals priors serve as competing alternatives to Minnesota priors, in terms of both forecasting and structural analysis.

For setting up priors on the parameters associated with model (1.2), let $\beta_i = vec([c_i, B_{ii,1}, \ldots, B_{ii,p}]'), i = 1, \ldots, N$, and $\beta^{\text{CSH}} = (\beta_1, \ldots, \beta_N)'$ be the collection of coefficients related to CSH restrictions. Due to the forecasting focus, we consider further splitting β^{CSH} into three blocks:

$$
\begin{aligned}
\beta_c^{\text{CSH}} &= (c_1', c_2', \cdots, c_N')', \\
\beta_{AR}^{\text{CSH}} &= (\text{diag}(B_{11,1}), \text{diag}(B_{22,1}), \ldots, \text{diag}(B_{NN,1}), \ldots, \text{diag}(B_{11,p}), \ldots, \text{diag}(B_{NN,p})), \\
\beta_o^{\text{CSH}} &= \beta^{\text{CSH}} \setminus \{\beta_c^{\text{CSH}}, \beta_{AR}^{\text{CSH}}\},
\end{aligned}
$$

where β_o^{CSH} includes the set of parameters related to cross-variable lags that are in β^{CSH} and not related to the intercept terms and own lags. Similarly, let $\beta_{il}^* = vec([B_{i1,l}, \ldots, B_{ii-1,l}, B_{ii+1,l}, \ldots, B_{iN,l}]'), i = 1, \ldots, N, l = 1, \ldots, p$, and $\beta^{\text{DI}_i} = (\beta_{i1}^*, \ldots, \beta_{ip}^*)'$ be the collections of coefficients related to DI restrictions in country i. Finally, define α as the free elements in A^{-1}. We have $N + 4$ blocks of coefficients: $\beta_c^{\text{CSH}}, \beta_{AR}^{\text{CSH}}, \beta_o^{\text{CSH}}, \beta^{\text{DI}_1}, \ldots, \beta^{\text{DI}_N}, \alpha$.

Using β as a generic notation for one block of coefficients, the hierarchical shrinkage we consider takes the form:

$$
\beta_j \sim N(0, \lambda\omega_j), \quad \omega_j \sim \mathcal{F}, \tag{1.9}
$$

where $j = 1, \ldots, \dim(\beta)$ and $\mathcal{F}$ denotes some pre-specified distribution for the local shrinkage parameter ω_j, which will be defined later. λ serves as a global shrinkage parameter, which can be specified as an additional hyper-prior to learn the values from the data. Taking β^{DI_1} as an example, because λ loads for all elements in β^{DI_1}, if $\lambda \to 0$, all β^{DI_1} are assumed to be identical (centered at zero), which implies that there is no dynamic interdependence for country 1. It is worth mentioning that, since both local and global shrinkage parameters are fully learned from the data, the hierarchical shrinkage approach offers more flexibility than the conventional Minnesota prior, in which all hyperparameters are fixed at some pre-specified values.

1.3.1 Choices of the priors

What remains is to specify $\mathcal{F}$ and hyper-priors on the global shrinkage parameters. Here we focus on three different scale mixtures of Normals priors, since they have been applied successfully in single-country

Bayesian VARs.[6]

The first prior we consider is the Horseshoe prior:

$$\beta_j|\omega_j^2 \sim N(0,\omega_j^2),\ \ \omega_j^2|\gamma_j^2 \sim \mathcal{G}(\frac{1}{2},\gamma_j^2),\ \ \gamma_j^2 \sim \mathcal{G}(\frac{1}{2},\lambda), \tag{1.10}$$

where $\mathcal{G}$ denotes the Gamma distribution.[7] We use the parameterization of the Horseshoe prior as in Armagan et al. [2011]. It can be shown that the marginal distribution of ω_j^2 follows $C^+(0,1)$, where $C^+(0,1)$ denotes a half-Cauchy distribution on $\mathcal{R}^+$ with scale parameter 1, and λ serves as the global shrinkage parameter, which is the original parameterization in Carvalho et al. [2010] and used in Follett and Yu [2019]. The prior (1.10) has computational advantages, since the conditional posteriors are conjugate (Makalic and Schmidt [2015]), making MCMC estimation straightforward. For the global shrinkage parameter λ, we also follow Armagan et al. [2011] and set $\lambda \sim C^+(0,1)$.

The second prior we consider is the Normal-Gamma prior:

$$\beta_j|\omega_j^2 \sim N(0,\omega_j^2),\ \ \omega_j^2 \sim \mathcal{G}(a^\omega,\frac{a^\omega\kappa^2}{2}). \tag{1.11}$$

The above, with a slightly different parameterization, is first introduced in Griffin and Brown [2010] and has recently been applied in single-country Bayesian VARs (Huber and Feldkircher [2019]) and time-varying parameter models (Bitto and Frühwirth-Schnatter [2019]). Using Monte-Carlo simulations, Bitto and Frühwirth-Schnatter [2019] find that a^ω controls the origins of the marginal prior distribution of $p(\beta_j)$ and κ^2 is the global shrinkage parameter. It can also be shown that (1.11) simplifies to Bayesian Lasso (Park and Casella [2008]) if $a^\omega = 1$ (Griffin and Brown [2010]). To infer hyperparameter values, we follow Bitto and Frühwirth-Schnatter [2019] to set $a^\omega \sim \mathcal{E}(b)$ and $\kappa^2 \sim \mathcal{G}(d_1,d_2)$, where $\mathcal{E}$ denotes the exponential distribution.

The final prior we consider is the Normal-Gamma-Gamma prior:

$$\beta_j|\tau_j^2,\lambda_j^2 \sim N\left(0,\phi\frac{\tau_j^2}{\lambda_j^2}\right),\ \ \tau_j^2 \sim \mathcal{G}(a,1),\ \ \lambda_j^2 \sim \mathcal{G}(c,1), \tag{1.12}$$

where $\phi = 2c/(a\kappa^2)$. The above, again with a slightly different parameterization, is first introduced in Griffin and Brown [2017]. Cadonna et al. [2020] apply it to time-varying parameter models. They show that specification in (1.12) is very general and nests some commonly used shrinkage priors. They also provide a comprehensive analysis of the properties of the Normal-Gamma-Gamma prior. It can be shown that a controls the origin and c controls the asymptotic tail behavior of the marginal prior distribution of $p(\beta_j)$. ϕ is the global shrinkage parameter. Using both simulated and real macroeconomic data, they

[6]Recently, there is a growing interest in the Dirichlet-Laplace prior; see Koop et al. [2020] for an application. However, we do not consider it here, since the Dirichlet-Laplace prior is the scale mixture of the Laplace prior, which is not the main focus of this paper.

[7]We use the parameterization of the $\mathcal{G}(a,b)$ distribution with pdf given by $f(y) \propto y^{a-1}\exp(-by)$.

find that the specification in (1.12) delivers relatively sparse solutions in time-varying parameter models. However, no attempts have been made so far to examine its performance for macroeconomic forecasting. Following Cadonna et al. [2020], we set $2a \sim \mathcal{B}(\alpha_a, \beta_a)$, $2c \sim \mathcal{B}(\alpha_c, \beta_c)$, and $\kappa^2|a, c \sim F(2a, 2c)$ to learn these hyperparameter values, where $\mathcal{B}$ denotes the Beta distribution and F is the standard F distribution.

1.3.2 Comparisons of the priors

Theoretically, the scale mixtures of Normals priors are typically compared in terms of the concentration properties at the origin and the asymptotic tail behavior. Results for the Horseshoe prior and the Normal-Gamma-Gamma prior can be found in Carvalho et al. [2010] and Cadonna et al. [2020], respectively. However, no results are available for the Normal-Gamma prior. In the following theorem, we formally characterize the tail behavior and concentration properties for the Normal-Gamma prior.

Theorem 1.3.1. *Let $\beta_j \sim \mathcal{NG}(a^\omega, \kappa^2)$, where $\mathcal{NG}$ is the Normal-Gamma prior parameterized as in* (1.11). *Then, the marginal density $\pi_{NG}(\beta_j)$ satisfies the following:*

- *Concentration properties: As $|\beta_j| \to 0$, we have*
  1. *if $a^\omega > \frac{1}{2}$, $\pi_{NG}(\beta_j) = O(1)$;*
  2. *if $0 < a^\omega < \frac{1}{2}$, $\pi_{NG}(\beta_j) = O\Big(\frac{1}{|\beta_j|^{\frac{1}{2}-a^\omega}}\Big)$;*
  3. *if $a^\omega = \frac{1}{2}$, $\pi_{NG}(\beta_j) = O\Big(\frac{1}{\log(|\beta_j|)}\Big)$;*
- *Asymptotic tail behavior: As $|\beta_j| \to \infty$, we have $\pi_{NG}(\beta_j) = O\left(\frac{|\beta_j|^{a^\omega - 1}}{\exp\left(\sqrt{a^\omega \kappa^2}|\beta_j|\right)}\right)$.*

Proof. See Appendix A. □

The results for Horseshoe, Normal-Gamma, and Normal-Gamma-Gamma priors in terms of both asymptotic tail behavior and concentration properties are summarized in Table 1.1. Clearly, Normal-Gamma and Normal-Gamma-Gamma priors share similar concentration properties, possessing unbounded density near the origin if either $0 < a^\omega < \frac{1}{2}$ or $0 < a < \frac{1}{2}$. Both priors diverge to infinity with a polynomial order, much faster than the Horseshoe prior (with a logarithmic order). For the tail behavior, it follows from straightforward calculation that $\lim_{|\beta_j| \to \infty} \pi_{NG}(\beta)/\beta^{-2} = 0$ and $\lim_{|\beta_j| \to \infty} \pi_{NGG}(\beta)/\beta^{-2} = \infty$ if $0 < c < \frac{1}{2}$, which implies that the Normal-Gamma prior has lighter tails than the Horseshoe prior, but the Normal-Gamma-Gamma prior has heavier tails than the Horseshoe prior. The Normal-Gamma-Gamma prior is the only one that can achieve a polynomial rate of convergence in both the tails and the origin. It extends the Normal-Gamma prior by having heavier tails. Compared to the Horseshoe prior, it puts more probability mass at the origin and offers more flexibility in modeling tails by the hyperparameter c.

	Tail Decay	Concentration at zero
Horseshoe	$O\left(\frac{1}{\beta_j^2}\right)$	$O\left(\log\left(\frac{1}{\lvert\beta_j\rvert}\right)\right)$
Normal-Gamma	$O\left(\frac{\lvert\beta_j\rvert^{a^\omega-1}}{\exp\left(\sqrt{a^\omega\kappa^2}\lvert\beta_j\rvert\right)}\right)$	$O(1)$ if $a^\omega > \frac{1}{2}$
		$O\left(\frac{1}{\lvert\beta_j\rvert^{\frac{1}{2}-a^\omega}}\right)$ if $0 < a^\omega < \frac{1}{2}$
		$O\left(\frac{1}{\log\left(\lvert\beta_j\rvert\right)}\right)$ if $a^\omega = \frac{1}{2}$
Normal-Gamma-Gamma	$O\left(\frac{1}{\beta_j^{2c+1}}\right)$	$O(1)$ if $a > \frac{1}{2}$
		$O\left(\frac{1}{\lvert\beta_j\rvert^{1-2a}}\right)$ if $0 < a < \frac{1}{2}$
		$O\left(\frac{1}{\log\left(\lvert\beta_j\rvert\right)}\right)$ if $a = \frac{1}{2}$

Table 1.1: Tail behavior and concentration around zero for Horseshoe, Normal-Gamma, and Normal-Gamma-Gamma priors

Empirically, Cadonna et al. [2020] use Euro area macroeconomic data and find that the Normal-Gamma-Gamma prior achieves more sparse parameter estimates than other shrinkage priors in a time-varying parameter model framework. However, as pointed out in Giannone et al. [2021], sparsity does not necessarily imply good forecasting performance. As we shall see, heavy tails is an important feature to obtain better forecasting performance in multi-country VARs, since the Horseshoe prior forecasts well in many cases. The extension to the Normal-Gamma-Gamma prior is less useful. The light-tailed Normal-Gamma prior is useful for output growth forecasts in some cases, but it is outperformed by Horseshoe and Normal-Gamma-Gamma priors for inflation and interest rate forecasts.

1.3.3 Other competing models

In addition to the multi-country VARs mentioned above, we also consider several alternative models which are commonly used in macroeconomic forecasting. These models include country-specific VARs with SV in which priors are specified as either Minnesota-type or hierarchical Normal-Gamma as in Chan [2021]; country-specific factor-augmented VAR (FAVAR) models with SV; global VAR (GVAR) with SV; and multi-country VAR-SV with factor shrinkage as in Canova and Ciccarelli [2009]. We use the country-specific VAR-SV with Minnesota prior as the benchmark, as it is the most commonly used model in the macroeconomic forecasting literature. A description of all the models under comparison is provided in Table 1.2. More details on the specification of the various models and associated priors can be found in the Appendix.

Model	Description
CVAR	country-specific VAR(p) with Minnesota prior
CVAR-H	country-specific VAR(p) with hierarchical shrinkage as in Chan [2021]
CFAVAR	country-specific factor-augmented VAR(p), with factors extracted from foreign variables
GVAR	Global VAR(p)
CC	parameters are assumed to follow an exact factor structure, as in Canova and Ciccarelli [2009]
MIN	priors are Minnesota-type as in Angelini et al. [2019]
SSSS	stochastic specification search and selection prior as in Korobilis [2016]
HS	hierarchical shrinkage with Horseshoe prior
NG	hierarchical shrinkage with Normal-Gamma prior
NGG	hierarchical shrinkage with Normal-Gamma-Gamma prior

Note: All the models include SV.

Table 1.2: List of competing models

1.4 Estimation Algorithms

1.4.1 Estimation outline

We estimate all of the models listed in Table 1.2 using Markov Chain Monte Carlo (MCMC) methods. All of our estimates are based on 30,000 posterior draws, with the first 10,000 discarded and the remaining 20,000 post-burn-in draws retained. This section provides a brief overview of our methods. The Appendix and the studies cited below provide more details on algorithms and priors.

For country-specific VARs with SV and factor-augmented VARs with SV, we use the non-conjugate Minnesota-type prior. The Gibbs sampling details of the country-specific VAR with SV are provided in Carriero et al. [2019]. The intercept and autoregressive coefficients are estimated by the corrected triangular algorithm proposed in Carriero et al. [2021].[8] Stochastic volatility is estimated with the algorithm in Del Negro and Primiceri [2015]. For free elements in A, we use the algorithm as in Cogley and Sargent [2005]. The Gibbs sampler used in country-specific VARs with SV can be easily extended to allow for augmented factors extracted from foreign variables. For the GVAR, we use a Minnesota-type prior similar to Huber [2016]; details of the algorithms are provided there. For the CVAR-H, CFAVAR, and GVAR specifications, we follow Chan [2021] and put hyper-priors on the overall shrinkage parameters. An additional step is needed to update these parameters. In the case of the CVAR-H model, we use the default setting as in Chan [2021], and sampling details can be found there.

To estimate the multi-country VAR with SV and the factor shrinkage approach (the CC model), we use an exact factorization as in Canova et al. [2007] and Korobilis [2016]. Algorithms are provided in these papers. SV can be easily added to this model and estimated similarly as in the country-specific case. In the

[8] A summary of the corrected triangular algorithm is provided in Appendix C.4.

case of the MIN specification that features a Minnesota-type prior, the algorithm can be obtained similarly, but for the Minnesota-type prior we use a specification similar to Angelini et al. [2019]. We also put a hyper-prior on the overall shrinkage parameter related to coefficients on lagged foreign variables. Finally, for the three hierarchical shrinkage approaches (HS, NG, and NGG), the Gibbs samplers are again very similar as in other models, but additional blocks of sampling are needed to update the hyperparameters. In particular, the details of the NG and NGG priors can be adapted as in Bitto and Frühwirth-Schnatter [2019] and Cadonna et al. [2020], respectively. For the HS prior, since we use a different parameterization, algorithms in Follett and Yu [2019] cannot be directly applied, but sampling schemes in Armagan et al. [2011] for univariate regression models can be extended. We provide a summary of the MCMC algorithms below.

Algorithm 1: MCMC inference for multi-country VARs with SV

Step 1: initialization;

Step 2: **for** $i = 1, \ldots, NG$ **do**

Use the corrected triangular algorithm in Carriero et al. [2021] to obtain posterior draws from VAR mean coefficients (intercepts and autoregressive coefficients);

end

Step 3: Use the algorithm in Cogley and Sargent [2005] to update the free elements in A;

Step 4: Update the hyperparameters in prior error covariance matrices, with conditional posteriors that depend on prior choices which can found in Appendix B.2;

Step 5: Use the algorithm in Del Negro and Primiceri [2015] to update the volatilities h_t and error variance of innovations Φ^h.

1.4.2 Computational efficiency

In this section, we briefly compare the computational efficiency of the MCMC algorithms for the multi-country VARs (note that we omit the CC specification, which is covered elsewhere in the literature). To assess the efficiency of the algorithms, we compute the potential scale reduction factors (PSRFs) detailed in Brooks and Gelman [1998]. A value of the PSRFs below 1.1 is generally taken as indication that the chain has satisfactory mixing properties. In Table 1.3, we report average PSRFs of parameters needed to construct forecasts and obtain predictive distributions: vec(B), α, and vech(Φ). We use the data set as in our forecasting exercises (21 variables) and all models include 4 lags. As is clear, our algorithms show satisfactory mixing and convergence properties.

	MIN	SSSS	HS	NG	NGG
vec(B)	1.005	1.002	1.002	1.002	1.008
α	1.004	1.004	1.002	1.002	1.004
vech(Φ)	1.061	1.040	1.046	1.077	1.069

Notes: The lags are set to 4 for all models except SSSS, in which we use 1 lag, to match the specification of our forecasting application.

Table 1.3: Mixing and convergence statistics (PRSFs) for multi-country VARs with 21 variables

As we discussed above, to handle CSH restrictions, the SSSS prior introduces dependence across equations, which prevents the use of an efficient sampling algorithm for the VAR's coefficients. Table 1.4 shows the computational time (in seconds) necessary to produce 10,000 draws from the posteriors of multi-country VARs including 21 variables. For this time comparison, all models include only 1 lag. Clearly, the specification with the SSSS prior takes much longer to estimate, roughly 5 times slower than the other specifications. In forecasting, computations can be extremely burdensome as estimation has to be done many times. Interestingly, the added blocks of sampling for hyperparameters in our proposed HS, NG, and NGG methods have very small additional computational costs compared to the Minnesota-type prior of the MIN specification.

MIN	SSSS	HS	NG	NGG
393	2077	367	324	427

Table 1.4: Time (in seconds) taken to obtain 10,000 posterior draws for various multi-country VARs with 21 variables and 1 lag

1.5 Data and Forecast Evaluation

1.5.1 Data

We examine the forecast performance of the various specifications using a G7 country data set: USA, UK, Germany (DEU), France (FRA), Italy (ITA), Japan (JPN), and Canada (CAN). In brief, we build a 3-variable data set for each country at the quarterly frequency, with a sample period of 1973Q1-2019Q4. The variables consist of real GDP growth, CPI inflation, and a short-term interest rate (the 3-month government bill rate). Table 1.5 presents the details of the data set along with the transformations of the variables and the corresponding sources.[9] Note that, like most other multi-country studies, we do not consider real-time data and use data from the last available vintage, owing to the lack of availability of real-time data for all seven countries.

[9] We obtain data from different sources due to data availability. In particular, the OECD provides real GDP growth data for Germany before 1991, and the GFD provides very long coverage for interest rate data in many countries.

Variable	Data source	Transformation
Real GDP growth	OECD	$4y_t$
CPI inflation	FRED	$400\log(y_t/y_{t-1})$
Interest rate	GFD	None

Notes: Because the OECD reports GDP growth as quarterly percent changes, we multiply the source data by 4 to obtain an approximate annual rate. FRED refers to the database maintained by the Federal Reserve Bank of St. Louis. CPI inflation is measured with the annualized quarterly percent change in the quarterly average level of the monthly CPI. GFD refers to Global Financial Database, from which we obtain the 3-month government bill rate and form the quarterly series as the average of values for the months of each quarter.

Table 1.5: Data description and variable transformation

1.5.2 Forecast evaluation

We consider both point and density forecasts at horizons up to 12 steps (three years) ahead. The forecasting exercise is performed in pseudo-real time; i.e., we do not use information that is not available when the forecast is made. Parameter estimation and forecasting are done recursively, using an expanding window of data for model estimation. The initial estimation sample runs from 1973Q1 to 1994Q4, the first available forecast is for 1995Q1, and forecasts are generated up to 12 quarters ahead. Our last estimation sample runs from 1973Q1 to 2016Q4, yielding forecasts from 2017Q1 to 2019Q4.

For all the models considered here, the full distribution of the forecasts is not available in closed form, and a simulation algorithm is required. At each post burn-in draw, we compute the implied path of $\hat{y}_{t+h}^{(j)}$ to generate a total of 20,000 draws from the predictive distribution.

Each point forecast is measured as the median of the predictive density. We evaluate them in terms of root mean squared forecast error (RMSFE). Letting $\hat{y}_{t+h}(M)$ be the forecast of the (scalar, for simplicity of notation here) target variable y_{t+h} made by model M and P be the total number of generated forecasts, the RMSFE made by model M for horizon h is

$$\text{RMSFE}_h^M = \sqrt{\frac{1}{P}\sum(\hat{y}_{t+h}(M) - y_{t+h})^2}. \tag{1.13}$$

In the case of the density forecasts, we use the continuous ranked probability score (CRPS) proposed by Gneiting and Raftery [2007], which is less sensitive to outliers than other density evaluation measures, such as the log score. The CRPS metric for the each variable at time t for horizon h is defined as

$$\text{CRPS}_t(F, y_{t+h}^o) = \mathbb{E}_F\left|y_{t+h}^d - y_{t+h}^o\right| - \frac{1}{2}\mathbb{E}_F\left|y_{t+h}^d - y_{t+h}^{dd}\right|, \tag{1.14}$$

where F denotes the cumulative distribution function associated with the predictive density f, y_{t+h}^o denotes the observed value, and y_{t+h}^d, y_{t+h}^{dd} are independent draws from the predictive posterior distribution. Following Smith and Vahey [2016], we compute (1.14) by numerical integration methods, which is shown to be more accurate and efficient. It is also worth mentioning that the lower the value of the CRPS, the more accurate the predictive density is.

Finally, to provide a statistical comparison of predictive accuracy, we apply the Diebold and Mariano [1995] (DM) test for equal forecast accuracy. Yet, our models are nested in many cases. It is well known that the DM test for nested models is undersized, and the results can be viewed as conservative for equal forecast accuracy in finite samples. We follow Coroneo and Iacone [2020] to apply fixed-smoothing asymptotics for the DM test, which is shown to deliver predictive accuracy tests that are correctly sized even when the number of out-of-sample observations are small.

1.6 Empirical Results

1.6.1 Overall forecast performance

We first provide a summary of the forecast evaluation exercise in Table 6. As we have 7 countries and forecasts are generated from 1 to 12 steps ahead, for each variable we have 84 forecasts from each model. The table reports the number of cases in which each model is best, for all horizons, short horizons ($h \leqslant 6$), and long horizons ($h > 6$).

Based on Table 6, the results can be summarized as follows. First, our proposed hierarchical shrinkage in multi-country VARs — used in the HS, NG, and NGG specifications — is quite helpful. This is particularly true with the HS prior. It has the most wins for inflation in terms of both point and density forecast, and it is the best performing model for output growth and the interest rate in terms of density forecasts. More specifically, for output growth, the HS prior is the best in 31 (out of 84) cases in terms of density forecasts, compared to 15 cases for the second-best performing model, which is the multi-country VAR with the NG prior. For inflation, the HS specification performs the best in more than half (44) of the cases in terms of point forecasts and exactly half of the cases in terms of density forecasts. The benefits are also more evident at long horizons, with 29 wins in point forecasts and 27 wins in density forecasts. For the interest rate, the HS prior also has the most wins (22 cases) for density forecasts, compared to 16 cases obtained from the factor shrinkage approach of the CC specification. Second, the NG specification is the best in nearly half (40) of the cases for output growth in terms of point forecasts, compared to 19 cases from the SSSS prior. The CC specification has the most wins for the interest rate in terms of point forecasts. However, the NG and CC specifications do not forecast well for other variables. For instance, for inflation at long horizons, the NG and CC specifications are never selected as best. SSSS never becomes the best for density forecasts of the interest rate. Third, including information across countries is very useful particularly for output growth; the single-country CVAR and CVAR-H specifications are never the best in point or density forecasts of output. Although the baseline single-country CVAR has the second best forecast performance for inflation, with 13 wins in terms of point forecast and 10 wins for density forecast, it is clearly outperformed by the multi-country HS specification. Regarding the CFAVAR and GVAR specifications, they perform better for inflation than the other variables but also fall short of the HS approach. The single-country CVAR-H specification that imposes hierarchical shrinkage works better for the interest rate than output growth or inflation, but still falls short of other models, including the multi-country HS specification.

Overall, the usefulness of hierarchical shrinkage for multi-country VARs in forecasting key macroeconomic variables emerges rather clearly. In general, the HS prior is better than the other two scale mixtures of Normals priors. As discussed in Section 3, with more hyperparameters controlling both origins and tails, the NGG prior is theoretically more flexible than the HS prior, which also provides a heavy-tailed extension of the NG prior. However, the theoretical advantages do not necessarily transfer to better forecasting performance, as it only ranks first (tied) in the case of density forecasts of the interest rate at short horizons.

The summary results should be interpreted with care, as they are based on deterministic comparisons (i.e., the best model could be not statistically better than the second best model). They also ignore the cross-country differences and the potential differences of model performance over time. Yet, they provide a broad overview of the models' performance. More detailed results and statistical comparisons are presented in the next subsection.

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	point	density	point	density	point	density	**Inflation**	point	density	point	density	point	density
CVAR	0	0	0	0	0	0	CVAR	13	10	7	7	6	3
CVAR-H	0	0	0	0	0	0	CVAR-H	0	2	0	2	0	0
CFAVAR	1	3	0	0	1	3	CFAVAR	8	6	3	2	5	4
GVAR	0	0	0	0	0	0	GVAR	3	10	2	3	1	7
CC	4	1	2	1	2	0	CC	0	0	0	0	0	0
MIN	5	7	1	2	4	5	MIN	8	9	8	8	0	1
SSSS	19	11	12	11	7	0	SSSS	4	3	4	3	0	0
HS	9	31	4	9	5	22	HS	44	42	15	15	29	27
NG	40	15	21	12	19	3	NG	1	2	1	2	0	0
NGG	6	16	2	7	4	9	NGG	3	0	2	0	1	0
Interest rate	point	density	point	density	point	density							
CVAR	6	10	3	4	3	6							
CVAR-H	6	13	6	9	0	4							
CFAVAR	0	0	0	0	0	0							
GVAR	0	1	0	1	0	0							
CC	30	16	11	3	19	13							
MIN	1	1	1	1	0	0							
SSSS	8	0	2	0	6	0							
HS	19	22	7	8	12	14							
NG	10	11	8	7	2	4							
NGG	4	10	4	9	0	1							

Notes: See Table 1.2 for a list of models and Section 1.5.2 for the evaluation criteria.

Table 1.6: Summary statistics: number of cases when one model becomes the best

1.6.2 Forecast evaluation: cross-country differences

Building on the previous section's overview of the forecast performance of the various model specifications, we turn now to a more quantitative assessment of forecast accuracy across models and countries. To this end, Tables 7-9 report relative RMSFEs and CRPSs for all G7 countries at selected horizons, $h = 1, 4, 8, 12$. Entries shaded in grey indicate the best performing model. RMSFEs and CRPSs in levels from the benchmark model are reported in the appendix's Table 1.17.

Consider first the results for output growth. In general, the best performing specifications are the multi-country VARs with scale mixtures of Normals priors (i.e., one of HS, NG, or NGG). In all but a few cases, these specifications improve on the point and density forecast accuracy of the CVAR benchmark. In general, hierarchical shrinkage of multi-country VARs, in particular with the HS prior, is rather a safe option for forecasters since it provides gains for both point and density forecasts in most of the cases. The other specifications do not fare as well in improving on the accuracy of the benchmark. One of the better alternatives is the SSSS specification, which is best for a few countries at short horizons, although in some other cases it is fairly strongly beaten by both other approaches (i.e., its performance is somewhat uneven). The MIN specification — a multi-country VAR estimated with a Minnesota-type prior — is only best in a few instances, all long-horizon forecasts for the USA. Similarly, the CC (factor-based shrinkage of coefficients) and CFAVAR (factor-augmented single-country models) are selected as best for no more than a few country/horizon/type of forecast combinations. Perhaps not surprisingly, the accuracy of the CVAR-H specification (single-country with hierarchical shrinkage) is relatively similar to the CVAR baseline, sometimes a little better and sometimes a little worse.

Moving to the inflation forecasts of Table 8, both some commonalities and different stories are evident. First, the multi-country HS specification continues to provide the best forecast in many cases, particularly for Canada, France, Italy, and Japan at longer horizons, as well as for density forecasts for the USA. Second, other competing specifications, including CFAVAR, GVAR, MIN, and SSSS, are occasionally the best model, but they are relatively less accurate in other cases. For example, the SSSS specification is the best for the USA at the 1-step-ahead and 4-steps-ahead horizons in terms of point forecast and 4-steps-ahead horizon in terms of density forecast, but it does not provide gains to forecasts for Italy. Third, results are somewhat different for the UK, perhaps due to historical inflation in the UK being rather different, with stronger peaks in the 1970s and a volatile period around the Black Wednesday crisis. In the case of the UK, the best-performing forecasting model is the benchmark specification, with RMSFE and CRPS ratios that exceed 1 in all but one case.

Turning to the interest rate forecasts, which are presented in Table 9, we again see similarities as well as some different patterns. Sorting through differences across countries, the multi-country HS specification continues to perform relatively well. For most, although not all, countries, forecasts from this model are more accurate than the benchmark, by margins as large as 32 percent. The other two multi-country scale mixtures of Normals priors — NG and NGG — don't offer any clear advantages over the HS specification, sometimes slightly to modestly improving accuracy and other times reducing accuracy (relative to the HS prior). Of the other multi-country VAR specifications, the CC model performs better in forecasting interest rates than output growth and inflation. For a few country/horizon combinations, the CC model is most accurate, whereas for some others, it is notably less accurate than the CVAR benchmark. The performance of the SSSS specification is also uneven, often much less accurate than the benchmark (e.g., for Canada) but occasionally more accurate (e.g., 4- and 8-steps-ahead forecasts for Germany). The performance of the GVAR is also inconsistent.

RMSFE					Canada										Germany			
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.993	0.903	0.940	1.008	1.000	0.845	0.927	0.865	0.861	0.981	0.891	0.937	1.084	0.981	0.853	0.972	0.899	0.921
$h = 4$	1.007*	1.009	0.986*	1.117	0.991	0.992	0.945	0.954	0.947	0.998	0.987	0.990	1.141	0.982	0.977	0.996	0.994	0.995
$h = 8$	1.006	0.993	1.002	1.237	1.003	1.032	0.976	0.975	0.982	0.997	0.996	1.001	1.193*	1.002	1.019	0.985	0.974	0.983
$h = 12$	1.000	0.988	0.992	1.268	1.013	1.031	0.988	1.003	0.994	0.999	0.996	1.001	1.186	0.996	1.025	0.974	0.977	0.976
					France										Italy			
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.989	0.962	0.971	1.038	1.024	0.954	0.971	0.934	0.929	0.981	0.980	1.008	0.959	1.075	0.981	1.021	0.984	1.000
$h = 4$	0.991	0.991	0.996	1.190	0.977	1.032	0.933	0.908	0.920	0.999	0.986	1.017	1.070	1.007	1.010	0.957	0.935	0.952
$h = 8$	0.985*	1.002	0.996	1.290	0.993	1.117	0.931	0.919	0.924	1.004	1.002	1.027*	1.020	1.016	1.022*	0.961	0.953	0.955
$h = 12$	0.983	1.001	0.994	1.374	1.001	1.143*	0.932	0.921*	0.940	1.008	0.998	1.027*	0.967	1.020	1.003	0.971	0.980	0.980
					Japan										UK			
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.005	0.983	0.993	1.128	1.014	0.948*	0.983	0.966	0.971	1.019	1.018	1.009	0.924	1.136	0.954	1.000	1.051	1.015
$h = 4$	0.991	0.986	0.996	1.122	0.988	0.978	0.978	0.954	0.966	1.003	1.002	1.005	1.075	0.974	0.943	0.949	0.930	0.936
$h = 8$	0.994	0.995	0.996	1.126	0.999	0.979	0.972	0.963	0.958	1.013*	1.006	1.002	1.099	0.981	0.959	0.972	0.965	0.961
$h = 12$	0.990	0.991	0.993	1.102	0.990	1.010	0.969*	0.960	0.956	1.007	1.002	1.001	1.206	1.001	0.959	0.989	1.003	0.988
					USA													
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG									
$h = 1$	0.989	0.972	1.010	1.080	1.027	0.921	0.977	0.936	0.947									
$h = 4$	0.985	1.016	1.025	1.223	0.968	0.965	0.973	0.966	0.965									
$h = 8$	0.994	1.016	1.020	1.386	0.962	0.962	0.984	1.013	0.992									
$h = 12$	0.986	1.027*	1.035	1.499*	0.974	0.976	0.986	1.001	0.990									
CRPS					Canada										Germany			
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.994	0.935	0.953	1.053	0.990	0.880	0.943	0.902	0.894	0.997	0.913	0.961	1.169*	0.972	0.880*	0.974	0.933	0.947
$h = 4$	1.009	1.011	0.978*	1.167	0.989	0.971	0.943	0.956	0.951	1.008	0.989	0.991	1.332*	0.982	0.969	1.002	1.005	1.004
$h = 8$	1.005	0.994	0.996	1.314	1.020	1.017	0.981	0.986	0.993	1.008	0.997	0.998	1.432*	1.000	1.011	0.977	0.982	0.980
$h = 12$	1.005	0.989	0.989	1.390*	1.020	1.054	0.990	1.016	0.997	1.009	1.000	0.998	1.480*	0.995	1.043	0.980	1.002	0.987
					France										Italy			
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.994	0.968	0.980	1.065	1.002	0.947	0.964	0.946	0.932	0.994	0.994	1.023	0.978	1.062	0.995	1.002	1.005	0.996
$h = 4$	0.984	0.997	0.990	1.233	0.968	1.007	0.922	0.907	0.915	1.000	0.996	1.025	1.098	1.013	1.003	0.958	0.944	0.957
$h = 8$	0.974*	1.006	0.996	1.383	0.988	1.089*	0.915	0.918	0.916	0.998	1.013	1.034	1.121	1.019	1.029	0.951	0.961	0.952
$h = 12$	0.973*	0.997	0.996	1.493	0.997	1.133*	0.924	0.921	0.929	1.007	0.992	1.030	1.150	1.008	1.039	0.954	0.976	0.963
					Japan										UK			
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.998	0.983	0.993	1.125	1.010	0.955*	0.979	0.969	0.960	1.031	1.020	1.008	0.976	1.114	0.970	1.000	1.055	1.018
$h = 4$	0.991	0.986	0.996	1.251*	0.990	0.978	0.973	0.952	0.958	1.008	1.005	1.003	1.149	0.971	0.945	0.944	0.937	0.947
$h = 8$	1.012	0.993	0.997	1.316*	1.009	0.973	0.974	0.972	0.959	1.013	1.004	0.998	1.245	0.986	0.977	0.968	0.975	0.976
$h = 12$	1.008	0.992	0.994	1.294*	0.995	0.987	0.970	0.975	0.959	1.013	0.999	1.004	1.384*	1.009	1.030	0.990	1.030	1.007
					USA													
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG									
$h = 1$	0.991	0.987	1.009	1.127*	1.004	0.945	0.965	0.942	0.944									
$h = 4$	0.989	1.024	1.024	1.319*	0.973	0.976	0.973	0.978	0.976									
$h = 8$	0.987	1.024	1.022	1.506*	0.958	0.970	0.978	1.007	0.992									
$h = 12$	0.990	1.033	1.035	1.628*	0.971	1.020	0.967	0.997	0.977									

Notes: The models are detailed in Table 1.2. For each specification, the upper panel presents the ratios of RMSFEs relative to the CVAR benchmark. The lower panel presents the ratios of CRPSs relative to the CVAR benchmark. Values below 1 indicate the model outperforms the benchmark and vice versa. Gray shading indicates the best performing model. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed-smoothing asymptotics as in Coroneo and Iacone [2020]. Differences in accuracy that are statistically different from zero are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 1.7: Out-of-sample output growth forecast performance: RMSFE and CRPS ratios in terms of CVAR benchmark, selected horizons

RMSFE	Canada									Germany								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.000	1.005	1.019	1.037	0.979	0.998	0.978	0.962	0.965	0.999	0.996	0.988	1.136*	0.996	1.062	1.004	1.023	1.020
$h = 4$	1.000	1.019	1.037	1.063	0.931	0.996	0.937	0.965	0.956	1.010	0.950	0.955	1.237*	0.947	1.010	0.968	1.004	0.982
$h = 8$	1.001	1.053	0.990	1.134	0.934	1.086	0.901	1.024	0.934	1.027	0.914	0.925	1.580*	0.929*	1.020	0.934	0.985	0.962
$h = 12$	0.978	1.029	0.955	1.228	0.932	1.185	0.845	0.998	0.887	1.017	0.900	0.899	1.937	0.934	1.091	0.897*	0.964	0.928
	France									Italy								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.003	0.999	1.006	1.039	0.998	1.048	0.996	1.019	0.999	1.004	0.998	1.014	1.040	1.005	1.021	1.010	1.030	1.002
$h = 4$	1.017	1.029	1.011	1.029	0.949	0.988	0.938	0.996	0.955	1.005	1.047	1.036	1.122	1.014	1.048	0.979	1.059	0.989
$h = 8$	1.009	1.058	1.011	1.252	1.006	1.041	0.932	1.058	0.971	0.980	1.081	1.049	1.261	1.090	1.177*	0.914	1.151	0.970
$h = 12$	1.017	1.082	0.990	1.535*	1.032	1.091	0.912	1.151	0.987	0.971	1.090	1.017	1.386	1.117	1.318	0.852	1.199*	0.952
	Japan									UK								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.997	0.989	0.985	1.138*	0.964	1.059	0.980	0.998	0.978	1.026	1.024	1.015	1.005	1.006	1.079	1.035	1.121	1.092
$h = 4$	1.004	1.026	0.978	1.115*	0.973	1.045	0.969	1.020	0.985	1.051	1.126	1.078	1.131	1.116*	1.192	1.130	1.363*	1.246
$h = 8$	1.008	1.026	0.968	1.256*	0.960	1.055*	0.947	1.007	0.962	1.058	1.170	1.053	1.235	1.196	1.292	1.088	1.407*	1.179
$h = 12$	1.003	1.000	0.963	1.369*	0.952	1.042	0.943	1.006	0.954	1.038	1.190	1.008	1.285	1.276	1.444*	1.036	1.489	1.151
	USA																	
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG									
$h = 1$	1.016	0.970	0.976	1.060	0.955	0.953	0.967	0.973	0.961									
$h = 4$	1.021	0.993	0.960	1.139*	0.975	0.943	0.973	1.010	0.982									
$h = 8$	1.021	1.045	0.976	1.297*	1.024	1.018	0.980	1.040	0.993									
$h = 12$	0.999	1.048	0.980	1.500*	1.018	1.021	0.967	1.017	0.966									

CRPS	Canada									Germany								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.002	1.010	1.014	1.040	0.979	0.995	0.987	0.972	0.972	0.989	0.991	0.980	1.122*	0.982	1.057*	0.999	1.010	1.011
$h = 4$	1.003	0.991	0.993	1.088*	0.911*	0.990	0.933	0.928	0.939	1.002	0.953	0.954	1.214*	0.944	1.029	0.968	0.995	0.981
$h = 8$	1.019	1.001	0.945	1.186*	0.902*	1.086	0.886	0.978	0.910	1.016	0.900	0.916	1.534*	0.926*	1.044	0.937	0.974	0.955
$h = 12$	1.008	0.987	0.916*	1.304*	0.889*	1.186	0.849*	0.956	0.873*	1.020*	0.899*	0.893*	1.925*	0.932*	1.183	0.906	0.966	0.931
	France									Italy								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.002	1.003	1.009	1.041	0.986	1.047	0.986	1.014	0.988	0.992	1.018	1.006	1.055*	1.013	1.030	1.019	1.069	1.023
$h = 4$	1.020	1.024	1.010	1.010	0.944*	1.006	0.930*	0.996	0.944	1.006	1.068	1.035	1.166*	1.004	1.049	0.999	1.074	1.009
$h = 8$	1.011	1.039	0.996	1.201	0.987	1.102*	0.913	1.017	0.942	0.989	1.083	1.045	1.279*	1.073	1.202*	0.919	1.115	0.961
$h = 12$	1.018	1.053	0.955*	1.469*	0.995	1.251*	0.888	1.101	0.950	0.980	1.083	0.995	1.380*	1.077	1.388*	0.847	1.142*	0.917
	Japan									UK								
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.994	0.992	0.985	1.146*	0.978	1.053	0.970	1.012	0.975	1.036	1.037	1.012	1.032	1.046	1.110*	1.030	1.114	1.073
$h = 4$	1.011	1.024	0.980	1.208*	0.998	1.070*	0.961	1.026	0.976	1.043*	1.091	1.040	1.142*	1.128	1.200*	1.077	1.302*	1.170*
$h = 8$	1.017	1.034	0.968	1.337*	0.989	1.098*	0.941	1.021	0.959	1.069	1.117	1.017	1.266*	1.181	1.319*	1.052	1.334*	1.128
$h = 12$	1.004	1.016	0.966	1.448*	0.983	1.156*	0.940	1.032	0.955	1.019	1.141	0.986	1.359*	1.187	1.489*	1.027	1.377*	1.127
	USA																	
	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG									
$h = 1$	1.029	0.988	0.984	1.098	0.987	0.992	0.993	1.029	0.996									
$h = 4$	1.047*	1.025	0.963	1.211*	0.993	0.957	0.971	1.039	0.990									
$h = 8$	1.048	1.052	0.962	1.379*	1.008	0.995	0.968	1.055	0.993									
$h = 12$	1.037	1.048	0.954	1.661*	0.994	1.059	0.940	1.022	0.960									

Notes: The models are detailed in Table 1.2. For each specification, the upper panel presents the ratios of RMSFEs relative to the CVAR benchmark. The lower panel presents the ratios of CRPSs relative to the CVAR benchmark. Values below 1 indicate the model outperforms the benchmark and vice versa. Gray shading indicates the best performing model. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed-smoothing asymptotics as in Coroneo and Iacone [2020]. Differences in accuracy that are statistically different from zero are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 1.8: Out-of-sample inflation forecast performance: RMSFE and CRPS ratios in terms of CVAR benchmark, selected horizons

RMSFE

Canada	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.993	0.994	1.044	1.036	1.084	1.210	1.021	1.075	1.044
$h = 4$	0.968	1.025	1.042	0.938	1.141	1.382	0.987	1.044	1.038
$h = 8$	0.961	1.042	1.040	0.815	1.147	1.708*	0.912*	0.985	0.970
$h = 12$	0.956	1.016	1.035	0.834	1.087	2.147*	0.830*	0.925	0.875*

Germany	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.011	0.939	0.939	1.065	0.980	1.005	0.946	0.947	0.929
$h = 4$	1.059	0.853*	0.826*	1.018	0.851*	0.839	0.791*	0.739*	0.774*
$h = 8$	1.080*	0.802*	0.777*	1.094	0.829*	0.953	0.727*	0.736*	0.775*
$h = 12$	1.067*	0.776*	0.763*	1.213	0.863*	1.201	0.700*	0.776*	0.789*

France	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.954	0.937	0.922	0.931	0.906	1.087	0.932	0.945	0.896
$h = 4$	0.983	0.929	0.881	0.822	0.863*	1.183	0.856	0.902	0.833
$h = 8$	0.989	0.928	0.878	0.767	0.886*	1.435	0.783	0.832*	0.778*
$h = 12$	0.986	0.966	0.900	0.883	0.966	1.918	0.777	0.908*	0.821*

Italy	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.991	0.965	0.971	1.044	0.942	1.136	0.948	1.063	0.959
$h = 4$	0.988	0.975	0.953	0.892	0.935	1.053	0.839	0.795	0.811
$h = 8$	0.953	0.971	0.945	0.741	0.972	1.106	0.779	0.762	0.763
$h = 12$	0.942	1.000	0.955	0.689	1.030	1.231	0.777	0.855	0.793

Japan	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.997	0.966	0.986	1.356	1.002	1.243	0.923	1.015	0.960
$h = 4$	1.025	0.924	0.943	0.753	0.971	0.841	0.852	0.900	0.850
$h = 8$	1.030	0.906	0.903	0.622	0.944	0.893	0.816	0.912	0.818
$h = 12$	1.018	0.909	0.891	0.789	0.954	1.280	0.818	0.979	0.847

UK	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.976	0.964	0.984	1.171	1.182	1.598*	0.951	1.146	0.986
$h = 4$	1.001	1.015	1.024	1.110	1.218	1.972*	0.994	1.048	1.000
$h = 8$	1.014	1.061	1.078	1.101	1.269*	2.661*	1.011	1.088	1.028
$h = 12$	1.017	1.061*	1.111	1.171	1.277*	3.521*	0.990	1.124	1.007

USA	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.003	1.091	1.032	1.257*	1.265	1.311	1.073	1.209*	1.123
$h = 4$	0.981	1.111	1.061*	1.089	1.154	1.017	1.024	1.045	1.053
$h = 8$	0.971	1.108	1.056*	0.995	1.105	0.876	0.992	0.975	0.985
$h = 12$	0.964	1.065	1.035*	1.036	1.034	0.856	0.947	0.902	0.915

CRPS

Canada	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.010	1.013	1.041	1.027	1.056	1.507*	1.013	1.075	1.042
$h = 4$	0.976	0.995	1.016	0.937	1.081	1.389*	0.953	0.979	0.986
$h = 8$	0.952	1.015	1.053	0.822	1.132	1.455*	0.877*	0.915	0.906*
$h = 12$	0.939*	1.004	1.084	0.804	1.100	1.584*	0.814*	0.873	0.844*

Germany	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.020	0.948	0.943	1.123	0.959	1.078	0.925	0.976	0.923
$h = 4$	1.063*	0.863*	0.817*	1.032	0.839*	0.882	0.770*	0.739*	0.762*
$h = 8$	1.066*	0.796*	0.761*	1.002	0.809*	0.877	0.690*	0.705	0.711*
$h = 12$	1.050*	0.761*	0.747*	1.053	0.863	1.008	0.677*	0.747*	0.721*

France	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.993	0.950	0.921*	1.026	0.911*	1.410*	0.908	0.957	0.887
$h = 4$	0.990	0.923	0.854*	0.853	0.834*	1.165	0.829*	0.847*	0.813*
$h = 8$	0.983	0.930	0.858*	0.752	0.873*	1.205	0.771*	0.784*	0.757*
$h = 12$	0.984	0.983	0.897	0.829	0.979	1.434*	0.775*	0.860*	0.795*

Italy	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	1.023	0.958	0.965	1.035	0.934	1.135	0.934	1.053	0.948
$h = 4$	1.021	0.983	0.962	0.901	0.928	1.006	0.838	0.782	0.803
$h = 8$	0.975	0.981	0.949	0.737	0.974	1.010	0.785	0.754	0.761
$h = 12$	0.954	1.030	0.979	0.675	1.051	1.098	0.774	0.833	0.787

Japan	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.987	0.998	0.980	1.624*	1.077	1.477*	0.980	1.146	1.041
$h = 4$	1.016	0.959	0.958	1.087	1.005	1.054	0.912	1.022	0.939
$h = 8$	1.011	0.938	0.933	0.926	0.972	1.084	0.867	0.994	0.891
$h = 12$	0.973	0.910	0.905	1.065	0.957	1.323	0.831	0.994	0.868

UK	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.984	1.005	1.004	1.405*	1.255*	1.884*	0.987	1.228*	1.043
$h = 4$	1.002	1.038	1.055*	1.175	1.260	1.789*	1.026	1.134	1.055
$h = 8$	1.017	1.080	1.108*	1.146	1.360*	2.073*	1.036	1.144	1.070
$h = 12$	1.021	1.085*	1.140*	1.150	1.386*	2.321*	1.019	1.172	1.054

USA	CVAR-H	CFAVAR	GVAR	CC	MIN	SSSS	HS	NG	NGG
$h = 1$	0.996	1.118	1.040	1.383*	1.313*	1.358*	1.069	1.209*	1.115
$h = 4$	0.968	1.167	1.071*	1.132	1.207	1.054	1.047	1.053	1.074
$h = 8$	0.960	1.189	1.067	0.994	1.174	0.961	1.041	1.008	1.034
$h = 12$	0.946	1.127	1.044	0.982	1.076	0.999	0.981	0.910	0.944

Notes: The models are detailed in Table 1.2. For each specification, the upper panel presents the ratios of RMSFEs relative to the CVAR benchmark. The lower panel presents the ratios of CRPSs relative to the CVAR benchmark. Values below 1 indicate the model outperforms the benchmark and vice versa. Gray shading indicates the best performing model. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed-smoothing asymptotics as in Coroneo and Iacone [2020]. Differences in accuracy that are statistically different from zero are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 1.9: Out-of-sample interest rate forecast performance: RMSFE and CRPS ratios in terms of CVAR benchmark, selected horizons

1.6.3 Investigating forecast performance over time

We have so far conducted a comprehensive evaluation of how different model specifications and prior choices affect forecast accuracy in the multi-country context, finding that multi-country VARs with hierarchical priors, in particular, the HS prior, are very helpful in forecasting inflation, as well as output growth and the interest rate. To get a better understanding of the source of the gains, we evaluate the models' forecasting performance over time. We plot in Figures 1-6 the cumulative sums of both RMSFEs and CRPSs at the selected horizons of 1,4,8, and 12 periods over the evaluation sample, averaged (arithmetic mean) across countries. Different colors with corresponding markers indicate different model specifications. The most recent Great Recession-financial crisis period (2007Q1-2009Q4) is highlighted in grey. For illustration, we only report results obtained from the benchmark and multi-country VARs with the three different scale mixtures of Normals priors (HS, NG, and NGG).

We first examine results obtained for output growth (Figures 1-2). The Great Recession-financial crisis clearly has a large effect on RMSFE and CRPS accuracy; all models' cumulative RMSFEs and CRPSs markedly increase after 2008. Before the crisis, the single-country CVAR benchmark performs similarly to multi-country VARs with hierarchical shrinkage and even does slightly better at long horizons. However, hierarchical shrinkage applied to multi-country VARs tends to be more beneficial after the crisis, which is particularly evident in the 4- and 8-steps-ahead density forecasts. Overall, in these aggregated measures, the NG specification is the best at short horizons, but the HS specification is better at long horizons, for both point and density forecasts.

When we compare the performance for inflation forecasts (Figures 3-4), there are several differences. First, compared to the results for output growth, we do not see as sharp an increase in cumulative RMSFEs and CRPSs during and after the financial crisis. There is some increase, but not as large as in the case of output growth forecasts. Second, when averaged across countries, before the crisis these models' forecasting performance is very similar at the 1-step-ahead horizon; after the crisis, the NG prior specification is slightly less accurate than the others. However, as the forecast horizon increases, the multi-country VAR with the HS prior becomes relatively more accurate, in both point and density forecasts. The light-tailed NG prior is clearly the worst among the three different scale mixtures of Normals priors and even worse than the single-country CVAR benchmark, particularly at longer horizons.

Moving to interest rate forecasts (Figures 5-6), the effects of the Great Recession-financial crisis are clear at multi-step forecast horizons, but less dramatic than for output growth forecasts. As interest rates in all G7 countries hit their effective lower bound, all models have difficulties in capturing the abrupt changes in short-term interest rates. At the 1-step-ahead horizon, the performance of the models is broadly similar, with the exception of the light-tailed NG prior; although comparable to others, the HS specification has slightly better accuracy. As the forecast horizon increases, the benefits obtained from the multi-country VARs with hierarchical priors become more evident; these specifications clearly outperform the single-country CVAR benchmark. The HS prior is better than the more flexible NGG prior, while the NG prior is clearly the worst among the three scale mixtures of Normals priors.

To conclude, we confirm that hierarchical shrinkage in multi-country VARs, especially coupled with the HS prior, delivers more accurate and robust forecasts over time for all three target variables. In these results aggregated across countries, for output growth, the NG prior is more preferable at short horizons, but the HS prior does better at long horizons. Gains are mainly obtained in the post-crisis evaluation period. For inflation, gains from the multi-country VAR with the HS prior are more evident as the forecast horizon increases. For the short-term interest rate, all models show difficulties in obtaining accurate forecasts as the horizon increases. Hierarchical shrinkage in multi-country VARs is generally better than the single-country benchmark as the horizon increases, and the HS prior tends to be more beneficial than the other scale mixtures of Normals priors.

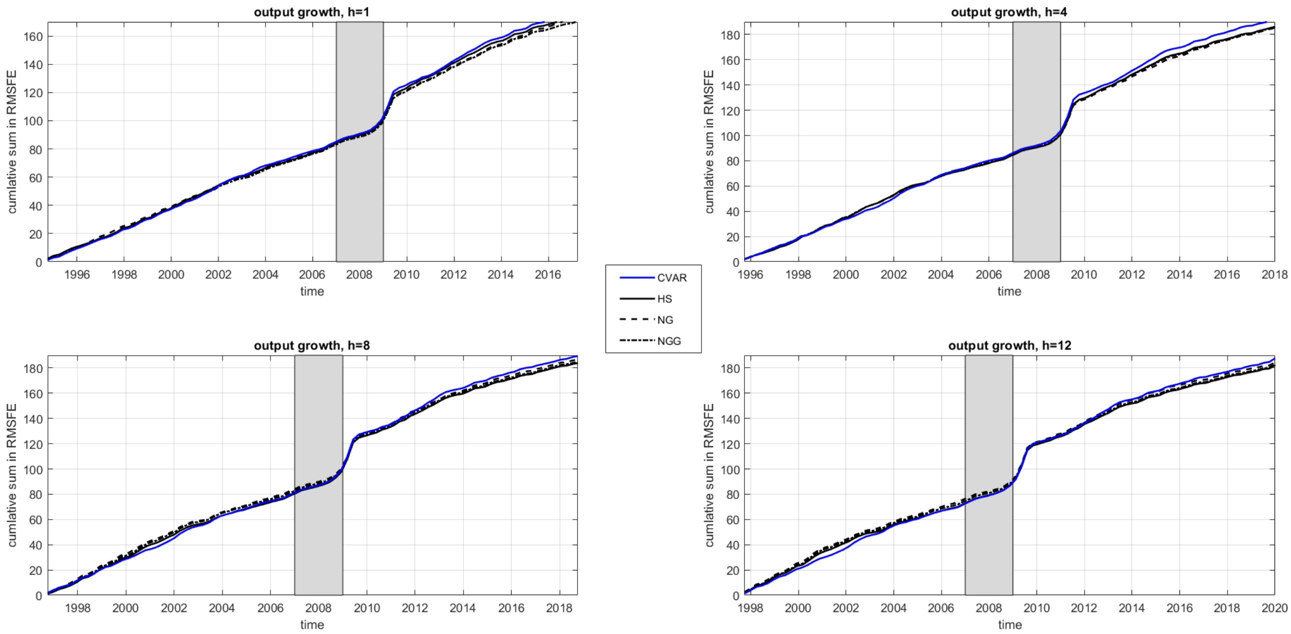

Figure 1.1: The figure presents cumulative sums (taken over time and averaged across countries) of RMSFEs for output growth forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 1.2.

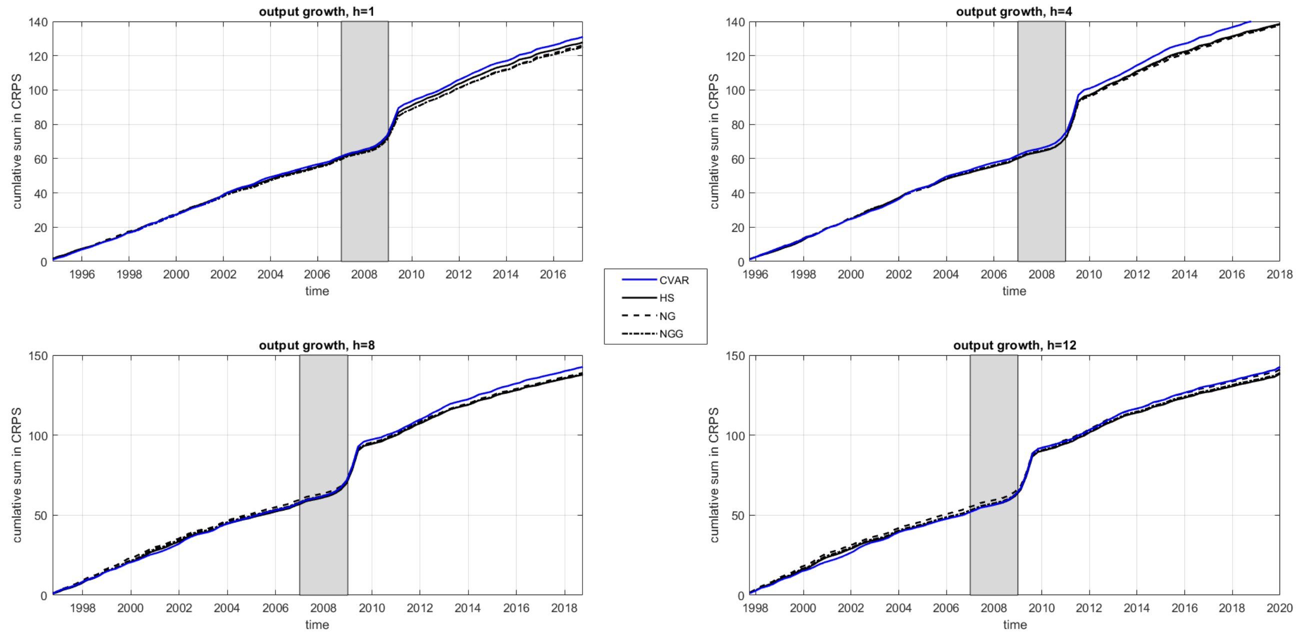

Figure 1.2: The figure presents cumulative sums (taken over time and averaged across countries) of CRPSs for output growth forecast at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 1.2.

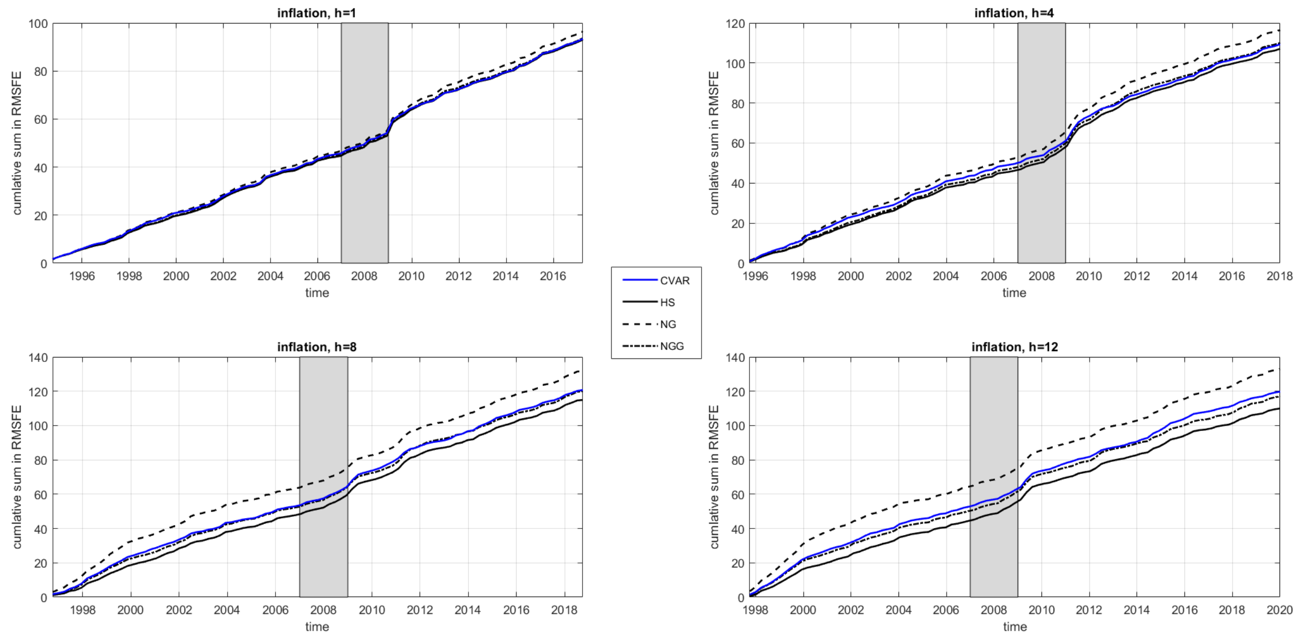

Figure 1.3: The figure presents cumulative sums (taken over time and averaged across countries) of RMSFEs for inflation forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 1.2.

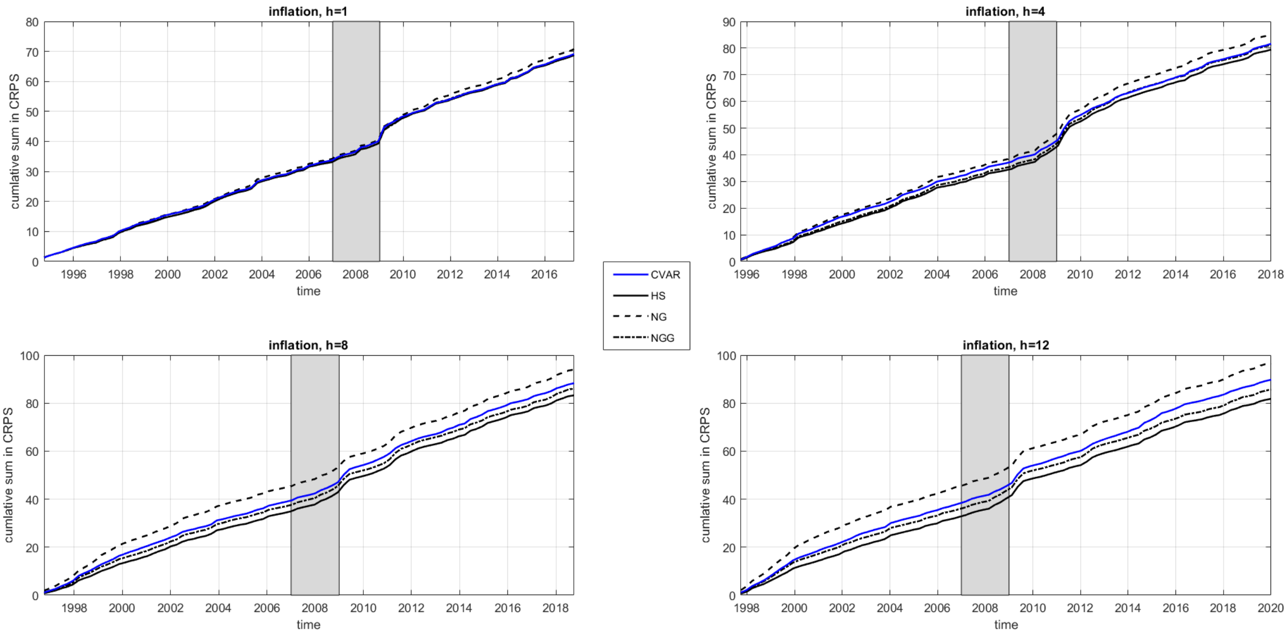

Figure 1.4: The figure presents cumulative sums (taken over time and averaged across countries) of CRPSs for inflation forecast sat selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 1.2.

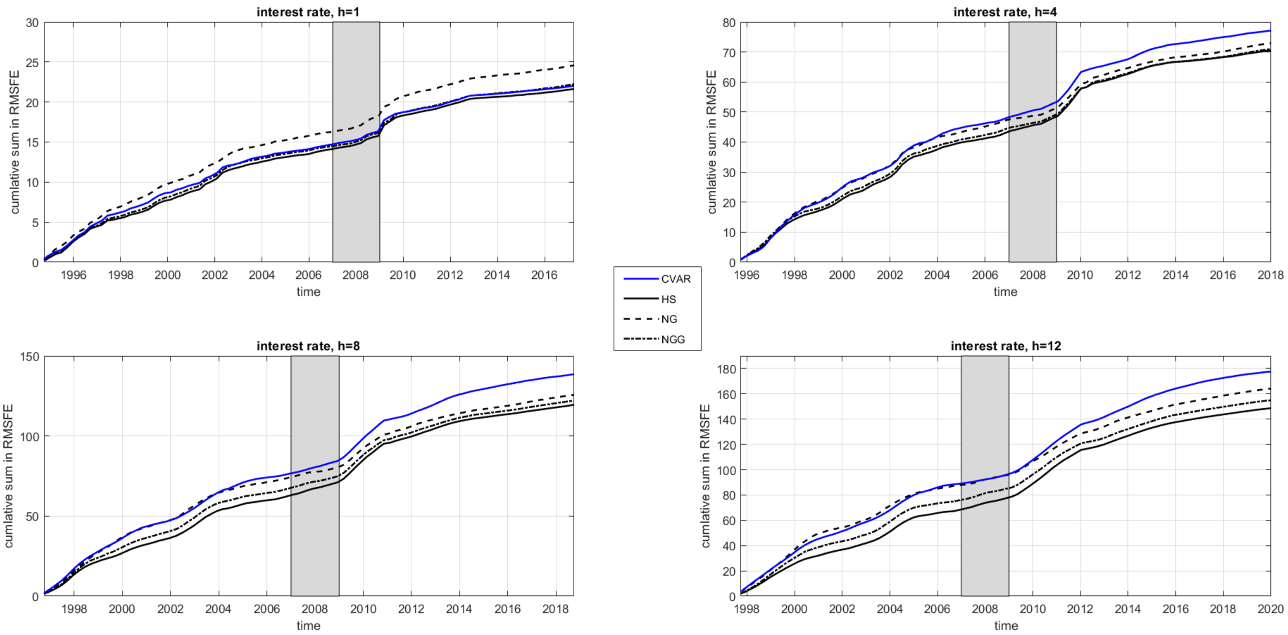

Figure 1.5: The figure presents cumulative sums (taken over time and averaged across countries) of RMSFEs for interest rate forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 1.2.

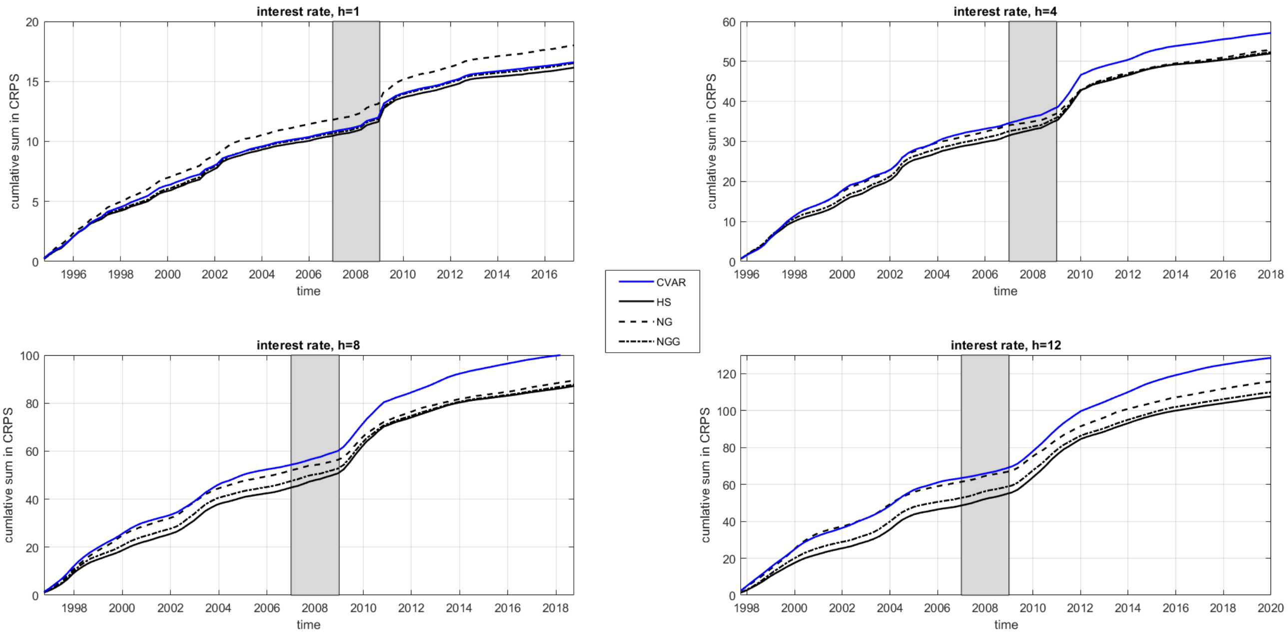

Figure 1.6: The figure presents cumulative sums (taken over time and averaged across countries) of CRPSs for interest rate forecasts at selected horizons: $h = 1, 4, 8, 12$. The models are detailed in Table 1.2.

1.6.4 Some robustness checks

In this subsection, we conduct several robustness checks of our main results presented above. We focus on the multi-country VAR with the Horseshoe prior (the HS specification) as overall it delivers the best forecast performance. In the interest of space, we briefly summarize the main findings; the details of results can be found in Appendix D.

Prior grouping of coefficients. As a check of the baseline prior's grouping of coefficients, we consider three alternative groupings of coefficients compared to the one used in the main results. First, we group all coefficients related to CSH together (HS-CSH) and assume that the priors of elements in vec(β^{CSH}) follow (1.9). Second, we do not make any attempt to search for restrictions but instead group both the intercepts and all autoregressive coefficients together (HS-A) and assume that all coefficients follow the same prior distributions as in (1.9). This has been examined in the single-country Bayesian VAR context by Cross et al. [2020], who find that scale mixtures of Normals priors do not improve the forecasting accuracy compared to conventional Minnesota priors. Finally, we group coefficients based on equations (HS-E) by assuming that priors for coefficients in each equation of (1.1) are the same. The equation-based shrinkage priors are more often seen in single-country Bayesian VARs. Follett and Yu [2019] and Huber and Feldkircher [2019] use this type of specification for the HS and NG priors, respectively. Cadonna et al. [2020] also propose an equation-wise specification for single-country time-varying parameter VARs.

In Tables 1.10-1.12, we report summary statistics on the percentage of gains for the alternative specifications described above compared to the specifications we use in the main results for all horizons, short horizons ($h \leqslant 6$), and long horizons ($h > 6$). The results can be summarized as follows. First, the HS-CSH, HS-A, and HS-E priors all improve forecast accuracy for the interest rate, especially at long horizons, with average percentage gains of more than 10 percent. Even though neither HS-A nor HS-E consider the underlying structure of model parameters, for the interest rate they forecast better than the HS-CSH specification. Second, these three alternatives are not helpful in forecasting output growth. The average gains are all negative, and they lead to loss of forecast accuracy in more than 75 percent of all cases. The HS-A prior has the overall worst forecast performance for output growth, and while output growth forecast performance from HS-CSH and HS-E is roughly similar, these approaches are outperformed by the HS specification used in the main results. Third, for inflation, the HS-CSH prior delivers forecast performance similar to our HS specification. The average gains are close to zero; in around half of all cases, the gains are positive. However, the HS-A and HS-E specifications generally reduce forecast accuracy (with HS-A the worst alternative), with impacts that are negative in more than 70 percent of all cases, and average losses of roughly 2 percent. Overall, these results suggest that it is not possible to improve our overall results by modifying our baseline grouping of coefficients.

Stochastic volatility. As another check, we also assess whether stochastic volatility is useful to improve forecast accuracy in the multi-country context. We modify the distributional assumption of u_t in (1.2) by assuming that $u_t \sim i.i.d.\ N(0, \Sigma)$. We specify a Normal prior for the VAR's coefficients and an Inverse Wishart prior for Σ and use the corrected triangular algorithms proposed in Carriero et al. [2021] to estimate and forecast. In Table 1.13, we provide summary statistics on the percentage differences in accuracy of forecasts from the models with and without SV. The results clearly indicate the usefulness of stochastic volatility. The constant volatility models are outperformed by stochastic volatility models for all horizons and all target variables. The benefits are particularly evident in forecasting inflation and the interest rate. For inflation, introducing stochastic volatility in multi-country VARs delivers gains in all cases at long horizons, and average gains are large: 34 percent for point forecasts and 27 percent for density forecasts. For the interest rate, stochastic volatility is more beneficial at short horizons. The average gains are around 12 (20) percent for point (density) forecasts, and gains are positive in nearly 80 (85) percent of cases for point (density) forecasts.

Estimation scheme. There is a long debate on the relative forecast performance of rolling and expanding window (recursive scheme) estimation in the forecasting literature. While rolling window estimates can be more robust to structural breaks, expanding window parameter estimates can be more efficient, helping forecast precision. In Table 1.14, we compare point and density forecasts from rolling and recursive schemes, taking as a benchmark the recursive scheme used in the paper's main results. The rolling scheme results use a window of 22 years of data, corresponding to the size of the sample used to generate the first forecast observation in our main results. On average, the rolling scheme forecasts are slightly better than the recursive forecasts, but the two methods perform broadly similarly. Compared to the recursive baseline, average gains to the rolling scheme are small and generally not statistically significant. By looking at the percentage of cases in which a given method outperforms the other, it appears that the rolling scheme does relatively better for inflation, for point forecasts for output growth, and interest rate forecasts at long horizons.

Univariate forecasting benchmarks. To understand the relative merits of our models with respect to univariate models, which are often tough benchmarks in the forecasting literature, we compare forecasts from the HS specification to those from univariate models with SV. We choose AR(p) models for output growth and the interest rate, with $p = 2$ and 4 lags, respectively, following Clark and Ravazzolo [2015]. For inflation, we choose an unobserved component model with SV as in Chan [2018]. Table 1.15 provides summary statistics for these accuracy comparisons. In these results, the multi-country HS specification yields more accurate forecasts of inflation and the interest rate. The average gains for the interest rate exceed 7 percent. Gains are positive in more than 90 percent of cases for density forecasts of inflation. For output growth, average gains from the HS specification are small but still positive. Overall, our main results based on a single-country VAR baseline still obtain when the baseline is changed to common univariate models.

Effective lower bound on interest rates. Since the 2007-2009 financial crisis put interest rates in all G7 countries at the effective lower bound for a number of years, one concern is that our reduced-form VAR models may forecast interest rates to be much higher than actual rates.[10] In Figures 1.7 and 1.8, we present point forecasts and associated 95 percent interval forecasts of the interest rate for all G7 countries obtained from the multi-country VAR with the Horseshoe prior (the HS specification) at horizons of 1 and 12 steps, respectively. We find that the ELB does not seem to be a major concern for short horizon (1-step-ahead) forecasts, as our model is able to track the true interest rate rather well even during the ELB period. However, some bias in the forecasts emerges during the ELB period when we look at 12-steps-ahead forecasts. Our model generally predicts the interest rate to be higher than the actual rate. While more evident in the ELB period, the problem is present even before the ELB period. However, in the case of Japan, we see that our model forecasts the interest rate to be much higher than the true value early in the sample, but the forecasts gradually decline and are able to track the realized values fairly well for much of the sample. We conclude from these results that the ELB has some impact on our interest rate forecasting results — as it likely will for most any VAR — but does not necessarily entirely distort them.

Directional forecasts. Finally, to provide a rough gauge on whether our forecasting models are also able to predict turning points, we use the nonparametric test developed by Pesaran and Timmermann [1992] to assess directional forecast performance for (1-step ahead) changes of output growth. In Table 1.16, we report test statistics and associated p-values for all 7 countries from both the multi-country VAR-SV model with the Horseshoe prior and the single-country VAR-SV benchmark. The results show that, except for Canada, for multi-country VAR-SV with Horseshoe prior, the test strongly rejects the null of no predictability. However, for the single-country VAR-SV benchmark, we cannot reject the null for all 7 countries. This provides clear evidence that the multi-country VAR-SV model with Horseshoe prior has

[10] Japan hit the ELB earlier than other countries. The short-term interest rate in Japan remains around zero in the entire forecast evaluation period.

better predictive power also for the directional forecasts of output growth.

1.7 Conclusions

In this paper, we use hierarchical shrinkage in multi-country Bayesian VARs and examine its macroeconomic forecasting ability. To make the approach operational, we consider three different scale mixtures of Normals priors, namely the Horseshoe prior, the Normal-Gamma prior, and the Normal-Gamma-Gamma prior, which have been shown to benefit macroeconomic forecasting in single-country settings. We also provide some new theoretical results for the Normal-Gamma prior. Empirically, we compare the forecast accuracy with country-specific VARs, country-specific factor-augmented VARs, global VARs, and alternative shrinkage approaches for multi-country VARs that have been used in the macroeconomic forecasting literature. All of our models include stochastic volatility, which is helpful to forecast accuracy. We confirm the benefits from enlarging single-country information sets to include information across countries. Hierarchical shrinkage in the multi-country VAR model with the Horseshoe prior has the overall best forecast performance.

Appendix A: Proof of Theorem 1

In the following proof, the notation $\sim$ indicates asymptotic equivalence. We say that a is asymptotically equivalent to b if $a/b = O(1)$.

As shown in equation (14) of Bitto and Frühwirth-Schnatter [2019], the marginal density for $\beta_j \sim \mathcal{NG}(\lambda, \kappa)$, given λ, κ, can be expressed as

$$\pi_{NG}(\beta_j) = \frac{(\sqrt{\lambda\kappa})^{\lambda+\frac{1}{2}}}{\sqrt{\pi}2^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\left|\beta_j\right|^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}\left(\sqrt{\lambda\kappa}\left|\beta_j\right|\right),$$

where $K_p(\cdot)$ is the modified Bessel function of the second kind of index p. Let us first consider the concentration properties. If $\lambda > \frac{1}{2}$, then $\lambda - \frac{1}{2} > 0$. By 9.6.9 in Abramowitz and Stegun [1965], as $\left|\beta_j\right| \to 0$,

$$K_{\lambda-\frac{1}{2}}\left(\sqrt{\lambda\kappa}\left|\beta_j\right|\right) \sim \frac{1}{2}\Gamma(\lambda - \frac{1}{2})\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\frac{1}{2}-\lambda}.$$

Then,

$$\begin{aligned}\pi_{NG}(\beta_j) &\sim \frac{(\sqrt{\lambda\kappa})^{\lambda+\frac{1}{2}}}{\sqrt{\pi}2^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\left|\beta_j\right|^{\lambda-\frac{1}{2}} \times \frac{1}{2}\Gamma(\lambda - \frac{1}{2})\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\frac{1}{2}-\lambda} \\ &= \frac{\sqrt{\lambda\kappa}}{\sqrt{\pi}}\frac{\Gamma(\lambda-\frac{1}{2})}{\Gamma(\lambda)} \times \frac{1}{2} = O(1).\end{aligned}$$

If $0 < \lambda < \frac{1}{2}$, recall that

$$K_{\lambda-\frac{1}{2}}\left(\sqrt{\lambda\kappa}\left|\beta_j\right|\right) = \frac{1}{2}\pi\frac{I_{\frac{1}{2}-\lambda}(\sqrt{\lambda\kappa}\left|\beta_j\right|) - I_{\lambda-\frac{1}{2}}(\sqrt{\lambda\kappa}\left|\beta_j\right|)}{\sin\left((\lambda-\frac{1}{2})\pi\right)},$$

where $I_p(\cdot)$ is the modified Bessel function of the first kind with index p. By 9.6.7 in Abramowitz and Stegun [1965], as $\left|\beta_j\right| \to 0$,

$$\begin{aligned}I_{\frac{1}{2}-\lambda}(\sqrt{\lambda\kappa}\left|\beta_j\right|) &\sim \frac{\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\frac{1}{2}-\lambda}}{\Gamma(\frac{3}{2}-\lambda)} \\ I_{\lambda-\frac{1}{2}}(\sqrt{\lambda\kappa}\left|\beta_j\right|) &\sim \frac{\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\lambda-\frac{1}{2}}}{\Gamma(\frac{3}{2}-\lambda)}.\end{aligned}$$

Thus,

$$\begin{aligned}\pi_{NG}(\beta_j) &\sim \frac{(\sqrt{\lambda\kappa})^{\lambda+\frac{1}{2}}}{\sqrt{\pi}2^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\left|\beta_j\right|^{\lambda-\frac{1}{2}} \times \frac{1}{2}\pi \times \frac{1}{\sin((\lambda-\frac{1}{2})\pi)} \times \left(\frac{\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\frac{1}{2}-\lambda}}{\Gamma(\frac{3}{2}-\lambda)} - \frac{\left(\frac{1}{2}\sqrt{\lambda\kappa}\left|\beta_j\right|\right)^{\lambda-\frac{1}{2}}}{\Gamma(\frac{3}{2}-\lambda)}\right)\\ &= C - \frac{(\sqrt{\lambda\kappa})^{2\lambda} \times (\frac{1}{2})^{2\lambda}\sqrt{\pi}}{\sin((\lambda-\frac{1}{2})\pi)\Gamma(\lambda)\Gamma(\lambda+\frac{1}{2})}\Big(\frac{1}{\left|\beta_j\right|}\Big)^{1-2\lambda}\\ &= O\left(\Big(\frac{1}{\left|\beta_j\right|}\Big)^{1-2\lambda}\right).\end{aligned}$$

Finally, if $\lambda = \frac{1}{2}$, by 9.6.8 in Abramowitz and Stegun [1965],

$$K_0\Big(\sqrt{\frac{1}{2}\kappa}\left|\beta_j\right|\Big) \sim -\log\Big(\sqrt{\frac{1}{2}\kappa}\left|\beta_j\right|\Big).$$

Then,

$$\pi_{NG}(\beta_j) \sim \frac{\sqrt{\frac{1}{2}\kappa}}{\pi} \times -\log\Big(\sqrt{\frac{1}{2}\kappa}\left|\beta_j\right|\Big) = O\Big(\frac{1}{\log(\left|\beta_j\right|)}\Big).$$

We now move to the asymptotic tail behavior. By 9.7.2 in Abramowitz and Stegun [1965], as $\left|\beta_j\right| \to \infty$,

$$K_{\lambda-\frac{1}{2}}\Big(\sqrt{\lambda\kappa}\left|\beta_j\right|\Big) \sim \sqrt{\frac{\pi}{2\sqrt{\lambda\kappa}\left|\beta_j\right|}}e^{-\sqrt{\lambda\kappa}\left|\beta_j\right|}.$$

Then,

$$\begin{aligned}\pi_{NG}(\beta_j) &\sim \frac{(\sqrt{\lambda\kappa})^{\lambda+\frac{1}{2}}}{\sqrt{\pi}2^{\lambda-\frac{1}{2}}\Gamma(\lambda)}\left|\beta_j\right|^{\lambda-\frac{1}{2}}\sqrt{\frac{\pi}{2}}(\sqrt{\lambda\kappa})^{-\frac{1}{2}}\left|\beta_j\right|^{-\frac{1}{2}}\exp\Big(-\sqrt{\lambda\kappa}\left|\beta_j\right|\Big)\\ &= O\left(\frac{\left|\beta_j\right|^{\lambda-1}}{\exp(\sqrt{\lambda\kappa}\left|\beta_j\right|)}\right),\end{aligned}$$

which completes the proof.

Appendix B: Model specifications and priors

Appendix B.1: Country-specific VARs

The country-specific VAR(p) model — denoted the CVAR specification in the paper's results — is specified as

$$
\begin{aligned}
y_{i,t} &= c_i + \sum_{l=1}^{p} B_{i,l} y_{i,t-l} + u_{i,t} && (1.15)\\
u_{i,t} &= A_i^{-1} H_{i,t}^{0.5} \epsilon_{i,t}, \epsilon_{i,t} \sim i.i.d.\ N(0, I_G), && (1.16)
\end{aligned}
$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, and the dimension of $y_{i,t}$, $u_{i,t}$ and $\epsilon_{i,t}$ is $G \times 1$. A_i^{-1} is a lower triangular matrix with diagonal elements equal to 1. $H_{i,t}$ is diagonal with generic j-th element $h_{ij,t}$ evolving as a random walk (RW):

$$\ln h_{ij,t} = \ln h_{ij,t-1} + e_{ij,t}, j = 1, \ldots, G, \tag{1.17}$$

where $e_{it} \sim N(0, \Phi_i)$ with a full covariance matrix Φ_i as in Primiceri [2005].

Letting $B_i = [c_i, B_{i,1}, \ldots, B_{i,p}]'$, the priors are specified as:

$$
\begin{aligned}
vec(B_i) &\sim N(0, \underline{\Omega}_{B_i})\\
vec(A_i) &\sim N(0, \underline{\Omega}_{A_i})\\
\Phi_i &\sim IW(Q_0, W_0).
\end{aligned}
$$

For the prior variances of the autoregressive coefficient matrices, we set them as in the Minnesota prior:

$$\underline{\Omega}_{B_{i,l}^{(mn)}} = \begin{cases} \frac{\lambda_1}{l^{\lambda_3}} \frac{1}{\sigma_n^2} & \text{for the coefficients on own lags} \\ \frac{\lambda_2}{l^{\lambda_3}} \frac{\sigma_m^2}{\sigma_n^2} & \text{for the coefficients on cross-variable lags} \\ \lambda_0 \sigma_m^2 & \text{for the intercept,} \end{cases} \tag{1.18}$$

where $m, n = 1, \ldots, G$. λ_1 measures the overall tightness to coefficients related to own lags. λ_2 is related to cross-variable shrinkage. We assume Gamma priors for these two hyperparameters: $\lambda_1 \sim \mathcal{G}(1, 0.04)$, $\lambda_2 \sim \mathcal{G}(1, 0.04^2)$. λ_3 determines the additional shrinkage for coefficients associated with higher order lags and is set to 2 (quadratic decay). The scale parameters σ_m^2, σ_n^2 are obtained from univariate AR(1) regressions. We elicit an uninformative prior for the intercept by setting $\lambda_0 = 100$. In the case of the free elements in the contemporaneous matrix A_i, we set the prior mean to 0 and the prior variance to be non-informative: $\underline{\Omega}_{A_i} = 10 \times I$. Finally, as in the previous section, we follow the literature and set a modestly informative prior for Φ: $\Phi \sim IW(Q_0, W_0)$, where Q_0, W_0 take very conservative values: $W_0 = 0.01 \times I$ and $Q_0 = G + 2$.[11]

[11] See, e.g., D'Agostino et al. [2013] and Clark and Ravazzolo [2015].

For the country-specific VAR with hierarchical shrinkage (CVAR-H), we follow exactly the approach in Chan [2021]. Following Chan, the reduced-form model (1.15) is expressed in structural form

$$A_i y_{i,t} = \tilde{c}_i + \sum_{l=1}^{p} \tilde{B}_{i,l} y_{i,t-l} + H_{i,t}^{0.5} \epsilon_{i,t},$$

where $\tilde{c}_i = Ac_i$, $\tilde{B}_{i,l} = AB_{i,l}$, and the innovations $e_{ij,t}$ in (1.17) are assumed to be independent across variables (equation $j = 1, \ldots, G$ of the VAR for country i): $e_{ij,t} \sim N(0, \sigma^2_{e_{ij}})$. The priors are specified as

$$\beta_{i,j} | \lambda_1, \lambda_2, \psi_{i,j}, C_{i,j} \sim N(0, 2\lambda_{i,j}\psi_{i,j}C_{i,j}),$$

where $\lambda_{i,j}$ equals λ_1 if $\beta_{i,j}$ are related to own lags but equals λ_2 for coefficients related to cross-variable lags. $C_{i,j}$ are specified according to

$$C_{i,l} = \left\{ \begin{array}{ll} \frac{1}{l^{\lambda_3}} \frac{1}{\sigma_n^2} & \text{for the coefficients on own lags} \\ \frac{1}{l^{\lambda_3}} \frac{\sigma_m^2}{\sigma_n^2} & \text{for the coefficients on cross-variable lags} \end{array} \right\}$$

and $\psi_{i,j}$ are assumed to follow a Gamma prior:

$$\psi_{i,j} \sim \mathcal{G}(\nu_\psi, \nu_\psi/2),$$

with an additional hyper-prior on $\nu_\psi \sim \mathcal{G}(1, 1)$. For $\sigma^2_{e_{ij}}$, priors are assumed to be $\sigma^2_{e_{ij}} \sim \mathcal{IG}(5, 0.04)$.

Appendix B.2: Factor augmented country-specific VARs

The factor-augmented country-specific VAR (CFAVAR) takes the form:

$$\begin{aligned} \begin{bmatrix} y_{i,t} \\ F_t \end{bmatrix} &= c_i + \sum_{l=1}^{p} B_{i,l} \begin{bmatrix} y_{i,t-l} \\ F_{t-l} \end{bmatrix} + u_{i,t} \\ Y_t^* &= \Lambda F_t + \varepsilon_t \\ F_t &= \sum_{l=1}^{q} \Pi_l F_{t-l} + v_t, v_t \sim i.i.d.\ N(0, \Sigma_v), \end{aligned}$$

where $Y_t^* = (y'_{1,t}, \ldots, y'_{i-1,t}, y'_{i+1,t}, y'_{N,t})'$ is the collection of foreign variables. F_t is a $r \times 1$ vector of weakly exogenous unobservable factors representing foreign information, which affect the variables in country i via the loadings $B^*_{i,l}, i = 1, \ldots, N, l = 1, \ldots, p$. Factors are estimated (recursively, as forecasting moves forward in time and the estimation sample expands) by principal components (see, e.g., Stock and Watson [2002a] and Stock and Watson [2002b]) and assumed to follow a VAR process with lag length q. In the VAR for $[y_{i,t}, F_t]$, the innovation vector $u_{i,t}$ includes the stochastic volatility structure previously indicated in the country-specific VAR's equation (1.16).

Priors for c_i and $B_{i,l}$ are specified in the same way as in the country-specific VARs. The same hyperpriors are imposed on (λ_1, λ_2), which are the overall tightness parameters on coefficients related to own lags and cross-variable lags. We specify the maximum number of factors and lag length to be $r^{\max} = 4$ and $q^{\max} = 4$, respectively. The number of factors is determined by Bai and Ng [2002]'s $IC2$ information criterion, and the number of lags is determined by the Bayesian Information Criterion (BIC). The VAR for the factors is separately estimated by Bayesian methods with non-informative priors. Specifically, letting $\pi = vec([\Pi_1, \ldots, \Pi_q]')$, we specify $\pi \sim N(0, 100 \times I_{r^2 q})$. Following Korobilis [2016], $\hat{\Sigma}_v$ is fixed at the OLS estimate to streamline computations (it also eliminates the uncertainty associated with covariance matrix estimation).

Appendix B.3: Global VARs

A GVAR model consists of a number of country-specific equations that are combined to form a global model. Assuming that the global economy consists of $N + 1$ countries, in the first step, we estimate the following country-specific VARX model for every country $i = 0, 1, ..., N$:

$$y_{i,t} = c_i + \sum_{l=1}^{p} B_{i,l} y_{i,t-l} + \sum_{l=0}^{p^*} B^*_{i,l} y^*_{i,t-l} + u_{i,t}, \tag{1.19}$$

$$u_{i,t} = A_i^{-1} H_{i,t}^{0.5} \epsilon_{i,t}, \quad \epsilon_{i,t} \overset{i.i.d.}{\sim} N(0, I_G), \tag{1.20}$$

where $t = 1, ..., T$, $y_{i,t}$ is a $G \times 1$ vector of endogenous variables in country i, c_i is a $G \times 1$ vector of intercept terms, $B_{i,l} (l = 1, ..., p)$ denotes the $G \times G$ matrix of parameters associated with lagged endogenous variables and $B^*_{i,l} (l = 0, 1, ..., p^*)$ is the matrix of parameters associated with contemporaneous and lagged weakly exogenous variables. The weakly exogenous foreign variables $y^*_{i,t}$ are constructed as a weighted average of the endogenous variables in other countries:

$$y^*_{i,t} = \sum_{j=0}^{N} w_{i,j} y_{j,t} \tag{1.21}$$

and the weights satisfy the following two restrictions: $w_{i,i} = 0$ and $\sum_{j=0}^{N} w_{i,j} = 1$. Weights are constructed from standardized bilateral trade flows. The data is available from Mohaddes and Raissi [2018].

In the second step, $N + 1$ country-specific VARX models are stacked to form a global model, which is given by

$$G y_t = c + \sum_{q=1}^{Q} H_q y_{t-q} + u_t, \tag{1.22}$$

where $y_t = (y'_{1,t}, ..., y'_{N,t})'$, $Q = \max(p, p^*)$, and G and H_q are both $NG \times NG$ dimensional coefficient matrices. Details on how to solve the global model can be found in Pesaran et al. [2009] and Huber [2016].

Priors for c_i and $B_{i,l}$ are specified in the same way as in the CFAVAR. More specifically, c_i and $B_{i,l}$ follow the same specification as in (1.18). For the prior on the elements of $B^*_{i,l}$, means are set to zero and variances are defined as: $\lambda_4 \frac{\sigma_m^2}{\sigma_n^2}$, where σ_m^2, σ_n^2 are obtained from univariate AR(1) regressions. We assume a Gamma prior for $\lambda_4 \sim \mathcal{G}(1, 0.02^2)$. Both p and p^* are set to 4.

Appendix B.4: Multi-country VARs

Appendix B.4.1: Factor shrinkage approach

The factor shrinkage approach used with the CC specification relies on the VAR written in system form. We define $X_t = I_{NG} \otimes x_t'$, where $x_t = (1, Y_{t-1}', \ldots, Y_{t-p}')'$, β_i is the $k \times 1$ vector containing coefficients related to each i, $k = NGp + 1$, and $\beta = (\beta_1', \ldots, \beta_N')'$ is the $NGk \times 1$ vector containing all coefficients. Write the VAR as

$$Y_t = X_t \beta + u_t, \tag{1.23}$$

where $u_t \sim i.i.d.\ N(0, \Sigma_t)$.

Canova and Ciccarelli [2009] assume that the vector of coefficients β can be expressed as:

$$\beta = \sum_{i=1}^{F} \Xi_i \theta_i \tag{1.24}$$

where $\Xi = [\Xi_1, \ldots, \Xi_F]$ are known matrices and $\theta = (\theta_1', \ldots, \theta_F')'$ is a low dimensional vector ($\dim(\theta) < K$, where $K = kNG$) of unknown parameters, and $\theta_1, \ldots, \theta_F$ are mutually orthogonal.[12]

We consider the factorization used in Canova et al. [2007] and Canova and Ciccarelli [2013]. We assume $F = 4$. θ_1 is a scalar factor which is common across all countries, $\theta_2' = (\theta_{2,1}, \ldots, \theta_{2,N})'$ is a $N \times 1$ vector of country-specific factors, $\theta_3' = (\theta_{3,1}, \ldots, \theta_{3,G})'$ is a $G \times 1$ vector of variable-specific factors and $\theta_4' = (\theta_{4,1}, \ldots, \theta_{4,p-1})'$ is a $(p-1) \times 1$ vector of lag-specific factors.[13] $\Xi_1, \ldots, \Xi_4$ are assumed to be known with elements associated with the corresponding original parameters equal to 1 and 0 otherwise. For example, consider a multi-country VAR model in (1) with $N = 2, G = 2, p = 1$. In this case, Ξ_1 is a 20×1 vector of ones, and Ξ_2 and Ξ_3 take the form:

$$\underset{20\times 2}{\Xi_2} = \begin{bmatrix} \iota_1 & 0 \\ \iota_1 & 0 \\ 0 & \iota_2 \\ 0 & \iota_2 \end{bmatrix}, \quad \underset{20\times 2}{\Xi_3} = \begin{bmatrix} \iota_3 & 0 \\ 0 & \iota_4 \\ \iota_3 & 0 \\ 0 & \iota_4 \end{bmatrix},$$

[12] A more general form is $\beta = \sum_{i=1}^{F} \Xi_i \theta_i + e$, where $e \sim N(0, \Sigma \otimes \sigma^2 I)$ is an approximation error uncorrelated with u_t. However, most of the literature assumes an exact factorization ($\sigma^2 = 0$); see, for example, Canova et al. [2007], Canova and Ciccarelli [2009], Dées and Güntner [2017]. Koop and Korobilis [2019] estimate σ^2 by a forgetting factor approach and find that it is very small (< 0.01). In some limited checks, we found that considering the approximation error harms forecasting performance.

[13] To avoid collinearity with θ_1, θ_4 can contain at most $p - 1$ components.

where $\iota_1 = (0,1,1,0,0)'$, $\iota_2 = (0,0,0,1,1)'$, $\iota_3 = (0,1,0,1,0)'$, and $\iota_4 = (0,0,1,0,1)'$. Thus, we can rewrite (1.23) as:

$$\begin{aligned} Y_t &= X_t\beta + u_t \\ &= X_t(\Xi\theta) + u_t = \tilde{X}_t\theta + u_t. \end{aligned} \tag{1.25}$$

In this case, $\dim(\theta) = N + G + p$. By construction, the $\tilde{X}_t$'s are linear combinations of the original right-hand-side variables in (1.23), and the parameterization above can capture comovement across lagged variables.

To incorporate SV, we decompose Σ_t as $\Sigma_t = A^{-1}H_tA'^{-1}$, where A is lower diagonal with diagonal elements equal to 1, and the diagonal elements in H_t evolve according to (1.17).

We specify the priors for θ, A, and Φ as (independent), Normal, Normal, and Inverse Wishart, respectively:

$$\theta \sim N(0, \underline{\Omega}_\theta), \quad a \sim N(0, \underline{\Omega}_a), \quad \Phi \sim IW(Q_0, W_0), \tag{1.26}$$

where a denotes the vector of free elements in A. The prior mean for θ is set to zero, and the prior covariance matrix $\underline{\Omega}_\theta$ is assumed to be diagonal. Letting $\omega_{\theta_{i,j}}$ be the elements in $\underline{\Omega}_\theta$ associated with the jth elements in θ_i, where $i = 1, \ldots, 4$, then

$$\underline{\omega}_{\theta_{i,j}} = \left\{ \begin{array}{ll} \sum_{m=1}^{NG} \sigma_m^2 & i = 1,2,3 \\ \frac{\sum_{m=1}^{NG} \sigma_m^2}{l^2}, & i = 4, l = 2, \ldots, p \end{array} \right\}.$$

The prior mean for a is set to 0, and the prior variance is set to $\underline{\Omega}_a = 10 \times I$. Q_0, W_0 are specified as $Q_0 = NG + 2$, $W_0 = 0.01 \times I$.

Appendix B.4.2: Prior specifications for other models

For the approach in Angelini et al. [2019] and the hierarchical shrinkage considered in this paper, the prior for free elements in A is assumed to be Normal with zero mean and variance equal to $10 \times I_{NG}$. The prior for Φ takes the form $\Phi \sim IW(Q_0, W_0)$, and Q_0, W_0 are specified as $Q_0 = NG + 2$, $W_0 = 0.01 \times I$.

For the prior in (1.18), σ_i^2, σ_j^2 are obtained from univariate AR(1) regressions. The prior for the intercept is assumed to be uninformative by setting the prior variance equal to $100 \times \sigma_i^2$, where σ_i^2 is again from a univariate AR(1) regression. The hyper-priors on overall shrinkage parameters are specified in the same way as in country-specific VARs. For the additional hyperparameter λ_4 controlling the tightness for coefficients related to cross-variable lags for foreign countries, we use a prior of $\lambda_4 \sim \mathcal{G}(1, 0.02^2)$.

For the SSSS prior, we follow Korobilis [2016] exactly. For (1.7) and (1.8), we set $\xi_{ij}^2 = \tau_{ij}^2 = 4$ and

$\underline{c}^{\text{DI}} = \underline{c}^{\text{CSH}} = 0.0025$. The prior for indicators are specified as

$$\gamma_{ij}^{\text{DI}} \sim Bernoulli(\pi_{ij}^{\text{DI}}),\ \ \pi_{ij}^{\text{DI}} \sim \mathcal{B}(1,1)$$
$$\gamma_{ij}^{\text{CSH}} \sim Bernoulli(\pi_{ij}^{\text{CSH}}),\ \ \pi_{ij}^{\text{CSH}} \sim \mathcal{B}(1,1).$$

For the Horseshoe prior, no more prior specifications are needed. For the Normal-Gamma prior, recall that we specify $a^{\omega} \sim \mathcal{E}(b)$ and $\kappa^2 \sim \mathcal{G}(d_1, d_2)$. We set b equal to the number of coefficients in each block and elicit a non-informative prior for κ^2 by setting $d_1 = d_2 = 0.01$. For the Normal-Gamma-Gamma prior, recall that $2a \sim \mathcal{B}(\alpha_a, \beta_a)$, $2c \sim \mathcal{B}(\alpha_c, \beta_c)$, and we set $\alpha_a = \alpha_c = 2, \beta_a = \beta_c = 1$.

Appendix B.5: Univariate models

For AR(p)-SV models applied to each scalar output growth or interest rate variable, generally denoted y_t, we have

$$\begin{aligned} y_t &= c + \sum_{\ell=1}^{p} \rho_\ell y_{t-\ell} + u_t, \\ u_t &= h_t^{0.5} v_t,\ \ v_t \overset{i.i.d.}{\sim} N(0,1), \\ \log h_t &= \log h_{t-1} + e_t,\ \ e_t \overset{i.i.d.}{\sim} N(0, \sigma_e^2). \end{aligned}$$

As in Clark and Ravazzolo [2015], lag length is set to 2 for output growth and 4 for the interest rate. Letting $\theta = (c, \rho_1, \ldots, \rho_p)'$, we specify the following priors:

$$\theta \sim N(0, V),\ \ \sigma_e^2 \sim IG(v_h, S_h),\ \ \log h_0 \sim N(a_0, b_0).$$

V is assumed to be diagonal with elements equal to $\frac{\theta_1}{\ell^{\theta_2}}$, $\ell = 1, \ldots, p$, for autoregressive coefficients and $100 \times \hat{\sigma}_y^2$ for the intercept. θ_1 is set to 0.04, θ_2 is set to 2, and $\hat{\sigma}_y^2$ is obtained from a univariate AR(1) regression. We use a modestly informative prior for σ_e^2 to control the time variation by setting v_h equal to 2 and S_h to 0.04. For the prior on initial conditions, we set $a_0 = 0$ and $b_0 = 10$.

For UCSV model, we have

$$\begin{aligned} y_t &= \tau_t + \varepsilon_t^y,\ \ \varepsilon_t^y \sim N(0, e^{h_t}), \\ \tau_t &= \tau_{t-1} + \varepsilon_t^\tau,\ \ \varepsilon_t^\tau \sim N(0, e^{g_t}), \\ h_t &= h_{t-1} + \varepsilon_t^h,\ \ \varepsilon_t^h \sim N(0, \omega_h^2), \\ g_t &= g_{t-1} + \varepsilon_t^g,\ \ \varepsilon_t^g \sim N(0, \omega_g^2), \end{aligned}$$

with initial conditions τ_0, h_0 and g_0 as unknown parameters. We can rewrite the above UCSV model in the non-centered parameterization:

$$\begin{aligned}
y_t &= \tau_t + e^{\frac{1}{2}(h_0+\omega_h\tilde{h}_t)}\tilde{\varepsilon}_t^y,\\
\tau_t &= \tau_{t-1} + e^{\frac{1}{2}(g_0+\omega_g\tilde{g}_t)}\tilde{\varepsilon}_t^\tau,\\
\tilde{h}_t &= \tilde{h}_{t-1} + \tilde{\varepsilon}_t^h,\\
\tilde{g}_t &= \tilde{g}_{t-1} + \tilde{\varepsilon}_t^g,
\end{aligned}$$

where $\tilde{h}_0 = \tilde{g}_0 = 0$ and $\tilde{\varepsilon}_t^y$, $\tilde{\varepsilon}_t^\tau$, $\tilde{\varepsilon}_t^h$, and $\tilde{\varepsilon}_t^g$ are all *i.i.d.* $N(0,1)$. We assume Normal priors for all model parameters: $\omega_h \sim N(0, 0.2^2)$, $\omega_g \sim N(0, 0.2^2)$, $h_0 \sim N(0, 10)$, $g_0 \sim N(0, 10)$, and $\tau_0 \sim N(0, 10)$.

Appendix C: Algorithms

Appendix C.1: Algorithms for VARs with Minnesota-type prior

For all the country-specific VARs, country-specific factor-augmented VARs, global VARs, and multi-country VARs with Minnesota prior, the MCMC samplers follow almost exactly the steps in Carriero et al. [2021], but an additional step is needed to update prior tightness parameters. We highlight three issues related to the sampler, and refer the interested readers to Appendix A.3 in their paper for other details.

Step 1: Update $\beta|\cdot$. We update the coefficients equation by equation, as in the corrected triangular algorithm in Carriero et al. [2021]. Details can be found in Appendix C.5.

Step 2: Update $\lambda_i|\cdot$, $i = 1, 2, 4$. Let S_{λ_i}, $i = 1, 2, 4$, be the collection of all indexes such that parameters associated with the overall shrinkage parameters belong to this set. It can easily shown that, with a Gamma prior, $\lambda_i \sim \mathcal{G}(1, c_i)$, conditional posteriors follow a Generalized Inverse Gaussian distribution:

$$\lambda_i|\cdot \sim \mathcal{GIG}\Big(1 - \frac{\dim(S_{\lambda_i})}{2}, 2c_i, \sum_{(i,j)\in S_{\lambda_i}} \frac{\beta_{i,j}^2}{2C_{i,j}}\Big).$$

The density of $x \sim \mathcal{GIG}(p, a, b)$ is given by $f(x) \propto x^{p-1}\exp\left(-(ax+b/x)/2\right)$. dim denotes the dimension of the set, and $C_{i,j}$ are the prior local variance parameters (the elements in (1.18) without an overall shrinkage parameter).

Step 3: Update the volatility. For the volatility estimation, let $\tilde{u}_t = Au_t$ denote the rescaled residuals. The elements of $\tilde{u}_t$ obey the following process:

$$\ln \tilde{u}_{ij,t}^2 = \ln h_{ij,t} + \ln \epsilon_{ij,t}^2, \quad i = 1, \ldots, N, j = 1, \ldots, G.$$

So, together with state equation (1.16), we have a non-linear and non-Gaussian state space system. To get the volatility estimates, we use the KSC algorithm, first introduced in Kim et al. [1998] and detailed for VAR models in Del Negro and Primiceri [2015]. We use a 10-state mixture of Normals to approximate the distribution of non-Gaussian errors $\ln \epsilon^2_{ij,t}$. The details of approximation are provided in Table 1 of Omori et al. [2007].

Step 4: Update the free elements in A. This can be done with the equation-by-equation approach of Cogley and Sargent [2005] or with the joint approach of Chan [2017]. For the latter, letting a denote the free elements in A, it can be shown that a can be interpreted as the coefficients from the regression:

$$u_t = K_t a + e_t, e_t \sim N(0, D_t),$$

where $D_t = diag(h_{1,t}, \ldots, h_{NG,t})$, and K_t is given as

$$K_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ -u_{1t} & 0 & 0 & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & -u_{1t} & -u_{2t} & 0 & 0 & \cdots & \cdots & \vdots \\ \vdots & & & \ddots & \ddots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & -u_{1t} & \cdots & -u_{(NG-1)t} \end{bmatrix}.$$

This permits drawing a jointly. Given the prior $a \sim N(0, \underline{\Omega}_a)$, the posterior is also Gaussian $a|\beta, h, \Phi, Y \sim N(\overline{\mu}_a, \overline{\Omega}_a)$, where

$$\overline{\Omega}_a = (\underline{\Omega}_a^{-1} + K'H^{-1}K)^{-1}$$
$$\overline{\mu}_a = \overline{\Omega}_a K'H^{-1}u.$$

This algorithm can be more efficient than the equation-by-equation approach, because a is updated jointly. However, the band matrix K_t does not have a fixed bandwidth (the number of non-zeros elements increases with model size). Thus, letting n denote the number of variables in the model, the complexity of this algorithm is still $O(n^3)$, and the estimation quickly becomes computationally demanding as the model size increases. Accordingly, for country-specific models, which are small ($n = N = 3$), we use this algorithm to update a. But for multi-country models, which are large ($n = NG = 21$), we use Cogley and Sargent [2005]'s algorithm to draw a equation by equation.

Step 5: Update $\Phi|\cdot$. Since we elicit a conditionally conjugate prior, the conditional posterior takes the same form, which can be shown to be:

$$\Phi|\cdot \sim IW\Big(Q_0 + T, W_0 + \sum_{t=1}^{T} \big(\log(h_t) - \log(h_{t-1})\big)\big(\log(h_t) - \log(h_{t-1})\big)'\Big).$$

Appendix C.2: Algorithm for multi-country VAR with factor shrinkage

Most of the steps of the algorithm for the CC specification follow from the previous section, except that we have to adapt step 1's treatment of the VAR's coefficients. With the transformation, we see that given

$\theta \sim N(0, \underline{\Omega}_\theta)$, the conditional posterior $\theta | Y, a, h, \Phi$ is multivariate Normal, $N(\overline{\mu}_\theta, \overline{\Omega}_\theta)$, with moments:

$$\overline{\Omega}_\theta = (\underline{\Omega}_\theta^{-1} + \tilde{Z}'\tilde{\Sigma}^{-1}\tilde{Z})^{-1}$$
$$\overline{\mu}_\theta = \overline{\Omega}_\theta \tilde{Z}'\tilde{\Sigma}^{-1} Y,$$

where $Y, \tilde{Z}$ are stacked versions of $Y_t, \tilde{Z}_t$ and $\tilde{\Sigma} = diag(\Sigma_1, \ldots, \Sigma_T)$.

Appendix C.3: Algorithms for multi-country VARs with hierarchical shrinkage

As in Algorithm 1 in the main text, the MCMC estimation involves 5 steps. The only new step compared to above is to update the prior variance parameters and associated hyperparameters. We provide details of the conditional posterior distributions for these parameters.

First, consider the Horseshoe prior:

$$\beta_j | \omega_j^2 \sim N(0, \omega_j^2), \;\; \omega_j^2 | \gamma_j^2 \sim \mathcal{G}(\frac{1}{2}, \gamma_j^2), \;\; \gamma_j^2 \sim \mathcal{G}(\frac{1}{2}, \lambda),$$

and $\lambda \sim C^+(0, 1)$. It follows from straightforward calculation that

$$\omega_j^2 | \cdot \sim \mathcal{GIG}(0, 2\gamma_j^2, \beta_j^2),$$

where $\mathcal{GIG}(p, a, b)$ denotes the Generalized inverse Gaussian distribution with pdf given by $f(x) \propto x^{p-1} \exp\big(-(ax + b/x)/2\big)$. For the conditional posterior of $\gamma_j^2|\cdot$, since the Gamma distribution is conjugate for the Gamma likelihood function, we have that

$$\gamma_j^2 | \cdot \sim \mathcal{G}(1, \lambda + \omega_j^2).$$

To update $\lambda|\cdot$, it follows the same as above since the prior admits the hierarchical representation: $\lambda \sim \mathcal{G}(\frac{1}{2}, \xi^2)$, $\xi \sim \mathcal{G}(\frac{1}{2}, 1)$.

Second, consider the Normal-Gamma prior:

$$\beta_j | \omega_j^2 \sim N(0, \omega_j^2), \;\; \omega_j^2 \sim \mathcal{G}(a^\omega, \frac{a^\omega \kappa^2}{2}),$$

and $a^\omega \sim \mathcal{E}(b)$ and $\kappa^2 \sim \mathcal{G}(d_1, d_2)$. It follows similarly as in the Horseshoe prior that

$$\omega_j^2 | \cdot \sim \mathcal{GIG}(a^\omega - 0.5, a^\omega \kappa^2, \beta_j^2).$$

The conditional posterior for $a^\omega|\cdot$ is not available in closed form. We use adaptive Random Walk Metropolis-

Hastings algorithms as in Roberts and Rosenthal [2009] with acceptance probability given by

$$\min\left\{1, \frac{p(a^{\omega,\text{new}})a^{\omega,\text{new}}}{p(a^{\omega})a^{\omega}} \prod_j \frac{p(\beta_j|a^{\omega,\text{new}},\kappa^2)}{p(\beta_j|a^{\omega},\kappa^2)}\right\},$$

where the marginal prior is given by

$$p(\beta_j|a^{\omega},\kappa^2) = \frac{(\sqrt{a^{\omega}\kappa^2})^{a^{\omega}+\frac{1}{2}}}{\sqrt{\pi}2^{a^{\omega}-\frac{1}{2}}\Gamma(a^{\omega})}\left|\beta_j\right|^{a^{\omega}-\frac{1}{2}} K_{a^{\omega}-\frac{1}{2}}\left(\sqrt{a^{\omega}\kappa^2}\left|\beta_j\right|\right),$$

and $K(\cdot)$ denotes a modified Bessel function of the second kind. At each iteration i, a new value $a^{\omega,\text{new}}$ is proposed according to

$$\log a^{\omega,\text{new}} = \log a^{\omega} + \epsilon_j, \quad \epsilon_j \sim N(0, \sigma^{2(i)}_{\psi_j}). \tag{1.27}$$

The variance of the increments is fixed at 1 for the first 50 iterations, and then updated by

$$\log \sigma^{2(i+1)}_{a^{\omega}} = \log \sigma^{2(i)}_{a^{\omega}} + \frac{1}{i^q}(\hat{\alpha} - \alpha^*), \tag{1.28}$$

where $\hat{\alpha}$ is the estimated acceptance probability of current draws and α^* is the desired acceptance probability. The parameter q controls the degree of vanishing adaption, which is necessary to make the adaptive algorithm valid.[14] This algorithm leads to an average acceptance rate that converges to α^*. Following Griffin and Brown [2017], we set $q = 0.55, \alpha^* = 0.3$. Then updating $\kappa^2|\cdot$ is quite straightforward since it again follows a Gamma distribution:

$$\kappa^2|\cdot \sim \mathcal{G}\Big(Ma^{\omega} + d_1, d_2 + a^{\omega}\sum_j \omega_j^2\Big),$$

where M denotes the number of parameters in each block.

Finally, consider the Normal-Gamma-Gamma prior:

$$\beta_j|\tau_j^2,\lambda_j^2 \sim N\left(0, \phi\frac{\tau_j^2}{\lambda_j^2}\right), \;\; \tau_j^2 \sim \mathcal{G}(a,1), \;\; \lambda_j^2 \sim \mathcal{G}(c,1),$$

where $\phi = 2c/(a\kappa^2)$, $2a \sim \mathcal{B}(\alpha_a,\beta_a)$, $2c \sim \mathcal{B}(\alpha_c,\beta_c)$, and $\kappa^2|a,c \sim F(2a,2c)$. We proceed as in Cadonna et al. [2020]. As we use marginalized distributions in each step to improve sampling efficiency, the steps described below are not interchangeable.

Step a: Update $a|\cdot$. Use the prior $p(\beta_j|\lambda_j^2,a,c)$, marginalized *w.r.t.* τ_j^2, to draw $a|\cdot$ via an adaptive Random Walk Metropolis-Hastings algorithm on $z = \log(a/(0.5-a))$. The variance of the increments is

[14] This means that the variances of increments are fixed as $i \to \infty$. Two conditions are provided in equations (1.1) and (1.2) of Roberts and Rosenthal [2009]. The condition in equation (1.2) in their paper is generally satisfied provided that ψ is bounded above.

updated as in the Normal-Gamma case. At each iteration m, letting a^* be the candidate draw and $a^{(m-1)}$ be the previous draw, the acceptance probability is given by

$$\min\left\{1, \frac{q_a(a^*)}{q_a(a^{(m-1)})}\right\}, \quad q_a(a) = p(a|\cdot)a(0.5-a).$$

Letting m be the number of parameters in each block, $\log q_a(a)$ is given by

$$\begin{aligned}\log q_a(a) = {} & a\Big(-m\log 2 + \frac{m}{2}\log\kappa^2 - \frac{m}{2}\log c + \frac{1}{2}\sum_j \log\lambda_j^2 + \frac{1}{2}\sum_j \log\beta_j^2\Big)\\ & + \frac{5}{4}m\log a + m\frac{a}{2}\log a - m\log\Gamma(a+1)\\ & + \sum_j \log K_{a-\frac{1}{2}}\Big(\beta_j\sqrt{\lambda_j^2\kappa^2 a/c}\Big)\\ & - \log\mathcal{B}(a,c) + a\Big(\log a + \log\big(\frac{\kappa^2}{2c}\big)\Big) - \log a - (a+c)\log\Big(1+\frac{a\kappa^2}{2c}\Big)\\ & + (\alpha_a - 1)\log(2a) - (\beta_a - 1)\log(1-2a)\\ & + \log a + \log(0.5-a).\end{aligned}$$

Step b: Update $\tau_j^2|\cdot$. This step is simple, as the conditional posterior is again $\mathcal{GIG}$:

$$\tau_j^2|\cdot \sim \mathcal{GIG}\Big(a - \frac{1}{2}, 2, \frac{\lambda_j^2\beta_j^2}{\phi}\Big).$$

Step c: Update $c|\cdot$. Use the prior $p(\beta_j|\tau_j^2, a, c)$, marginalized *w.r.t.* λ_j^2, to draw $c|\cdot$ via an adaptive Random Walk Metropolis-Hastings algorithm on $z = \log(c/(0.5-c))$. The variance of the increments is updated as in the Normal-Gamma case. At each iteration m, letting c^* be the candidate draw and $c^{(m-1)}$ be the previous draw, the acceptance probability is given by

$$\min\left\{1, \frac{q_c(c^*)}{q_c(c^{(m-1)})}\right\}, \quad q_c(c) = p(c|\cdot)c(0.5-c).$$

Letting m be the number of parameters in each block, $\log q_c(c)$ is given by

$$\begin{aligned}\log q_c(c) &= m \log \Gamma(c+0.5) - m \log \Gamma(c+1) + \frac{m}{2} \log c \\ &- (c+0.5)\Big(\sum_j \log\left(4c\tau_j^2 + \beta_j^2 \kappa^2 a\right) - \sum_j \log(4c\tau_j^2)\Big) \\ &- \log \mathcal{B}(a,c) - (a-1)\log c - (a+c)\log\left(1 + \frac{a\kappa^2}{2c}\right) \\ &+ (\alpha_c - 1)\log(2c) + (\beta_c - 1)\log(1-2c) \\ &+ \log c + \log(0.5 - c).\end{aligned}$$

Step d: Update $\lambda_j^2|\cdot$. This step is simple; the conditional posterior is $\mathcal{G}$:

$$\lambda_j^2|\cdot \sim \mathcal{G}\Big(\frac{1}{2} + c, \frac{\beta_j^2}{2\phi\tau_j^2} + 1\Big).$$

Step e: Update $\kappa^2|\cdot$. Notice that the prior of κ^2 admits the following hierarchical representation: $\kappa^2|a \sim \mathcal{G}(a, d_2)$, $d_2|a, c \sim \mathcal{G}(c, \frac{2c}{a})$. Then updating $\kappa^2|\cdot$ involves first sampling from

$$d_2|\cdot \sim \mathcal{G}(a + c, \kappa^2 + \frac{2c}{a}),$$

then sampling from (m is the number of parameters in each block)

$$\kappa^2|\cdot \sim \mathcal{G}(\frac{m}{2} + a, \frac{a}{4c}\sum_j \frac{\lambda_j^2}{\tau_j^2}\beta_j^2 + d_2).$$

Appendix C.4: Corrected triangular algorithm

Consider an n-variable reduced-form $VAR(p)$ model as in Carriero et al. [2021]:

$$y_t = \Pi' x_t + A^{-1}\Lambda_t^{0.5}\epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0, I_n),$$

where $t = 1, \ldots, T$, x_t is a $(np+1) \times 1$ dimensional vector containing the lags of y_t and an intercept, $\Pi = (\Pi_0, \Pi_1, \ldots, \Pi_p)'$ is a $(np+1) \times n$ matrix of coefficients, A^{-1} is a unit lower triangular matrix, and $\Lambda_t^{0.5}$ is diagonal with the log of the generic j-th element following a random walk process.

Defining $\tilde{y}_t = Ay_t$ with generic j-th element $\tilde{y}_{j,t} = y_{j,t} + a_{j,1}y_{1,t} + \cdots + a_{j,j-1}y_{j-1,t}$, consider the triangular representation of the system:

$$\tilde{y}_t = A\Pi' x_t + \Lambda_t^{0.5}\epsilon_t = A(x_t'\Pi)' + \Lambda_t^{0.5}\epsilon_t,$$

which can be expressed as the following system of equations:

$$\begin{aligned}
\tilde{y}_{1,t} &= x_t'\pi^{(1)} + \lambda_{1,t}^{0.5}\epsilon_{1,t} \\
\tilde{y}_{2,t} &= a_{2,1}x_t'\pi^{(1)} + x_t'\pi^{(2)} + \lambda_{2,t}^{0.5}\epsilon_{2,t} \\
\tilde{y}_{3,t} &= a_{3,1}x_t'\pi^{(1)} + a_{3,2}x_t'\pi^{(2)} + x_t'\pi^{(3)} + \lambda_{3,t}^{0.5}\epsilon_{3,t} \\
&\vdots \\
\tilde{y}_{n,t} &= a_{n,1}x_t'\pi^{(1)} + \cdots + a_{n,n-1}x_t'\pi^{(n-1)} + x_t'\pi^{(n)} + \lambda_{n,t}^{0.5}\epsilon_{n,t},
\end{aligned}$$

where $\pi^{(j)}$ denotes the coefficients of the j-th equation. Clearly, $\pi^{(j)}$ appears not only in equation j but also in equations $j+1$ through n. Letting $z_{j+l,t} = \tilde{y}_{j+l,t} - \sum_{i\neq j, i=1}^{j+l} a_{j+l,i}x_t'\pi^{(i)}$, for $l = 0, ..., n-j$, and $a_{i,i} = 1$, consider the following system of equations:

$$\begin{aligned}
z_{j,t} &= x_t'\pi^{(j)} + \lambda_{j,t}^{0.5}\epsilon_{j,t} \\
z_{j+1,t} &= a_{j+1,j}x_t'\pi^{(j)} + \lambda_{j+1,t}^{0.5}\epsilon_{j+1,t} \\
&\vdots \\
z_{n,t} &= a_{n,j}x_t'\pi^{(j)} + \lambda_{n,t}^{0.5}\epsilon_{n,t}.
\end{aligned}$$

Then, using the above triangular representation, the full conditional posterior of $\pi^{(j)}|\cdot$ follows immediately from standard Bayesian linear regression results (assuming that prior means are zero):

$$\pi^{(j)}|\cdot \sim N\left(\overline{\mu}_{\pi^{(j)}}, \overline{\Omega}_{\pi^{(j)}}\right),$$

where

$$\begin{aligned}
\overline{\Omega}_{\pi^{(j)}}^{-1} &= \underline{\Omega}_{\pi^{(j)}}^{-1} + \sum_{i=j}^{n} a_{i,j}^2 \sum_{t=1}^{T} \frac{1}{\lambda_{i,t}} x_t x_t' \\
\overline{\mu}_{\pi^{(j)}} &= \overline{\Omega}_{\pi^{(j)}} \times \Big(\sum_{i=j}^{n} a_{i,j} \sum_{t=1}^{T} \frac{1}{\lambda_{i,t}} x_t z_{i,t} \Big),
\end{aligned}$$

with $a_{i,i} = 1$.

Appendix C.5: Algorithms for SSSS prior

The algorithms described in Appendix A.3 of Korobilis [2016] can be easily extended to our case with SV. Only step 1 has to be modified. In particular, let $Y = (y_1 \ \cdots \ y_T)'$, $x_t = (1, y_{t-1}')'$, and $X = (x_1 \ \cdots \ x_T)'$, and write the model as

$$Y = XB + U,$$

where $U = (u_1 \quad \cdots \quad u_T)'$. The sampler involves the following steps:

Step a: Update $\text{vec}(B)|\cdot$. It can be shown that

$$\text{vec}(B)|\cdot \sim N(\Gamma \times \mu_B, D_B),$$

where

$$D_B = \Big(V + \sum_{t=1}^{T} (\Sigma_t^{-1} \otimes x_t' x_t)\Big)^{-1}, \ \ \mu_B = D_B\Big(\text{vec}(\sum_{t=1}^{T} x_t y_t' \Sigma_t^{-1})\Big).$$

The diagonal matrix V contains prior variances; details of constructing the indicator matrix Γ can be found in Korobilis [2016].

Steps b,c,d,e: These follow exactly as in steps 2,3,4,5 in Korobilis [2016].

Steps f,g,h: Update free elements in A, stochastic volatility, and related parameters. These steps follow the corresponding steps used for the multi-country VAR with the Minnesota-type prior.

Appendix C.6: Algorithms for country-specific VAR with hierarchical shrinkage

We follow exactly the algorithms described in Chan [2021]. Estimation for the intercept, autoregressive coefficients, free elements in A, and stochastic volatility are very similar to the algorithms used in this paper. It is worth mentioning that, as in Chan [2021], the model has been first transformed to structural form, and then estimation is performed equation by equation. For hyperparameters related to the Normal-Gamma prior, since a slightly different parameterization is used there, the updating of hyperparameters is slightly different. The conditional posterior for $\psi_{i,j}|\cdot$ is also $\mathcal{GIG}$, but with a slightly different parameterization. An independent Metropolis-Hastings algorithm is used to update $\nu_\psi|\cdot$. We refer the reader to section 4 and Appendix B in that paper for more details.

Appendix C.7: Algorithms for univariate models

We use the algorithms as described in Clark and Ravazzolo [2015] to estimate AR(p)-SV models. The steps to draw intercept and autoregressive parameters follow from standard linear regression results. To estimate stochastic volatility and related parameters, we follow the procedures described in section 7.1 in Chan [2017]. For the UCSV model, we estimate it in non-centered parameterization and then transform back to the centered parameterization to perform predictive simulation. Estimation details can be found in Appendix B in Chan [2018] and in section 7.2 in Chan [2017].

Appendix D: Additional empirical results

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	-1.229	-1.999	-0.623	-1.236	-1.836	-2.761	Mean	-0.204	0.087	-1.018	-0.759	0.610	0.933
Median	-1.053	-2.029	-0.610	-1.183	-2.209	-3.580	Median	0.325	0.264	0.287	0.077	0.349	0.639
Min	-6.251	-7.402	-3.926	-5.022	-6.251	-7.402	Min	-8.772	-7.086	-8.772	-7.086	-6.593	-5.164
Max	3.282	3.078	2.128	2.284	3.282	3.078	Max	10.556	10.025	3.519	4.223	10.556	10.025
% > 0	32.143	28.571	35.714	28.571	28.571	28.571	% > 0	57.143	55.952	54.762	52.381	59.524	59.524
$\%p <= 0.05$	0	2.381	0	2.381	0	2.381	$\%p <= 0.05$	0	1.190	0	2.381	0	0
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	7.951	6.192	3.960	2.191	11.942	10.193							
Median	6.169	5.192	3.180	1.804	10.788	8.174							
Min	-4.416	-5.618	-4.416	-5.618	0.337	0.383							
Max	28.668	23.514	21.308	14.492	28.668	23.514							
% > 0	90.476	86.905	80.952	73.810	100	100							
$\%p <= 0.05$	8.333	10.714	2.381	2.381	14.286	19.048							

Notes: "HS-CSH" is the multi-country VAR model in which all the parameters related to CSH restrictions follow the same Horseshoe prior specification. The table provides summary statistics for the performance of this alternative model compared to the multi-country HS specification. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone [2020].

Table 1.10: Comparison of HS-CSH and baseline HS: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	-1.592	-3.194	-1.386	-2.642	-1.798	-3.746	Mean	-2.180	-1.956	-2.282	-2.190	-2.079	-1.721
Median	-1.626	-3.109	-1.160	-2.488	-2.434	-4.376	Median	-2.037	-1.909	-1.173	-1.290	-2.533	-3.450
Min	-7.806	-12.719	-7.806	-10.600	-7.287	-12.719	Min	-12.187	-10.419	-12.187	-10.419	-9.568	-8.276
Max	4.116	3.688	2.804	3.197	4.116	3.688	Max	9.258	10.280	3.404	4.276	9.258	10.280
% > 0	33.333	28.571	33.333	28.571	33.333	28.571	% > 0	28.571	25	28.571	21.429	28.571	28.571
$\%p <= 0.05$	1.190	4.762	2.381	4.762	0	4.762	$\%p <= 0.05$	0	0	0	0	0	0
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	10.697	10.033	4.370	3.699	17.023	16.367							
Median	9.769	10.128	4.496	3.532	17.011	16.445							
Min	-7.071	-10.052	-7.071	-10.052	6.479	7.464							
Max	25.761	26.416	17.567	15.201	25.761	26.416							
% > 0	90.476	86.905	80.952	73.810	100	100							
$\%p <= 0.05$	21.429	38.095	9.524	23.810	33.333	52.381							

Notes: "HS-A" is the multi-country VAR model in which all the parameters follow the same Horseshoe prior specification. The table provides summary statistics for the performance of this alternative model compared to the multi-country HS specification. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone [2020].

Table 1.11: Comparison of HS-A and baseline HS: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	-1.154	-2.286	-1.212	-2.101	-1.095	-2.470	Mean	-1.756	-2.309	-2.332	-2.680	-1.181	-1.939
Median	-1.454	-2.461	-1.105	-1.768	-1.608	-2.767	Median	-1.109	-2.087	-1.109	-2.370	-1.036	-1.991
Min	-7.244	-10.132	-7.244	-9.228	-6.265	-10.132	Min	-13.790	-15.398	-13.790	-15.398	-9.156	-10.805
Max	4.161	3.987	3.126	3.642	4.161	3.987	Max	11.063	9.459	3.316	4.634	11.063	9.459
% > 0	33.333	28.571	38.095	28.571	28.571	28.571	% > 0	26.190	26.190	28.571	30.952	23.810	21.429
$\%p <= 0.05$	0	1.190	0	2.381	0	0	$\%p <= 0.05$	0	3.571	0	7.143	0	0
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	11.280	10.700	4.059	3.313	18.501	18.087							
Median	10.568	10.984	2.596	2.770	16.977	17.372							
Min	-7.634	-11.672	-7.634	-11.672	6.737	7.870							
Max	34.537	28.239	25.968	19.593	34.537	28.239							
% > 0	83.333	82.143	66.667	64.286	100	100							
$\%p <= 0.05$	13.095	20.238	2.381	9.524	23.810	30.952							

Notes: "HS-E" is the multi-country VAR model in which all the parameters in each equation follow the same Horseshoe prior specification. The table provides summary statistics for the performance of this alternative model compared to the multi-country HS specification. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone [2020].

Table 1.12: Comparison of HS-E and baseline HS: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	-1.852	-3.445	-0.893	-2.525	-2.811	-4.364	Mean	-22.974	-19.720	-11.820	-12.712	-34.127	-26.727
Median	-1.962	-3.434	-1.123	-2.603	-2.722	-4.309	Median	-16.394	-15.737	-7.538	-8.186	-24.002	-19.492
Min	-9.097	-11.549	-6.032	-8.837	-9.097	-11.549	Min	-98.999	-69.770	-56.606	-52.922	-98.999	-69.770
Max	5.197	4.580	5.197	4.580	2.362	0.835	Max	4.018	0.548	4.018	0.548	-7.184	-7.248
%> 0	28.571	19.048	38.095	28.571	19.048	9.524	%> 0	4.762	1.190	9.524	2.381	0	0
$\%p <= 0.05$	3.571	5.952	0	2.381	7.143	9.524	$\%p <= 0.05$	15.476	15.476	16.667	16.667	14.286	14.286
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	-9.429	-13.515	-12.149	-20.332	-6.709	-6.698							
Median	-7.592	-6.062	-8.375	-14.525	-7.590	-2.534							
Min	-57.608	-73.394	-57.608	-73.394	-27.100	-47.395							
Max	10.402	14.686	9.970	11.987	10.402	14.686							
%> 0	26.190	28.571	21.429	14.286	30.952	42.857							
$\%p <= 0.05$	11.905	28.571	16.667	42.857	7.143	14.286							

Notes: The table provides summary statistics for the performance of the alternative model with SV compared to the multi-country HS specification with SV. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone [2020].

Table 1.13: Comparison of Horseshoe priors with and without SV: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	1.095	0.342	1.229	0.707	0.960	-0.023	Mean	3.268	1.872	3.989	2.046	2.546	1.698
Median	0.849	0.160	1.102	0.528	0.759	0.002	Median	3.006	1.788	3.533	2.128	2.742	1.208
Min	-1.632	-1.930	-1.632	-1.930	-0.688	-1.772	Min	-1.479	-3.063	-1.318	-2.326	-1.479	-3.063
Max	7.256	4.464	7.256	4.464	3.948	3.023	Max	10.905	5.850	10.905	5.495	7.575	5.850
$\%> 0$	83.333	54.762	80.952	59.524	85.714	50	$\%> 0$	92.857	85.714	95.238	88.095	90.476	83.333
$\%p <= 0.05$	3.571	4.762	7.143	9.524	0	0	$\%p <= 0.05$	0	0	0	0	0	0
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	1.928	1.330	0.539	-0.077	3.317	2.738							
Median	1.435	0.461	-0.338	0.043	3.237	1.713							
Min	-6.350	-6.095	-6.350	-6.095	-2.769	-4.450							
Max	12.168	12.503	9.226	9.878	12.168	12.503							
$\%> 0$	59.524	57.143	47.619	50	71.429	64.286							
$\%p <= 0.05$	10.714	4.762	7.143	0	14.286	9.524							

Notes: The table provides summary statistics for the performance of the HS model estimated with a rolling approach relative to the paper's baseline recursive approach. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone [2020].

Table 1.14: Comparison of Horseshoe priors with expanding versus rolling windows: descriptive statistics for all horizons

	All horizons		$h \leqslant 6$		$h > 6$			All horizons		$h \leqslant 6$		$h > 6$	
Output growth	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS	**Inflation**	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS
Mean	0.535	0.732	0.290	0.396	0.779	1.067	Mean	1.279	3.874	0.252	0.927	2.306	6.822
Median	-0.513	-0.461	-0.342	-0.406	-0.750	-0.744	Median	2.022	3.073	1.740	1.213	2.572	7.117
Min	-3.642	-4.655	-3.616	-3.813	-3.642	-4.655	Min	-13.484	-7.475	-13.484	-7.475	-11.329	-3.670
Max	7.571	10.119	5.646	6.972	7.571	10.119	Max	13.247	17.693	6.875	7.681	13.247	17.693
$\%> 0$	38.095	40.476	40.476	42.857	35.714	38.095	$\%> 0$	70.238	77.381	59.524	64.286	80.952	90.476
$\%p <= 0.05$	13.095	14.286	11.905	7.143	14.286	21.429	$\%p <= 0.05$	0	8.333	0	0	0	16.667
Interest rate	RMSFE	CRPS	RMSFE	CRPS	RMSFE	CRPS							
Mean	8.982	7.799	8.442	7.370	9.522	8.227							
Median	10.504	10.207	8.651	8.779	14.056	16.663							
Min	-19.264	-28.167	-15.896	-22.757	-19.264	-28.167							
Max	28.698	29.027	24.232	25.927	28.698	29.027							
$\%> 0$	78.571	63.095	85.714	69.048	71.429	57.143							
$\%p <= 0.05$	11.905	25	9.524	23.810	14.286	26.190							

Notes: This table presents descriptive statistics on comparisons of forecasting performance for the multi-country VAR-SV model with the Horseshoe prior (the paper's HS specification) relative to univariate models with SV. For output growth and the interest rate, we use an AR(p)-SV model, with $p = 2$ for output growth and $p = 4$ for the interest rate. For inflation, we use an unobserved component model with SV, as in Chan [2018]. Descriptive statistics include average, median, minimum, maximum, percentage of cases in which gains are above 0, and the percentage gains in which the forecasts from the competing models are statistically different according to the Diebold-Mariano (1995) test with fixed-smoothing asymptotics as in Coroneo and Iacone [2020].

Table 1.15: Comparison with univariate models with HS baseline featuring SV: descriptive statistics for all horizons

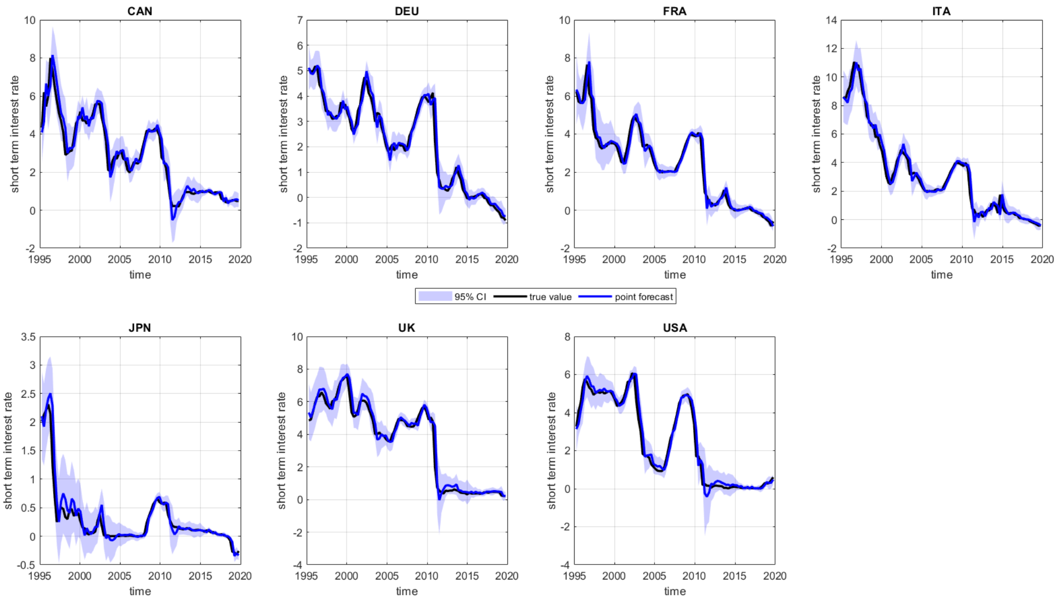

Figure 1.7: The figures present 1-step-ahead short-term interest rate forecasts for all G7 countries. The blue line and shaded areas are point forecasts and associated 95 percent forecast intervals. The black line shows the true values.

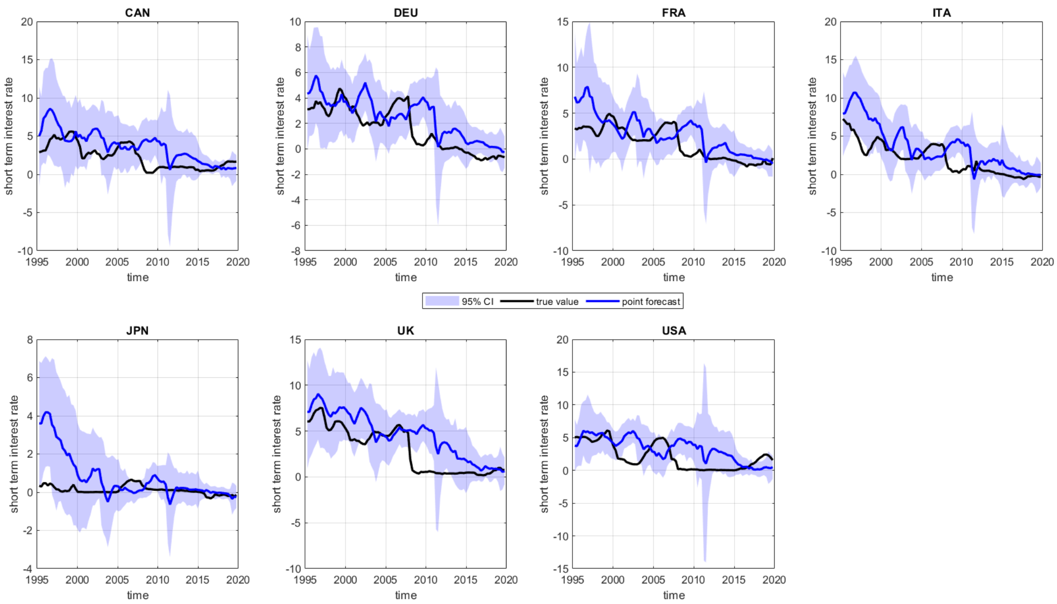

Figure 1.8: The figures present 12-steps-ahead short-term interest rate forecasts for all G7 countries. The blue line and shaded areas are point forecasts and associated 95 percent forecast intervals. The black line shows the true values.

	CAN	DEU	FRA	ITA	JPN	UK	USA
HS	-0.380	6.715	2.508	5.870	7.556	8.174	9.225
	(0.648)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)
CVAR	-2.183	0.421	-2.928	-2.307	-2.318	-1.668	-3.797
	(0.986)	(0.337)	(0.998)	(0.990)	(0.990)	(0.952)	(0.999)

Notes: This table presents test statistics and associated p-values for directional predictive performance of 1-step ahead changes of output growth from multi-country VAR-SV model with Horseshoe prior and single-country VAR-SV benchmark. The test statistics are computed according to equation (6) in Pesaran and Timmermann [1992].

Table 1.16: Directional forecast: 1-step ahead changes of output growth

	RMSFE				CRPS			
Output growth	$h = 1$	$h = 4$	$h = 8$	$h = 12$	$h = 1$	$h = 4$	$h = 8$	$h = 12$
CAN	2.319	2.712	2.637	2.647	1.245	1.466	1.419	1.402
DEU	3.592	3.595	3.569	3.470	1.832	1.805	1.798	1.764
FRA	1.630	2.068	2.116	2.144	0.892	1.115	1.134	1.149
ITA	2.539	2.985	2.964	2.935	1.335	1.580	1.553	1.515
JPN	4.185	4.167	4.156	4.251	2.208	2.187	2.159	2.261
UK	2.070	2.542	2.509	2.534	1.080	1.311	1.282	1.287
USA	2.335	2.548	2.594	2.542	1.266	1.364	1.393	1.367
Inflation	$h = 1$	$h = 4$	$h = 8$	$h = 12$	$h = 1$	$h = 4$	$h = 8$	$h = 12$
CAN	1.870	1.722	1.816	1.809	0.998	1.006	1.060	1.085
DEU	1.140	1.277	1.377	1.376	0.663	0.743	0.814	0.794
FRA	1.118	1.410	1.439	1.456	0.618	0.783	0.848	0.875
ITA	0.934	1.498	1.690	1.777	0.503	0.830	0.946	1.005
JPN	1.652	1.797	1.846	1.858	0.891	0.978	1.015	1.023
UK	0.982	1.215	1.384	1.358	0.542	0.701	0.778	0.813
USA	2.151	2.227	2.223	2.193	0.988	1.102	1.185	1.160
Interest rate	$h = 1$	$h = 4$	$h = 8$	$h = 12$	$h = 1$	$h = 4$	$h = 8$	$h = 12$
CAN	0.474	1.327	2.130	2.588	0.231	0.704	1.187	1.502
DEU	0.323	1.068	1.795	2.202	0.162	0.589	1.083	1.374
FRA	0.416	1.298	2.083	2.445	0.192	0.675	1.182	1.419
ITA	0.461	1.502	2.518	3.190	0.234	0.777	1.393	1.817
JPN	0.177	0.740	1.342	1.577	0.068	0.278	0.536	0.676
UK	0.418	1.233	1.884	2.289	0.189	0.630	1.012	1.295
USA	0.353	1.188	2.076	2.644	0.173	0.648	1.205	1.588

Table 1.17: Loss function levels for the benchmark CVAR specification

Chapter 2

Estimation and inference in large heterogeneous panels with stochastic time-varying coefficients

[1]

2.1 Introduction

Since the study by Pesaran and Smith [1995], large heterogeneous panel data models have received a lot of attention in both theoretical work and practical applications. Surveys of the literature on large heterogeneous panels are provided by Hsiao and Pesaran [2008] and Chapter 28 of Pesaran [2015]. Double-index panel data models enable researchers to explore both dynamic information over the time-span and heterogeneity over cross-sections, which may be difficult to examine by applying purely cross sectional or time series models. It is now quite common to have panels with both large cross-sectional units (N) and time-series periods (T) and it has been found that neglected heterogeneity may lead to misleading inferences [Ul Haque et al., 1999].

Various methods have been proposed to identify and handle structural change in econometric models. As parameter instability is pervasive [Stock and Watson, 1996], allowing coefficients to vary over time would offer benefits for flexible modeling of the true relationship between economic and financial variables. Evolutions of parameters can be either discrete and abrupt, such as in Markov Switching models [e.g., Hamilton, 1989], or continuous and smooth. Continuous and smooth time variation can be driven by observed variables, as in smooth transition models [Teräsvirta, 1994], or by unobserved shocks, as in random coefficient models [e.g., Nyblom, 1989]. In these models, parameters typically evolve as random walk or autoregressive processes and are mostly estimated by the Kalman Filter in classical context or by

[1]This is a joint work with Massimiliano Marcellino and George Kapetanios.

Bayesian Markov Chain Monte Carlo (MCMC) methods.

Yet another strand of the vast and growing literature on dealing with parameter instability allows for a smooth evolution of parameters without specifying the form of parameter time variation. The evolution can be either deterministic, as in Robinson [1991] and Chen and Hong [2012], or stochastic, as in Giraitis et al. [2014, 2018]. These papers have provided theoretical, Monte Carlo and empirical results to justify their estimation methods and showed that they indeed perform very well in finite samples. Such estimates are nonparametric and can have computational advantages over MCMC and other simulation-based methods.[2] The approach has also been extended to panel data models. Chen and Huang [2018] propose methods to estimate and test smooth structural changes in panel data models with exogenous regressors and homogenous time-varying coefficients. Liu et al. [2018] and Liu et al. [2020] develop methods to estimate time-varying coefficients in large panel data models with cross-sectional dependence, but focus on the case of exogenous regressors.

When the parameters of interests are coefficients attached to endogenous variables, endogeneity bias invalidates least square estimation and instrumental variable (IV) estimation comes to play a role. A usual assumption made when carrying out IV estimation is that the parameters in the entertained model are constant over time. This assumption is clearly restrictive, because relations between economic variables as well as instruments and endogenous variables may vary over time. Recently, some papers have attempted to develop estimation methods in the IV framework that account for the possible presence of parameter instability. Hall et al. [2012] develop inferential theory for linear GMM estimator with endogenous regressors in the structural change context. Chen [2015] extends the framework of deterministic smooth evolution of parameters to the IV case. Giraitis et al. [2020a] propose non-parametric kernel-based estimation and inferential theory for time-varying IV regression, with either deterministic or random coefficients. There is also limited but growing attention in the panel data literature to models allowing for both endogenous regressors and parameter instability. Baltagi et al. [2016] and Baltagi et al. [2019] develop an estimation procedure for large heterogeneous panels with cross-sectional dependence in the structural change context by extending the work of Pesaran [2006] and Harding and Lamarche [2011].

This paper makes the following contributions to the literature. First, we introduce a new class of large heterogeneous panels in which parameters are not only heterogeneous but also vary stochastically over time. The model extends both standard random coefficient panel data model as in Hsiao and Pesaran [2008] and panel random coefficient autoregression model as in Horváth and Trapani [2016]. We propose non-parametric kernel-based mean group and pooled estimators, derive their asymptotic distributions and show the uniform consistency and asymptotic normality of the path coefficients. Second, we extend the work of Giraitis et al. [2020a] to large heterogeneous panels. We show that both kernel-based mean group and pooled estimators can be extended to settings with possibly endogenous regressors. We also derive the properties of time-varying IV mean group and pooled estimators and show their uniform consistency

[2] See, for instance, Kapetanios et al. [2019a], for a comparison between kernel-based methods and simulation-based methods in a vector autoregression context.

and asymptotic normality under similar conditions to the case of time-varying least square estimators. We further propose a pointwise time-varying version of the Hausman exogeneity test in a large heterogeneous panels context, which compares time-varying OLS and IV estimators, possibly also allowing for changes in the endogeneity status of regressors over time. The finite sample performance of proposed estimators and the time-varying Hausman test is evaluated in an extensive Monte Carlo study. For the estimators, we evaluate the biases of point estimates and coverage probabilities of the coefficients paths under the scenarios of both exogenous and endogenous regressors. We also compute both the size and power of the time-varying Hausman test. The results are encouraging, and can be also used to provide guidelines on the choice of the kernel bandwidth parameters.

Finally, we provide an empirical application to explore in practice the use of our proposed estimators. We estimate panel versions of time-varying hybrid Phillips curves with 19 Eurozone countries over the period 2000M1–2019M12. We find trade-off between unemployment and inflation is time-varying, but the coefficients are small and only significant roughly around the period of year 2005 and 2014–2016. Endogeneity issues may arise not only because inflation expectation is not observed, but also the fact that inflation is measured with error. In general, IV delivers much larger estimates than OLS for persistent parameters. Backward-looking is a dominating feature for Eurozone inflation, except for the period around year 2015.

The remainder of this paper is organized as follows. Section 2 describes our framework and the time-varying least square estimators, and derives the related theoretical results. Section 3 extends the work to the case of possibly endogenous regressors, proposes time-varying IV estimators, derives their theoretical properties and introduces the time-varying Hausman test. In Section 4 we evaluate our proposed estimators and pointwise Hausman exogeneity test in an extensive Monte Carlo study, under the scenarios of both exogenous and endogenous regressors. Section 5 presents the empirical application related to multi-country Phillips curves. Section 6 summarizes our main results and concludes the paper. The proofs of all results are presented in the appendices.

NOTATION: The letter C stands for a generic finite positive constant, $||A||_{sp} = \sqrt{\lambda_{\max}(A'A)}$ is the spectrum norm of matrix A, where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of $\cdot$. $||\cdot||_p$ denotes the L^p norm, $||\cdot||$ is the Euclidean norm. $|\cdot|_p$ and $|\cdot|$ denote the associated norm when $\cdot$ is one dimensional. $a_n = O(b_n)$ states that the deterministic sequence $\{a_n\}$ is at most of order b_n. $a_n = o(b_n)$ states that the deterministic sequence $\{a_n\}$ is of smaller order than b_n. $x_n = O_p(y_n)$ states that the vector of random variables x_n is at most of order y_n in probability, and $x_n = o_p(y_n)$ is of smaller order than y_n in probability. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution. $(N, T) \to \infty$ denotes joint convergence of N and T.

2.2 Theoretical considerations

In this section, we present our model and set out the proposed estimators and their properties. We consider the following model:

$$y_{it} = x_{it}'\beta_{it} + u_{it}, \quad i = 1, 2, ..., N, \ \ t = 1, 2, ..., T, \tag{2.1}$$

where y_{it} denotes the explained variable, x_{it} is a $k \times 1$ dimensional vector of explanatory variables, one of them being a constant (e.g., $x_{1,it} = 1$), and u_{it} is the disturbance term. We make the following assumptions on this model.

Assumption 2.2.1. *Elements in x_{it}, u_{it} have the properties:*

(i) There exists $\theta > 4$ such that $E\left|x_{\ell,it}\right|^{\theta} < \infty$ and $E\left|u_{\ell,it}\right|^{\theta} < \infty$, uniformly over ℓ, i, t;

(ii) $\forall(\ell, i, t)$, $(x_{\ell,it} - Ex_{\ell,it})$, and $(u_{\ell,it})$ are strong-mixing processes with mixing coefficients α_k^j satisfying

$$\alpha_k^j \leqslant c_j \phi_j^k, \ \ k \geqslant 1 \tag{2.2}$$

for some $0 < \phi_j < 1$ and $c_j > 0$, where $j = \{x, u\}$.

Assumption 2.2.2. *The coefficients β_{it} follow the random coefficient model:*

$$\beta_{it} = \beta_t + e_{it}, \ \ i = 1, 2, \cdots, N, \ \ t = 1, 2, \cdots, T, \tag{2.3}$$

(i) Let $e_i = (e_{i1}, e_{i2}, \cdots, e_{iT})'$, $u_i = (u_{i1}, u_{i2}, \cdots, u_{iT})'$, $x_i = (x_{i1}', x_{i2}', \cdots, x_{iT}')'$, $(x_i, u_i, e_i)'$ are independently distributed over i;

(ii) $\forall(\ell, i, t)$, $E[e_{\ell,it}|\beta_{\ell,t}, x_{\ell,it}] = 0$;

(iii) $\forall(\ell, i, t)$, $E[u_{\ell,it}|x_{\ell,it}, \beta_{\ell,t}, e_{\ell,it}] = 0$;

(iv) $\forall t$, we have

$$\Omega_{e,t} = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it}\Big) = O(1)$$

Assumption 2.2.3. *$\forall(\ell, i, t)$, elements in $\beta_t = (\beta_{\ell,t})$, $e_{it} = (e_{\ell,it})$ satisfy the following smoothness condition:*

$$\left|\beta_{\ell,t} - \beta_{\ell,s}\right| \leqslant \Big(\frac{|t-s|}{T}\Big)^{\gamma_1} r_{\ell,ts}^{(1)}, \quad \left|e_{\ell,it} - e_{\ell,is}\right| \leqslant \Big(\frac{|t-s|}{T}\Big)^{\gamma_1} r_{\ell,i,ts}^{(2)}, \ \ t, s = 1, 2, \cdots, T \tag{2.4}$$

for some $0 < \gamma_1 < 1$ and the distribution of each variable in $X_u^{(1)} = \{\beta_{\ell,t}, e_{\ell,it}, r_{\ell,ts}^{(1)}, r_{\ell,i,ts}^{(2)}\}$ has a thin tail $\mathcal{E}(\alpha)$:

$$\mathbb{P}(\left|X_u^{(1)}\right| \geqslant \omega) \leqslant \exp(-c_1|\omega|^{\alpha}), \ \ \omega > 0,$$

for $c_1 > 0$, $\alpha > 0$ which does not depend on t, s, T.

Assumption 2.2.4. *The matrix* $\Sigma_{xx,it} = E(x_{it}x_{it}')$ *is such that*

$$(i)\ \max_{t=1,\cdots,T}\left\|\Sigma_{xx,it}\right\|_{sp} < \infty, \quad \forall i; \qquad (ii)\ \left\|\Sigma_{xx,it} - \Sigma_{xx,is}\right\|_{sp} \leqslant C\,\frac{|t-s|}{T}, \quad \forall i, t.$$

Assumption 2.1(i) impose some moment conditions on regressors and error terms. Assumption 2.1(ii) are strong mixing conditions to control temporal dependence, which are weaker than the conditions imposed in Chen and Huang [2018] since we allow (u_{it}) and $(x_{it}u_{it})$ to be serially correlated sequences. In Assumption 2.2(i), as in (2.3), β_t is a $K \times 1$ vector of stochastic time-varying coefficients and e_{it} is a $K \times 1$ vector of random variables. If $\beta_t = \beta$ and e_{it} is *i.i.d.* across t for every i, we have the model as in Horváth and Trapani [2016]. If $\beta_t = \beta$ and $e_{it} = e_i$, $\forall t$, the model simplifies to the standard random coefficient model settings with time-invariant coefficient [see Hsiao and Pesaran, 2008]. Assumption 2.2(ii) and (iii) impose exogeneity conditions for x_{it}. Assumption 2.2(ii) rules out correlated random effects. Assumption 2.2(iii) imposes contemporaneous exogeneity condition on x_{it}. x_{it} are neither correlated with error terms nor with idiosyncratic components in β_{it}. Assumption 2.2(iv) is the condition to ensure that asymptotic variance of the proposed estimators are positive definite. Assumption 2.3 implies that elements in (β_t), (e_{it}) are smoothly varying persistent stochastic processes. If $r^{(1)}_{\ell k,t}$ and $r^{(2)}_{\ell k,t}$ are equal to 1 and $\gamma_1 = 1$, we obtain the case of deterministic smoothly time varying coefficient model. For example, consider an array of random processes (bounded random walk) defined as $\beta_{\ell,t} = \frac{1}{\sqrt{T}}u_{\ell,t}$, where $u_{\ell,t}$ are random walk processes: $u_{\ell,t} - u_{\ell,t-1} \overset{i.i.d.}{\sim} N(0,1)$, satisfying Assumption 2.2(ii). As shown in Lemma 1 of Dendramis et al. [2020], Assumption 2.2(ii) is satisfied if $\omega_{\ell,t} = u_{\ell,t} - u_{\ell,t-1}$ is α-mixing and has a thin tail.[3] Other allowable processes are discussed by Giraitis et al. [2014, 2018]. Assumption 2.4(i) states that the second order moment of (x_{it}), $\Sigma_{xx,it}$, is uniformly bounded for each i. In Assumption 2.4(ii), we allow $\Sigma_{xx,it}$ to vary over time but the variations have to be smooth, which are essential to establish weak law of large numbers (WLLN, see, Lemma A1).

Remark 1. Consider $x_{1,it} = 1$, let $x_{-1,it}$ be the remaining regressors in x_{it} and $\beta_{-1,it}$ be the associate coefficients, then, the model (2.1) can be rewritten as $y_{it} = \beta_{1,t} + x_{-1,it}'\beta_{-1,it} + e_{1,it} + u_{it}, \quad i = 1, 2, ..., N, \ \ t = 1, 2, ..., T$. As shown in Theorem 1, the component $\beta_{1,t}$ can be consistently estimated by our proposed kernel methods. $e_{1,it}$ captures unobserved heterogeneity across both i and t, which bridges our model and the conventional random effects panel data model. In standard random effects panel data model, $e_{1,it}$ are either specified in additive form $e_{1,it} = e_i^1 + e_t^{(1)}$ or interactive form $e_{1,it} = (e_i^1)'e_t^{(1)}$. However, in our framework it is not required to specify the forms to model unobserved heterogeneity. We just need to assume that it is independent over i and change smoothly over time.

Remark 2. The condition in (2.4) are the key assumptions to ensure consistency and applicability of central limit theorem (CLT). Unlike Chen and Huang [2018], we focus on path coefficients β_t, $t = 1, 2, \cdots, T$ under conditions on their increments rather than estimating deterministic functions $\beta(\frac{t}{T})$. Our assumptions

[3] Dendramis et al. [2020] derive this under the case $\gamma_1 = 0.5$, but extensions are quite straightforward.

also differ from the micro panel literature (large N, fixed T). For example, unlike Graham and Powell [2012], we do not require β_t to be stationary and even $i.i.d.$ across t.

The main objective in this section is to construct estimators for β_t and derive uniform consistency rates and asymptotic distributions for the estimators. The individual specific estimates can be obtained from a time-varying parameter least-square estimator (TVP-OLS):

$$\hat{\beta}_{i,t} = \Big(\sum_{j=1}^{T} b_{j,t}(H) x_{ij} x_{ij}'\Big)^{-1} \Big(\sum_{j=1}^{T} b_{j,t}(H) x_{ij} y_{ij}\Big). \tag{2.5}$$

The kernel weights $b_{j,t}(H)$ are defined as

$$b_{j,t}(H) = K\Big(\frac{|j-t|}{H}\Big),$$

where the bandwidth parameter H satisfies $H = o(T)$ as $T \to \infty$. $K(x)$ is a non-negative continuous function with either bounded or unbounded support satisfying

$$|K(x)| \leqslant C(1 + x^{\nu})^{-1}, \quad |(d/dx)K(x)| \leqslant C(1 + x^{\nu})^{-1},$$

for some $C > 0$ and $\nu \geqslant 2$. Examples include $K(x) = \frac{1}{2} I\{|x| \leqslant 1\}$, $K(x) = \frac{3}{4}(1 - x^2) I\{|x| \leqslant 1\}$ and $K(x) \propto \exp(-cx^{\alpha})$ with $c > 0$, $\alpha > 0$.

As in the literature of large heterogenous panels, we propose two types of estimators. The first is a mean group estimator (TVP-OLS-MG), given by

$$\hat{\beta}_{MG,t} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_{i,t}, \tag{2.6}$$

where $\hat{\beta}_{i,t}$ is defined in (2.5). The second is a pooled estimator (TVP-OLS-P):

$$\hat{\beta}_{P,t} = \Big(\sum_{i=1}^{N} \sum_{j=1}^{T} b_{j,t}(H) x_{ij} x_{ij}'\Big)^{-1} \Big(\sum_{i=1}^{N} \sum_{j=1}^{T} b_{j,t}(H) x_{ij} y_{ij}\Big). \tag{2.7}$$

Before analyzing the theoretical properties of these estimators, we assume that the bandwidth parameter H satisfies:

$$c_1 T^{1/(\theta/4-1)+\delta_1} \leqslant H \leqslant c_2 T^{1-\delta_2} \tag{2.8}$$

for some $\theta > 4$ as in Assumption 2.1, and $c_1, c_2 > 0$, $\delta_1, \delta_2 > 0$ are sufficiently small. In the next theorem, we establish uniform consistent rate, asymptotic distributions for (2.6) and (2.7) and how to obtain consistent estimates of asymptotic covariance matrices.

Theorem 2.2.1. *Under Assumptions 2.1–2.4 and assuming that the bandwidth parameter H satisfies* (2.8)*,* *as $(N,T) \to \infty$, we have the following:*

(1) Uniform consistency: $\hat\beta_{MG,t}$, $\hat\beta_{P,t}$ have the property

$$\max_{t=1,2,\cdots,T}\left\|\hat\beta_{\int,t} - \beta_t\right\| = O_p(r^1_{N,T,H,\gamma,\alpha}),$$

where $r^1_{NT,H,\alpha} = (\frac{H}{T})^{\gamma_1}\log^{1/\alpha} T + \sqrt{\frac{\log T}{NH}} + \frac{\log^{1/\alpha} T}{\sqrt{N}}$ *for* $\int = \{MG, P\}$.

(2) Asymptotic normality: Suppose that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$,

(i) Mean group estimator:

$$\sqrt{N}\left(\Omega_{e,t}\right)^{-1/2}\left(\hat\beta_{MG,t} - \beta_t\right) \xrightarrow{d} N(0, I_k),$$

where $\Omega_{e,t}$ is given by

$$\Omega_{e,t} = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\Big).$$

(ii) Pooled estimator: Suppose that $\overline{\Sigma}_{xx,t} = \text{plim}_{(N,T)\to\infty}\frac{1}{NK_t}\sum_{i=1}^N\sum_{j=1}^T b_{jt}x_{ij}x'_{ij} = O_p(1)$*, then*

$$\sqrt{N}\,\overline{\Sigma}_{xx,t}\ R_t^{-1/2}\left(\hat\beta_{P,t} - \beta_t\right) \xrightarrow{d} N(0, I_k),$$

where R_t is given by

$$R_t = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{xx,it}e_{it}\Big).$$

Even though both $\Omega_{e,t}$ and R_t contain e_{it}, which is not observed, in Appendix B1, we show that they can be consistently estimated by $\hat\Omega_{e,t}$ and $\hat R_t$:

$$\hat\Omega_{e,t} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat\beta_{i,t} - \hat\beta_{MG,t}\right)\left(\hat\beta_{i,t} - \hat\beta_{MG,t}\right)' \tag{2.9}$$

$$\hat R_t = \frac{1}{N}\sum_{i=1}^{N}\left[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\Big)(\hat\beta_{i,t} - \hat\beta_{MG,t})(\hat\beta_{i,t} - \hat\beta_{MG,t})'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\Big)\right]. \tag{2.10}$$

Three comments are in order. First, the uniform consistency rate is different from the time series setting ((19) in Giraitis et al. [2020a]). The first term in $r^1_{NT,H,\alpha}$ is the same as before, since it only relates to the properties of the parameter of interests (Assumption 2.3). The second term is different, since N appears in the rate. The third term is new, due to the heterogeneity in model parameters e_{it}. Indeed, panel dimension N is helpful to improve the rate. Second, as in the setting of standard random coefficient model (Hsiao et al. [1998]), both estimators have standard root N rate of convergence. However, the requirement $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ is different from time series setting, which is $(\frac{H}{T})^{\gamma_1} = o(H^{-1/2})$. Third, Inference on Mean group estimator $\hat{\beta}_{MG,t}$ is straightforward, since the asymptotic variance only depends on the variance of e_{it}. Inference on pooled estimator $\hat{\beta}_{P,t}$ is slightly more involved, since asymptotic variance relies on additional assumption and the second order moments of (x_{it}).

The use of estimates $\hat{\beta}_{MG,t}$ and $\hat{\beta}_{P,t}$ makes it necessary to choose the bandwidth parameter H. Whereas uniform consistency holds under minimal restrictions on H (see (2.8)), asymptotic normality results require stronger restrictions: $(\frac{H}{T})^{\gamma_1} = o(H^{-1/2})$. This condition implies that $\sqrt{N}(\frac{H}{T})^{\gamma_1} \to 0$ when both $(N, T) \to \infty$. As in Giraitis et al. [2014, 2020a], a practical suggestion for H is to set $H = T^{\overline{\alpha}}$, for some $0 < \overline{\alpha} < 1$. Then, the condition simplifies to $\sqrt{N}/T^{\gamma_1(1-\overline{\alpha})} \to 0$. If we assume that β_t follows the bounded random walk process and set $H = T^{0.5}$, the condition becomes $\sqrt{N}/T^{0.25} \to 0$ as $(N, T) \to \infty$. A practically meaningful implication is that T has to diverge at a faster rate than N.

2.3 Endogenous regressors

Consider again the model proposed in section 2:

$$
\begin{aligned}
y_{it} &= x_{it}'\beta_{it} + u_{it}, \\
\beta_{it} &= \beta_t + e_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T.
\end{aligned}
$$

In Assumption 2.2(iii), we impose contemporaneous exogeneity condition on x_{it}, which is rather restrictive. This condition is likely to be violated in many empirical applications, due to simultaneity, measurement error, or omitted variables.

In this section, we consider an extension of the proposed TVP-OLS type estimators to the Instrumental variable regression (IVR) context. Let us consider the following model:

$$y_{it} = x_{it}'\beta_{it} + u_{it}, \tag{2.11}$$

$$x_{it} = \Psi_{it}' z_{it} + v_{it}, \tag{2.12}$$

$$\beta_{it} = \beta_t + e_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T, \tag{2.13}$$

where $z_{it} = (z_{1,it}, z_{2,it}, \cdots, z_{p,it})'$ is a $p \times 1$ vector of instruments, $\Psi_{it}' = (\psi_{\ell k,it})$ is a $k \times p$ parameter matrix and $v_{it} = (v_{1,it}, v_{2,it}, \cdots, v_{p,it})'$ is a $k \times 1$ vector of error terms. We shall make the following additional assumptions.

Assumption 2.3.1. *Elements in z_{it}, v_{it} have the properties:*

(i) There exists $\theta > 4$ such that $E\left|z_{\ell,it}\right|^{\theta} < \infty$ and $E\left|v_{\ell,it}\right|^{\theta} < \infty$, uniformly over ℓ, i, t;

(ii) $\forall \ell, i, t$, $(z_{\ell,it} - Ez_{\ell,it})$ and $(v_{\ell,it})$ are strong-mixing processes with mixing coefficients α_k^j satisfying

$$\alpha_k^j \leqslant c_j \phi_j^k, \quad k \geqslant 1 \tag{2.14}$$

for some $0 < \phi_j < 1$ *and* $c_j > 0$*, where* $j = \{z, v\}$*.*

Assumption 2.3.2. *The coefficients* Ψ_{it} *follow the random coefficient model:*

$$\Psi_{it} = \Psi_t + \Upsilon_{it}, \quad i = 1, 2, \cdots, N, \quad t = 1, 2, \cdots, T, \tag{2.15}$$

(i) Let $\Upsilon_i = (\Upsilon_{i1}, \Upsilon_{i2}, \cdots, \Upsilon_{iT})'$*,* $v_i = (v_{i1}, v_{i2}, \cdots, v_{iT})'$*,* $z_i = (z'_{i1}, z'_{i2}, \cdots, z'_{iT})'$*,* $(x_i, z_i, u_i, v_i, v_i, \Upsilon_i)$ *are independently distributed over i;*

(ii) $\forall \ell, k$*,* $E[u_{\ell,it}|z_{\ell,it}, \beta_{\ell,t}, e_{\ell,it}, \Psi_{\ell k,t}] = 0$*,* $E[v_{\ell,it}|z_{\ell,it}, \Psi_{\ell k,t}, \Upsilon_{\ell k,it}] = 0$*,* $\forall (i, t)$*;*

(iii) $\forall \ell, k$*,* $E[\Upsilon_{\ell k,it}|\Psi_{\ell k,t}, z_{\ell,it}] = 0$*,* $E[e_{\ell,it}|\beta_{\ell,t}, x_{\ell,it}, z_{\ell,it}, \Psi_{\ell k,t}, \Upsilon_{\ell k,it}] = 0$*,* $\forall (i, t)$*.*

Assumption 2.3.3. $\forall \ell, k, i, t$*, elements in* $\Psi_t = (\psi_{\ell k,t})$*,* $\Upsilon_{it} = (\upsilon_{\ell k,it})$ *satisfy the following smoothness condition:*

$$\left|\psi_{\ell k,t} - \psi_{\ell k,s}\right| \leqslant \Big(\frac{|t-s|}{T}\Big)^{\gamma_2} q^{(1)}_{\ell k,ts}, \quad \left|\upsilon_{\ell k,it} - \upsilon_{\ell k,is}\right| \leqslant \Big(\frac{|t-s|}{T}\Big)^{\gamma_2} q^{(2)}_{\ell k,i,ts}, \quad t, s = 1, 2, \cdots, T \tag{2.16}$$

for some $0 < \gamma_2 < 1$ *and the distribution of each variable in* $X^{(2)}_u = \{\psi_{\ell k,t}, \upsilon_{\ell k,t}, q^{(1)}_{\ell k,ts}, q^{(2)}_{\ell k,i,ts}\}$ *has a thin tail* $\mathcal{E}(\alpha)$*:*

$$\mathbb{P}(\left|X^{(2)}_u\right| \geqslant \omega) \leqslant \exp(-c_1|\omega|^{\alpha}), \quad \omega > 0,$$

for $c_1 > 0$*,* $\alpha > 0$ *which does not depend on* t, s, T*.*

Assumption 2.3.4. *The matrices* $\Sigma_{zz,it} = E(z_{it}z'_{it})$*,* $\Sigma_{\Psi zz\Psi,it} = \Psi'_t\Sigma_{zz,it}\Psi_t$ *and* $\Sigma_{zx,it} = E(z_{it}x'_{it})$ *are such that*

(i) $\max_{t=1,\cdots,T}\left\|\Sigma_{zz,it}\right\|_{sp} < \infty$*,* $\forall i$*;* $\left\|\Sigma_{zz,it} - \Sigma_{zz,is}\right\|_{sp} \leqslant C\,\frac{|t-s|}{T}$*,* $\forall i, t$*;*

(ii) $\max_{t=1,\cdots,T}\left\|\Sigma_{zx,it}\right\|_{sp} < \infty$*,* $\forall i$*;* $\left\|\Sigma_{zx,it} - \Sigma_{zx,is}\right\|_{sp} \leqslant C\,\frac{|t-s|}{T}$*,* $\forall i, t$*;*

(iii) $\max_{t=1,\cdots,T}\left\|\Sigma_{\Psi zz\Psi,it}\right\|_{sp} < \infty$*,* $\forall i$*, a.s..*

Since we introduce the first stage regression in our IV context, restrictions on moments and mixing conditions are imposed for z_{it} and v_{it} as in Assumptions 3.1. In Assumption 3.2, we impose similar conditions for random coefficients Ψ_{it} as for β_{it}. Additional assumptions, as in Assumption 3.2(ii)-(iii), are required for identification[4]. Assumption 3.3 imposes similar smooth conditions on newly introduced time-varying components: ψ_t and Υ_{it}, but we have a different smoothness parameter γ_2 and random components $q^{(1)}_{\ell k,ts}$, $q^{(2)}_{\ell k,ts}$. Assumption 3.4(i) parallels Assumptions 2.4(i) and 2.4(ii). Assumption 3.4(ii) ensures that, given i, $E(z_{it}x'_{it})$ has full column rank, $\forall t$, which implies that identification condition holds for all t. Assumption 3.4(iii) is taken from (20) in Giraitis et al. [2020a], which is a high level assumption to ensure that certain sums are uniformly bounded in probability.

[4]Note that, unlike the micro panel literature, these assumptions rule out the correlated random coefficient panel data model. In our context, endogeneity is caused by the fact that x_{it} are correlated with u_{it}.

The main objectives in this section are to construct consistent estimates for β_t and to derive asymptotic distributions for the estimator of β_t. We aim to generalize the estimators proposed in Chen [2015] and Giraitis et al. [2020a] to the panel setting. The individual specific estimator can be obtained as the kernel-based two-stage least square estimator (2SLS):

$$\hat{\beta}^{IV}_{i,t} = \Big(\sum_{j=1}^{T} b_{j,t}(H)\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)^{-1}\Big(\sum_{j=1}^{T} b_{j,t}(H)\hat{\Psi}'_j z_{ij} y_{ij}\Big), \tag{2.17}$$

where $b_{j,t}(H)$ is the kernel function with bandwidth H and $\hat{\Psi}_j$ are the consistent estimates of Ψ_j. At the first stage, $\hat{\Psi}_t$ can be easily obtained as either the TVP-OLS-MG estimator

$$\hat{\Psi}_{MG,t} = \frac{1}{N}\sum_{i=1}^{N}\hat{\Psi}_{i,t}, \tag{2.18}$$

where

$$\hat{\Psi}_{i,t} = \Big(\sum_{j=1}^{T} b_{j,t}(L) z_{ij} z'_{ij}\Big)^{-1}\Big(\sum_{j=1}^{T} b_{j,t}(L) z_{ij} x'_{ij}\Big),$$

or the TVP-OLS-P estimator

$$\hat{\Psi}_{P,t} = \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{j,t}(L) z_{ij} z'_{ij}\Big)^{-1}\Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{j,t}(L) z_{ij} x'_{ij}\Big), \tag{2.19}$$

where $b_{j,t}(L)$ is the kernel function with bandwidth L, where L can be different from H. We assume that both bandwidth parameters H, L satisfy (2.8).

Define

$$r_{N,T,H,\gamma,\alpha} = \Big(\frac{H}{T}\Big)^{\gamma}\log^{1/\alpha} T + \frac{\log^{1/\alpha} T}{\sqrt{N}} + \sqrt{\frac{\log T}{NH}}. \tag{2.20}$$

In the next lemma, we establish uniform consistency results for the estimates $\hat{\Psi}_{MG,t}$ and $\hat{\Psi}_{P,t}$. Since Assumptions 2.1–2.4 are satisfied for the first stage regression, it follows immediately from Theorem 1(1), with a possibly different bandwidth L and smoothness parameter γ_2.

Lemma 2.3.1. *Under Assumptions 3.1–3.4 and assuming that the bandwidth parameters L satisfy* (2.8)*, as* $(N, T) \to \infty$*, we have*

$$\max_{t=1,2,\cdots,T}\big\|\hat{\Psi}_{\int,t} - \Psi_t\big\|_{sp} = O_p(r_{N,T,L,\gamma_2,\alpha}).$$

Then, at the second stage, we propose two estimators for β_t. As in the previous section, we consider both mean group and pooled estimators. The TVP-IV-MG estimator is defined as

$$\hat{\beta}^{IV}_{MG,t} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{IV}_{i,t}, \tag{2.21}$$

where $\hat{\beta}_{i,t}^{IV}$ is given by (2.17). The TVP-IV-P estimator can be computed as

$$\hat{\beta}_{P,t}^{IV} = \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{j,t}(H)\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)^{-1}\Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{j,t}(H)\hat{\Psi}'_j z_{ij} y_{ij}\Big). \tag{2.22}$$

In the next theorem, we establish uniform consistency rates and asymptotic distributions for (2.21) and (2.22).

Theorem 2.3.2. *Under Assumptions 2.1–2.3 (except Assumption 2.2(iii)), Assumptions 3.1–3.4 and assuming that the bandwidth parameters H, L satisfy* (2.8)*, then as $(N, T) \to \infty$ we have the following:*

1. *Uniform consistency: $\hat{\beta}_{MG,t}^{IV}$, $\hat{\beta}_{P,t}^{IV}$ have the property*

$$\max_{t=1,2,\cdots,T}\left\|\hat{\beta}_{\int,t} - \beta_t\right\| = O_p\left(\frac{1}{\sqrt{N}}\Big((\log T)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\Big)\right),$$

   where $\int = \{MG, P\}$ and $r_{NT,H,\alpha}$, $r_{NT,L,\alpha}$ are defined as in (2.20)*;*

2. *Asymptotic normality: Suppose that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$, $L = o\left(L/\log^{\frac{\gamma_2}{\alpha}} T\right)$, $\log^{1/\alpha} T = o(N^{-1/2})$, $k = p$,*

   (i) Mean group estimator:

$$\sqrt{N}\,\Psi_t\left(\Omega_{e,t}^{IV}\right)^{-1/2}\left(\hat{\beta}_{MG,t}^{IV} - \beta_t\right) \xrightarrow{d} N(0, I_k),$$

   where $\Omega_{e,t}^{IV}$ is given by

$$\Omega_{e,t}^{IV} = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{zz,it}^{-1}\Sigma_{ze,it}e_{it}\Big).$$

   (ii) Pooled estimator: Suppose that $\overline{\Sigma}_{\Psi zz\Psi,t} = \text{plim}_{(N,T)\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\Psi'_j z_{ij} z'_{ij}\Psi_j = O_p(1)$, then

$$\sqrt{N}\,\overline{\Sigma}_{\Psi zz\Psi,t}\,\Psi_t^{-1}\left(R_{P,t}^{IV}\right)^{-1/2}\left(\hat{\beta}_{P,t}^{IV} - \beta_t\right) \xrightarrow{d} N(0, I_k),$$

   where $R_{P,t}^{IV}$ is given by

$$R_{P,t}^{IV} = \lim_{N\to\infty} Var\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{ze,it}e_{it}\Big).$$

Several comments are in order. First, the uniform rates derived are different from (21) in Giraitis et al. [2020a] in the time series setting. The first term in the parenthesis $(\log T)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha}$ is nearly identical to (21) in Giraitis et al. [2020a], except for the fact that $\gamma_1 = 1/2$ there. The second term is different, since first stage estimates $\hat{\Psi}_t$ are obtained from panel regression, N appears in the rates. The overall rates are also scaled by root N, indicating that panel dimension is useful to improve the rates. Second, as in Giraitis et al. [2020a], we only present the exact identification case: $k = p$. If the model is overidentified: $p > k$, Ψ_t is not invertible and appears in the limiting distribution, inference is complex. Third, since asymptotic normality requires that the first stage estimates $\hat{\Psi}_t$ are uniformly consistent, additional two conditions, $L = o\left(L/\log^{\frac{\gamma_2}{\alpha}} T\right)$ and $\log^{1/\alpha} T = o(N^{-1/2})$, are needed to achieve asymptotic normality. The later condition implies that T cannot diverge too fast compared to N. Finally, unlike Theorem 1, for both $\hat{\beta}^{IV}_{MG,t}$ and $\hat{\beta}^{IV}_{P,t}$, asymptotic normality relies on weighted averages of e_{it}. Normal approximations of $\hat{\beta}^{IV}_{MG,t}$ and $\hat{\beta}^{IV}_{P,t}$ are very different. The finite sample performance may differ more compared to OLS case. For pooled estimator, we do require another condition $\overline{\Sigma}_{\Psi zz\Psi,t} = O_p(1)$, making inference more involved.

As in (2.9), the above inference results can be operationalised by replacing Ψ_t with $\hat{\Psi}_t$ and the fact that

$$\hat{\Omega}^{IV}_{e,t} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\beta}^{IV}_{i,t} - \hat{\beta}^{IV}_{MG,t}\right)\left(\hat{\beta}^{IV}_{i,t} - \hat{\beta}^{IV}_{MG,t}\right)' \xrightarrow{p} \Psi_t \Omega^{IV}_{e,t} \Psi'_t$$

$$\hat{R}^{IV}_{P,t} = \frac{1}{N}\sum_{i=1}^{N}\left[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)(\hat{\beta}^{IV}_{i,t} - \hat{\beta}^{IV}_{MG,t})(\hat{\beta}^{IV}_{i,t} - \hat{\beta}^{IV}_{MG,t})'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)\right] \xrightarrow{p} \Psi_t R^{IV}_{P,t}\Psi'_t.$$

As the endogeneity of x_{it} can change over time, we can apply the Hausman test for the null hypothesis of exogeneity for each t. The pointwise null hypothesis is formally stated as: $\mathcal{H}_0 : E(u_{it}v_{it}) = 0, \forall i$, for each t, $t = 1, 2, \cdots, T$. A common formulation of the Hausman test is in terms of the quadratic differences between the (time-varying) OLS and IV estimators. In our case, it is given by

$$\mathcal{H}_t = N\left(\hat{\beta}^{IV}_{\int,t} - \hat{\beta}_{\int,t}\right)' \hat{V}_H^{-1}\left(\hat{\beta}^{IV}_{\int,t} - \hat{\beta}_{\int,t}\right), \tag{2.23}$$

where $\hat{V}_H = \text{Avar}\left(\hat{\beta}^{IV}_{\int,t} - \beta_t\right) - \text{Avar}\left(\hat{\beta}_{\int,t} - \beta_t\right)$, $\int = \{MG, P\}$ and Avar is the asymptotic variance given in Theorems 1 and 2. Under Assumptions in Theorems 1 and 2, it can easily be shown that for each t, $\mathcal{H}_t \xrightarrow{d} \chi^2_k$, under the pointwise null hypothesis, that the regressors are exogenous for each t.

Of course, all the other test statistics in the IV framework can be generalized to the panel data model with time-varying coefficients. For example, a time-varying version of the $\mathcal{J}$-test for overidentifying restrictions is given by

$$\mathcal{J}_t = N\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt} z'_{ij}\hat{u}_{ij}\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt} z_{ij} z'_{ij}\Big)^{-1}\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt} z'_{ij}\hat{u}_{ij}\Big), \tag{2.24}$$

where $\hat{u}_{ij} = y_{ij} - \mathbf{x}'_{ij}\hat{\beta}_t$. It can be easily shown that for each t, $\mathcal{J}_t \xrightarrow{d} \chi^2_{p-k}$, under the null hypothesis of valid overidentifying restrictions and assumptions in deriving Theorem 2.

2.4 Monte Carlo study

In this section, we conduct Monte-Carlo experiments to evaluate the finite sample performance of the time-varying OLS and IV estimators (2.6), (2.7), (2.21), (2.22) and the pointwise time-varying Hausman test (2.23). We generate data using the model defined in (2.11) with one regressor x_{it}:

$$\begin{aligned} y_{it} &= x_{it}'\beta_{it} + u_{it}, \\ x_{it} &= \Psi_{it}'z_{it} + c_1 v_{it}, \quad i = 1, 2, ..., N, \;\; t = 1, 2, ..., T, \end{aligned}$$

where we introduce an additional parameter c_1 to control the strength of instruments and we set it to 0.5. We introduce time-varying correlation between u_{it} and v_{it} by specifying them as

$$u_{it} = a(\alpha_{it} + s)e_{1,it} + e_{2,it}, \quad v_{it} = (\alpha_{it} + s)e_{1,it} + e_{3,it},$$

where $e_{j,it}$, $j = 1, 2, 3$ are generated independently from $N(0, 1)$ and $s = 1$. We set $a = 1$ if x_{it} is endogenous, otherwise (if exogenous) we set $a = 0$. We also introduce a time-varying component α_{it} to measure the time-varying correlations between u_{it} and v_{it}.

The time-varying parameters β_{it}, Ψ_{it} and α_{it} are generated according to

$$\begin{aligned} \beta_{it} &= \beta_t + e_{it} \\ \Psi_{it} &= \Psi_t + \Upsilon_{it} \\ \alpha_{it} &= \alpha_t + \iota_{it}, \end{aligned}$$

where elements in $X_{MC}^{(1)} = \{\beta_t, e_{it}, \psi_t, \Upsilon_{it}, \alpha_t, \iota_{it}\}$ are generated from the scaled random walk processes, such that $X_{\ell,t} = \xi_{\ell,t}/\sqrt{t}$, for $\xi_{\ell,t} - \xi_{\ell,t-1} \overset{i.i.d.}{\sim} N(0, 1)$, $t = 1, 2, \cdots, T$. The instrument z_{it} is again generated from $N(0, 1)$ and it is independent from $e_{j,it}$, $j = 1, 2, 3$ and elements in $X_{MC}^{(1)}$. Each experiment was replicated 1,000 times for each (N, T) pair with $N, T = 50, 100, 200, 500$.

To compute both TVP-OLS and TVP-IV type estimators,[5] we use a two-sided normal kernel $K(x) = \exp(-x^2/2)$ with bandwidth set to take values $T^{\overline{\alpha}}$ for $\overline{\alpha} = 0.2, 0.4, 0.5, 0.7$. Lower values of $\overline{\alpha}$ increase the robustness of estimates to parameter changes but decrease efficiency, and in the panel case lower values for $\overline{\alpha}$ also make it more likely the condition $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ holds. Thus, it is interesting to evaluate the impact of the bandwidth on the performance of estimators and of the pointwise Hausman test for exogeneity. The global performance of the estimators is evaluated by both the average of median absolute deviations (MADs), $\frac{1}{M}\sum_{r=1}^{M} med_{t=1,2,\cdots,T}\left|\hat{\beta}_{r,t} - \beta_{r,t}\right|$, and 95% coverage rates, computed starting from the second-half of the sample period, $t = [T/2] + 1, \cdots, T$. For the time-varying Hausman test, we report both the size and power of the test evaluated at the middle point $t = [T/2]$.

To get a rough idea of the estimates and 95% confidence intervals, we report a single replication of

[5] To compute those estimators, we normalize all kernel weights so that $\sum_{j=1}^{T} b_{jt}(H) = 1$ and $\sum_{j=1}^{T} b_{jt}(L) = 1$.

the estimates for $N = 50, T = 200$ and $H = T^{0.5}$, $L = H = T^{0.5}$ in Figures 1 and 2. Evidently, TVP-OLS estimators perform well when x_{it} is exogenous, but they lead to substantial bias and poor coverage rates when x_{it} is endogenous. The performance of TVP-IV estimators is quite satisfactory under the case of endogeneity of x_{it}.

Table 1 reports the average MAD and coverage probability of both TVP-OLS-MG and TVP-OLS-P estimators for the normal kernel with various values of bandwidth H. A number of comments can be made. First, both cross-sectional and time series dimensions are useful to reduce the bias of the estimates. All values become closer to zero as (N, T) increases. Second, regarding the bandwidth parameter $H = T^{\overline{\alpha}}$, smaller values of $\overline{\alpha}$ often yield the lowest values of MAD and higher coverage probabilities. However, results are very similar if $\overline{\alpha}$ is replaced with a value lower than 0.5. If $\overline{\alpha}$ is larger than 0.5, MAD increases and coverage probability decreases. Recall that the bandwidth parameter $H = T^{\overline{\alpha}}$ is required to be such that $H = o(T)$ and asymptotic normality requires $\frac{\sqrt{N}}{T^{(1-\overline{\alpha})\gamma}} \to 0$. This means that we cannot set $\overline{\alpha}$ to be a value close to 1 otherwise these conditions will fail. Finally, TVP-OLS-MG and TVP-OLS-P estimates deliver very similar results.

Tables 2 and 3 present average MADs and coverage probabilities, respectively, in the case of endogeneity. For the bandwidth parameters, we set $H = L$. We report the results from TVP-IV type estimates obtained either from mean group or pooled estimators at the first stage. For comparison purposes, we also include OLS results in the bottom panel of each table. First, TVP-OLS estimates clearly yield much large biases and lower coverage probabilities when x_{it} is endogenous. There is no sign of convergence as (N, T) increases, as expected. Second, TVP-IV estimates have much better performance with significantly lower MADs and higher coverage probabilities. Regarding the bandwidth, it seems that setting $H = T^{0.2} \sim T^{0.4}$ leads to the lowest MAD and $H = T^{0.2}$ has the best overall performance. However, smaller bandwidth is often associated with larger confidence intervals, which implies higher uncertainty around point estimates. In general, $H = T^{0.4}$ and $H = T^{0.5}$ deliver comparable results with $H^{0.2}$. Recall that asymptotic normality again requires $\frac{\sqrt{N}}{T^{(1-\overline{\alpha})\gamma}} \to 0$. If we let $\overline{\alpha}$ increase, we should also let T be much larger than N to make the above condition hold. As shown in Tables 2 and 3, when $\overline{\alpha}$ gets larger, the lowest MADs are obtained when $(N, T) = (500, 500)$ and the best coverage probabilities are always obtained when $(N, T) = (50, 500)$. Third, it does not seem to make a difference whether $\hat{\Psi}_j$ is obtained either by mean group or pooled estimates at the first stage. However, when the bandwidth is small, $H = T^{0.2}$, the pooled estimator seems to be slightly worse than the mean group estimator for MAD at the second stage, but not for coverage probability. As the bandwidth H increases, both mean group and pooled estimators deliver similar results in terms of both MAD and coverage probability.

Table 4 reports the size and power of the time-varying Hausman test with nominal size equal to 5%. The bandwidth parameters are set to $H = L$. First, we find that setting $H = T^{0.2}$ or $H = T^{0.4}$ leads to sizes close to the nominal value. As the bandwidth H increases, size distortion becomes sizable, especially when N is larger than T. Recall that an asymptotically χ^2 distribution of the test statistic requires asymptotic normality of the estimators to hold. As explained in the previous paragraph, if we increase α, we also need very large T (compared to N) to make the condition $\frac{\sqrt{N}}{T^{(1-\bar{\alpha})\gamma}} \to 0$ hold. For instance, consider the case when $\hat{\Psi}_j$ is obtained by pooled estimator and $H = T^{0.4}$; empirical size is 0.033 when $(N, T) = (50, 500)$, which is close to 0.05. However, when $(N, T) = (500, 50)$, empirical size becomes 0.256, which is clearly oversized. Second, in terms of power, because the convergence rate of the estimator is the square root of N, we expect that power increases with N, and this is confirmed by Table 4. The power also increases slightly with bandwidth H, and with larger (N, T). Setting $H = T^{0.5}$ leads to the highest power when $(N, T) = (500, 500)$. In general, test power is very similar across different values of bandwidth H.

2.5 Empirical application

In this section, we consider an empirical application on modeling inflation dynamics. We estimate a multi-country version of the Phillips curve that links inflation to unemployment and also possibly to inflation expectations. The main goals are to understand whether unemployment is indeed significantly related to inflation, when and if forward looking behavior of inflation dominates backward looking behavior and whether there are changes in these features over time.

We use monthly data for 19 Eurozone countries[6] over the period 2000M1–2019M12. We consider the hybrid Phillips curve, along the lines of Galí and Gertler [1999],

$$\pi_{i,t} = c_{i,t} + \gamma_{i,t}\pi_{i,t-1} + \alpha_{i,t}u_{i,t} + \rho_{i,t}\pi^e_{i,t+1} + \nu_{i,t}. \tag{2.25}$$

Let $\epsilon_{i,t} = \rho_{i,t}(\pi^e_{i,t+1} - \pi_{i,t+1}) + \nu_{i,t}$, we can also write the above as

$$\pi_{i,t} = c_{i,t} + \gamma_{i,t}\pi_{i,t-1} + \alpha_{i,t}u_{i,t} + \rho_{i,t}\pi_{i,t+1} + \epsilon_{i,t}. \tag{2.26}$$

It is clear that, since $\pi^e_{i,t+1}$ is not observed, we have to replace it by $\pi_{i,t+1}$, endogeneity arises again due to measurement error. In view of Hansen and Lunde [2014] and Galí and Gambetti [2019], endogeneity may also arise due to measurement error of $\pi_{i,t}$ and simultaneity of $\pi_{i,t}$ and $u_{i,t}$. We use an intercept, three lags of inflation and two lags of unemployment as instruments: $(1, \pi_{i,t-2}, \pi_{i,t-3}, \pi_{i,t-4}, u_{i,t-1}, u_{i,t-2})'$[7].

Figure 2.3 provides plots of estimates for ρ_t, α_t and γ_t. Left panel of Figure 2.3 provides plots of time-varying estimates for ρ_t, the common part of coefficients on inflation expectations. OLS estimates are small (less than 0.2), only significant around 2005–2013 and remain almost stable over time. IV estimates are more volatile, and are significant in shorter periods, around 2004 and 2014. Middle panel of Figure 2.3 provides plots of time-varying estimates of $\hat{\alpha}_t$, the common part of coefficients on unemployment. Point estimates generally fluctuate below the zero line, which indicates that there is a trade-off between unem-

[6] Data are taken from Eurostat. The countries are Austria, Belgium, Cyprus, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Portugal, Slovenia, Slovakia and Spain.

[7] Model (2.26) is a dynamic heterogeneous panel and lagged variables are used as instruments. As is well known in the literature [e.g., Pesaran and Smith, 1995], pooling gives inconsistent estimates. Thus, we only report results for mean group estimates.

ployment and inflation, but the coefficients are very small (less than -0.1) over the whole sample period. The estimates from both OLS and IV are roughly similar (except around 2008–2014), but IV estimates are smaller and are less "significant" than OLS estimates. Both estimates are statistically significant and very similar around 2014–2018. Moreover, IV estimates are significant around 2005, but OLS estimates are significant around 2005-2010. Turning to the persistence parameter estimates, which are shown in the right panel of Figure 2.3, we see that OLS estimates are small (less than 0.2) and significant from 2016–2013. However, IV estimates are much larger until 2015. Our findings are in line with Hansen and Lunde [2014], who find that OLS estimates are biased towards zero for persistent parameter when time series are measured with error.

Plots of p-values of time-varying pointwise Hausman test from (2.26) are provided in Figure 2.4. The null of exogeneity is rejected in most of the period (until 2016). Indeed, we see from Figure 2.3 that OLS and IV estimates are very different for ρ_t and γ_t, but estimates for α_t from OLS and IV are almost identical after 2016.

As a robustness check, in Appendix C, we report estimation results and p-value of time-varying pointwise Hausman test from both backward-looking Phillips curve and forward-looking Phillips curve, obtained either by setting $\rho_{i,t} = 0$ or $\gamma_{i,t} = 0$. The results are similar compared to Figures 2.3 and 2.4, but there are also some differences, particularly for forward-looking Phillips curve. First, IV estimates for α_t are also significant around 2005–2008. Second, there are shorter periods in which the null of exogeneity is rejected, round around 2005, 2010–2012 and 2014.

In summary, this simple but economically interesting empirical application highlights the importance of allowing for parameter time variation and usefulness of the IV method. There is a small but varying impact of unemployment on inflation. OLS deliver smaller persistence parameter estimates than IV. Forward-looking feature only dominates backward-looking feature for inflation around year 2015. Endogeneity may not only come from inflation expectation but also measurement error of inflation series. In terms of sources of endogeneity, measurement error is likely to be more important than unobservable of inflation expectation and simultaneity between inflation and unemployment. Ignoring these features lead to bias and misleading results.

2.6 Conclusion

Large heterogeneous panel data models are becoming increasingly popular in empirical applications, but the parameters are typically assumed to be constant over time and regressors are treated as exogenous. However, the vast literature on panels, structural change and parameter instability has highlighted the importance of considering both time variation of parameters and endogeneity. In this paper, we introduce a new class of large heterogeneous panel data models whose parameters are not only heterogeneous but also vary stochastically over time. We propose time-varying mean group and pooled least square and IV estimators, taking a non-parametric approach in order to remain as agnostic as possible regarding the type of parameter evolution.

We derive theoretical properties for the proposed time-varying mean group and pooled estimators in both the least square and IV contexts. We show the uniform consistency and derive the asymptotic distributions of the proposed estimators. We also propose a pointwise time-varying Hausman exogeneity test, which compares time-varying least square and IV estimators, possibly also allowing for changes in the endogeneity status of the regressors over time.

Next, we evaluate the finite sample properties of the estimators and size and power of the time-varying Hausman tests in an extensive Monte Carlo study. The results show that least square type estimates perform very well when regressors are exogenous, but have large biases and low coverage probabilities when regressors are endogenous. The IV type estimates have small finite-sample biases and satisfactory coverage probabilities when regressors are endogenous, especially if the bandwidth is chosen to be a value smaller than $T^{0.5}$. The size of the time-varying Hausman test statistic is also reasonable if bandwidth is smaller than $T^{0.5}$. The test has a good power and is not strongly affected by the bandwidth choice.

Finally, we provide an empirical application to illustrate in practice the use of time-varying mean group and pooled estimators. We estimate the panel version of time-varying hybrid Phillips curves for 19 Eurozone countries over the period 2000M1–2019M12. This simple but economically meaningful empirical application highlights the relevance of allowing for both parameter time variation and endogeneity in the panel framework.

Figures and Tables

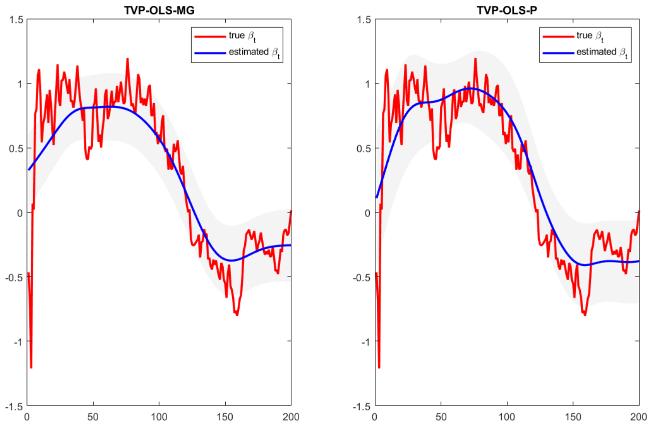

Figure 2.1: Realization of β_t, TVP-OLS-MG and TVP-OLS-P estimates with a two-sided normal kernel and $H = T^{0.5}$ for $(N, T) = (50, 200)$. The solid red lines show the true realization of β_t. The solid blue lines show the point estimates and the grey shaded areas show the 95% pointwise confidence intervals.

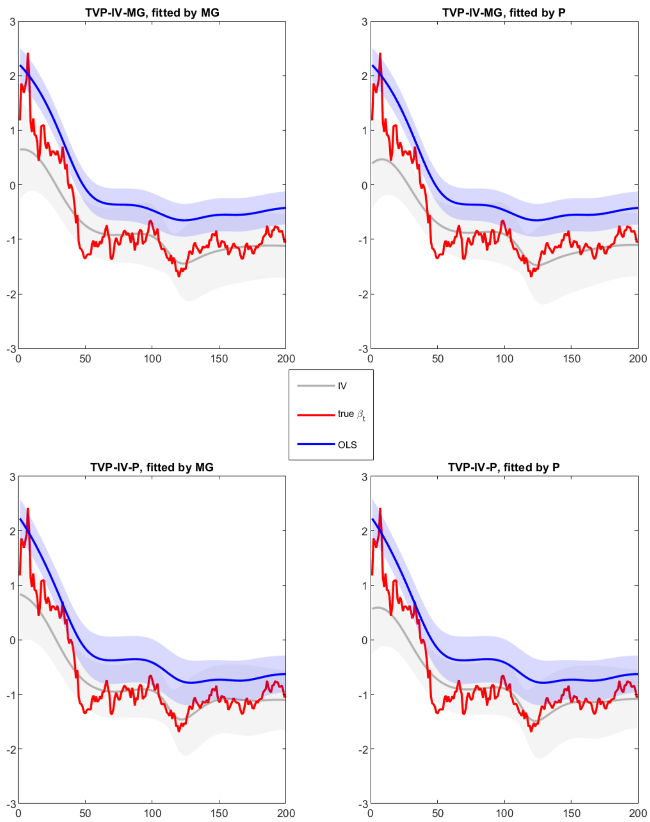

Figure 2.2: Realization of β_t, TVP-OLS-MG, TVP-IV-MG, TVP-OLS-P and TVP-IV-P estimates with a two-sided normal kernel and $H = T^{0.5}$, $H = L = T^{0.5}$ for $(N, T) = (50, 200)$. The solid red lines show the true realization of β_t. The solid blue lines show the OLS point estimates and the blue shaded areas show the 95% pointwise confidence intervals. The solid grey lines show the IV point estimates and the grey shaded areas show the 95% pointwise confidence intervals.

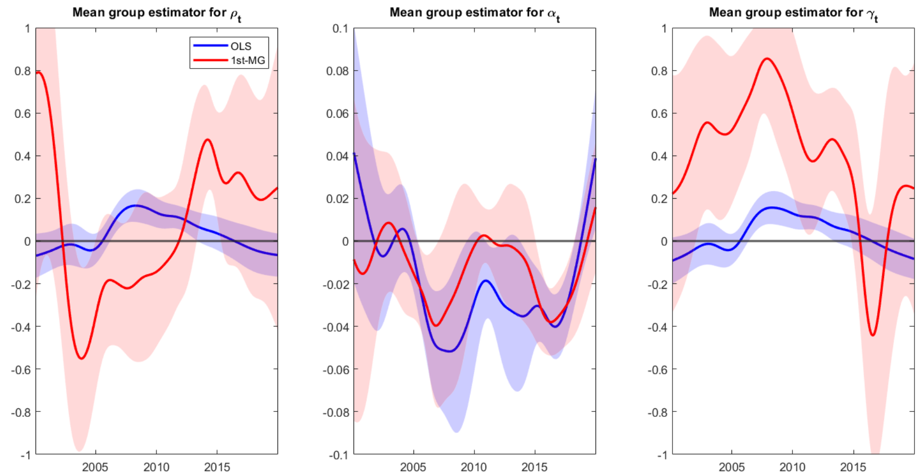

Figure 2.3: Empirical results for model (2.26). The solid blue lines show the point TVP-OLS estimates and the blue shaded areas show the 95% pointwise confidence intervals. The red solid lines show the point TVP-IV estimates and the red shaded areas show the 95% pointwise confidence intervals.

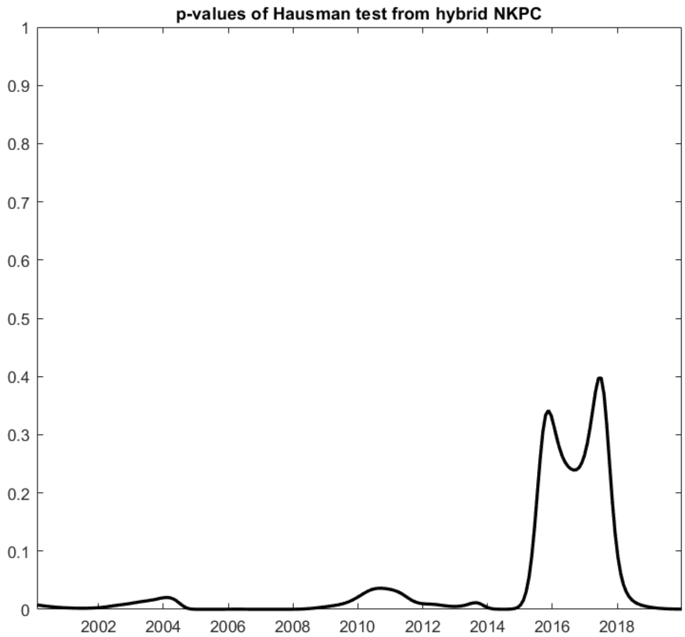

Figure 2.4: p-values of time-varying Hausman test for model (2.26).

(N,T)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
MAD	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-OLS-MG																
50	0.157	0.143	0.131	0.126	0.194	0.165	0.150	0.141	0.217	0.193	0.172	0.158	0.287	0.271	0.248	0.227
100	0.132	0.119	0.104	0.091	0.179	0.146	0.129	0.112	0.204	0.176	0.154	0.132	0.279	0.260	0.238	0.212
200	0.117	0.106	0.086	0.071	0.172	0.136	0.116	0.097	0.199	0.169	0.145	0.120	0.276	0.258	0.232	0.204
500	0.109	0.095	0.073	0.055	0.167	0.129	0.108	0.086	0.196	0.164	0.139	0.112	0.280	0.253	0.232	0.201
TVP-OLS-P																
50	0.161	0.147	0.136	0.131	0.195	0.167	0.151	0.142	0.217	0.193	0.172	0.158	0.287	0.270	0.248	0.227
100	0.134	0.121	0.107	0.095	0.177	0.146	0.129	0.112	0.202	0.176	0.154	0.132	0.277	0.259	0.238	0.212
200	0.116	0.106	0.087	0.074	0.169	0.135	0.116	0.097	0.197	0.168	0.144	0.120	0.274	0.257	0.232	0.204
500	0.105	0.094	0.073	0.056	0.164	0.127	0.107	0.086	0.193	0.162	0.138	0.112	0.278	0.252	0.231	0.201
Coverage probability	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-OLS-MG																
50	0.875	0.889	0.916	0.917	0.765	0.828	0.865	0.880	0.702	0.756	0.807	0.840	0.530	0.565	0.612	0.664
100	0.813	0.847	0.887	0.923	0.652	0.748	0.795	0.851	0.578	0.652	0.708	0.784	0.410	0.439	0.485	0.548
200	0.720	0.753	0.838	0.898	0.522	0.625	0.695	0.771	0.447	0.515	0.585	0.672	0.305	0.320	0.363	0.428
500	0.552	0.596	0.717	0.838	0.359	0.451	0.526	0.622	0.298	0.359	0.416	0.505	0.190	0.220	0.243	0.284
TVP-OLS-P																
50	0.888	0.894	0.920	0.921	0..783	0.838	0.873	0.884	0.702	0.756	0.807	0.840	0.543	0.573	0.619	0.666
100	0.838	0.860	0.894	0.926	0.673	0.763	0.805	0.856	0.578	0.652	0.708	0.784	0.423	0.446	0.489	0.551
200	0.763	0.783	0.858	0.907	0.552	0.646	0.712	0.779	0.447	0.515	0.585	0.672	0.314	0.325	0.368	0.430
500	0.607	0.637	0.749	0.857	0.385	0.475	0.544	0.635	0.298	0.359	0.416	0.505	0.198	0.224	0.246	0.285

Note: See Section 4 for details on the computations of mean absolute deviation (MAD) and coverage probability.

Table 2.1: Small sample properties of TVP-OLS-MG and TVP-OLS-P estimators in the case of an exogenous regressor: Average MAD and coverage probability

	MAD															
(N,T)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
IV, MG in first stage	$H=T^{0.2}$				$H=T^{0.4}$				$H=T^{0.5}$				$H=T^{0.7}$			
TVP-IV-MG																
50	0.332	0.289	0.274	0.264	0.359	0.301	0.277	0.260	0.391	0.333	0.298	0.273	0.516	0.456	0.404	0.367
100	0.249	0.233	0.205	0.196	0.289	0.259	0.221	0.202	0.318	0.298	0.255	0.220	0.434	0.427	0.377	0.332
200	0.213	0.189	0.165	0.148	0.268	0.219	0.189	0.166	0.305	0.264	0.225	0.193	0.435	0.394	0.355	0.303
500	0.177	0.154	0.125	0.106	0.244	0.193	0.161	0.136	0.285	0.240	0.205	0.169	0.428	0.376	0.333	0.303
TVP-IV-P																
50	0.352	0.307	0.293	0.283	0.372	0.314	0.284	0.264	0.401	0.340	0.302	0.276	0.529	0.458	0.408	0.368
100	0.262	0.242	0.217	0.207	0.293	0.264	0.225	0.206	0.321	0.301	0.257	0.222	0.436	0.431	0.376	0.333
200	0.220	0.197	0.173	0.156	0.273	0.224	0.192	0.169	0.307	0.267	0.226	0.194	0.438	0.396	0.355	0.304
500	0.180	0.158	0.129	0.111	0.245	0.196	0.163	0.138	0.285	0.241	0.207	0.170	0.428	0.376	0.333	0.303
IV, P in first stage	$H=T^{0.2}$				$H=T^{0.4}$				$H=T^{0.5}$				$H=T^{0.7}$			
TVP-IV-MG																
50	0.333	0.291	0.276	0.267	0.357	0.301	0.277	0.260	0.386	0.333	0.296	0.273	0.516	0.455	0.402	0.366
100	0.251	0.233	0.207	0.198	0.288	0.258	0.222	0.202	0.317	0.299	0.256	0.220	0.430	0.427	0.372	0.332
200	0.213	0.188	0.167	0.149	0.267	0.220	0.190	0.166	0.303	0.263	0.224	0.193	0.429	0.394	0.355	0.302
500	0.177	0.154	0.125	0.107	0.241	0.193	0.161	0.136	0.284	0.239	0.204	0.169	0.424	0.376	0.332	0.303
TVP-IV-P																
50	0.337	0.295	0.282	0.272	0.361	0.307	0.280	0.261	0.394	0.336	0.300	0.274	0.517	0.455	0.405	0.366
100	0.251	0.236	0.209	0.200	0.289	0.261	0.222	0.204	0.316	0.300	0.255	0.221	0.430	0.429	0.371	0.332
200	0.212	0.192	0.168	0.151	0.268	0.221	0.190	0.168	0.304	0.264	0.225	0.193	0.431	0.394	0.355	0.303
500	0.175	0.154	0.125	0.108	0.241	0.194	0.161	0.136	0.283	0.240	0.205	0.169	0.424	0.376	0.332	0.303
OLS	$H=T^{0.2}$				$H=T^{0.4}$				$H=T^{0.5}$				$H=T^{0.7}$			
TVP-OLS-MG																
50	0.590	0.589	0.583	0.586	0.586	0.588	0.581	0.584	0.586	0.586	0.580	0.584	0.595	0.592	0.582	0.585
100	0.582	0.585	0.579	0.578	0.580	0.583	0.577	0.576	0.579	0.581	0.576	0.575	0.588	0.586	0.579	0.572
200	0.583	0.573	0.589	0.578	0.580	0.572	0.587	0.577	0.580	0.571	0.586	0.575	0.581	0.576	0.587	0.571
500	0.589	0.569	0.583	0.578	0.585	0.568	0.581	0.577	0.585	0.567	0.580	0.576	0.593	0.574	0.580	0.573
TVP-OLS-P																
50	0.532	0.528	0.518	0.523	0.537	0.532	0.520	0.524	0.542	0.537	0.524	0.526	0.562	0.556	0.540	0.542
100	0.516	0.517	0.507	0.503	0.525	0.520	0.511	0.506	0.529	0.526	0.515	0.510	0.549	0.545	0.534	0.524
200	0.512	0.500	0.512	0.498	0.521	0.506	0.516	0.502	0.526	0.512	0.521	0.505	0.542	0.532	0.539	0.519
500	0.514	0.495	0.507	0.498	0.523	0.500	0.512	0.502	0.528	0.506	0.516	0.506	0.548	0.528	0.533	0.520

Note: See Section 4 for details on the computations of mean absolute deviation (MAD) and coverage probability. For bandwidth parameters for TVP-IV estimators, we set $H = L$.

Table 2.2: Small sample properties of TVP-OLS and TVP-IV types of estimators in the case of an endogenous regressor: Average MAD

	Coverage probability															
(N, T)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
IV, MG in first stage	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-IV-MG																
50	0.939	0.953	0.956	0.961	0.892	0.931	0.940	0.950	0.857	0.896	0.915	0.933	0.758	0.779	0.808	0.841
100	0.928	0.941	0.955	0.963	0.845	0.900	0.922	0.940	0.793	0.842	0.874	0.910	0.647	0.676	0.706	0.752
200	0.893	0.910	0.936	0.957	0.758	0.839	0.871	0.912	0.687	0.748	0.792	0.855	0.536	0.562	0.570	0.644
500	0.807	0.835	0.898	0.946	0.616	0.708	0.775	0.846	0.538	0.592	0.657	0.749	0.380	0.406	0.429	0.483
TVP-IV-P																
50	0.940	0.948	0.954	0.959	0.896	0.929	0.940	0.949	0.866	0.895	0.916	0.934	0.763	0.781	0.809	0.842
100	0.930	0.944	0.952	0.961	0.850	0.903	0.922	0.941	0.800	0.842	0.877	0.911	0.657	0.678	0.709	0.754
200	0.899	0.913	0.939	0.957	0.768	0.844	0.876	0.914	0.698	0.756	0.796	0.858	0.545	0.570	0.574	0.645
500	0.823	0.849	0.908	0.948	0.634	0.721	0.786	0.850	0.557	0.602	0.664	0.753	0.391	0.412	0.433	0.485
IV, P in first stage	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-IV-MG																
50	0.933	0.951	0.954	0.959	0.889	0.930	0.939	0.949	0.854	0.895	0.914	0.933	0.756	0.779	0.808	0.841
100	0.923	0.940	0.952	0.961	0.842	0.898	0.920	0.940	0.789	0.840	0.872	0.910	0.648	0.674	0.708	0.752
200	0.889	0.907	0.934	0.955	0.758	0.838	0.870	0.911	0.686	0.746	0.792	0.855	0.537	0.559	0.569	0.644
500	0.804	0.834	0.895	0.944	0.615	0.708	0.774	0.846	0.536	0.593	0.657	0.748	0.377	0.404	0.429	0.482
TVP-IV-P																
50	0.943	0.952	0.956	0.962	0.899	0.932	0.941	0.951	0.867	0.898	0.916	0.935	0.762	0.782	0.809	0.843
100	0.933	0.946	0.955	0.963	0.853	0.904	0.921	0.942	0.802	0.844	0.877	0.913	0.660	0.677	0.711	0.754
200	0.902	0.918	0.942	0.959	0.771	0.846	0.878	0.915	0.699	0.756	0.798	0.858	0.546	0.567	0.573	0.645
500	0.832	0.853	0.912	0.952	0.636	0.725	0.787	0.852	0.555	0.604	0.667	0.754	0.390	0.411	0.433	0.484
OLS	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
TVP-OLS-MG																
50	0.220	0.203	0.194	0.200	0.216	0.200	0.187	0.189	0.226	0.208	0.195	0.195	0.236	0.225	0.228	0.218
100	0.112	0.088	0.094	0.082	0.130	0.097	0.101	0.085	0.140	0.114	0.116	0.096	0.162	0.150	0.149	0.136
200	0.052	0.045	0.026	0.027	0.077	0.061	0.039	0.035	0.091	0.077	0.055	0.047	0.119	0.107	0.096	0.091
500	0.020	0.015	0.012	0.007	0.040	0.027	0.019	0.012	0.051	0.039	0.029	0.020	0.067	0.067	0.058	0.049
TVP-OLS-P																
50	0.426	0.416	0.428	0.430	0.387	0.393	0.398	0.399	0.372	0.375	0.385	0.390	0.334	0.339	0.358	0.353
100	0.310	0.282	0.302	0.293	0.276	0.261	0.277	0.262	0.268	0.252	0.266	0.255	0.242	0.240	0.249	0.247
200	0.191	0.185	0.166	0.176	0.179	0.177	0.155	0.160	0.179	0.176	0.158	0.163	0.178	0.171	0.162	0.172
500	0.090	0.084	0.067	0.059	0.099	0.090	0.074	0.063	0.104	0.099	0.083	0.074	0.105	0.108	0.101	0.098

Note: See Section 4 for details on the computations of mean absolute deviation (MAD) and coverage probability. For bandwidth parameters for TVP-IV estimators, we set $H = L$.

Table 2.3: Small sample properties of TVP-IV-MG and TVP-IV-P estimators in the case of an endogenous regressor: Coverage probability, $t = [T/2]$

(N,T)	50	100	200	500	50	100	200	500	50	100	200	500	50	100	200	500
Mean group in first stage																
Size	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
Mean group																
50	0.037	0.036	0.031	0.034	0.062	0.049	0.030	0.033	0.090	0.081	0.043	0.039	0.157	0.155	0.142	0.111
100	0.034	0.031	0.027	0.020	0.088	0.059	0.034	0.027	0.143	0.090	0.073	0.042	0.213	0.239	0.194	0.173
200	0.046	0.034	0.028	0.028	0.156	0.074	0.049	0.038	0.222	0.166	0.109	0.079	0.296	0.343	0.334	0.259
500	0.092	0.063	0.041	0.028	0.256	0.156	0.094	0.064	0.383	0.268	0.190	0.123	0.493	0.470	0.478	0.384
Pooled																
50	0.027	0.034	0.025	0.021	0.055	0.045	0.033	0.020	0.086	0.065	0.048	0.030	0.153	0.169	0.158	0.109
100	0.046	0.034	0.025	0.019	0.088	0.058	0.040	0.026	0.144	0.099	0.083	0.035	0.219	0.270	0.227	0.187
200	0.054	0.040	0.026	0.014	0.155	0.081	0.051	0.034	0.233	0.169	0.121	0.071	0.320	0.358	0.346	0.298
500	0.094	0.059	0.035	0.019	0.279	0.176	0.105	0.065	0.412	0.294	0.225	0.130	0.513	0.485	0.501	0.427
Power	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
Mean group																
50	0.415	0.460	0.421	0.405	0.528	0.516	0.500	0.489	0.550	0.571	0.532	0.516	0.570	0.600	0.579	0.600
100	0.595	0.594	0.580	0.566	0.712	0.653	0.659	0.639	0.730	0.693	0.703	0.668	0.717	0.694	0.707	0.767
200	0.725	0.728	0.741	0.698	0.789	0.789	0.791	0.778	0.803	0.806	0.806	0.790	0.790	0.799	0.822	0.809
500	0.838	0.858	0.850	0.832	0.905	0.891	0.905	0.884	0.896	0.894	0.929	0.904	0.875	0.908	0.889	0.901
Pooled																
50	0.363	0.416	0.394	0.346	0.497	0.485	0.468	0.424	0.534	0.546	0.508	0.463	0.553	0.596	0.575	0.569
100	0.518	0.524	0.525	0.491	0.645	0.595	0.609	0.584	0.677	0.648	0.652	0.622	0.684	0.676	0.689	0.740
200	0.663	0.681	0.670	0.642	0.752	0.737	0.743	0.725	0.760	0.767	0.768	0.743	0.775	0.783	0.796	0.800
500	0.795	0.822	0.810	0.787	0.868	0.859	0.878	0.865	0.866	0.868	0.901	0.883	0.858	0.889	0.874	0.879
Pooled in first stage																
Size	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
Mean group																
50	0.040	0.039	0.035	0.034	0.057	0.050	0.032	0.034	0.095	0.077	0.040	0.040	0.158	0.158	0.141	0.111
100	0.028	0.037	0.026	0.025	0.087	0.059	0.038	0.028	0.138	0.090	0.075	0.040	0.212	0.237	0.197	0.175
200	0.051	0.033	0.036	0.027	0.144	0.077	0.053	0.036	0.221	0.162	0.106	0.076	0.297	0.339	0.337	0.261
500	0.094	0.062	0.045	0.030	0.257	0.148	0.091	0.062	0.372	0.265	0.187	0.116	0.491	0.465	0.481	0.380
Pooled																
50	0.023	0.028	0.019	0.023	0.053	0.044	0.028	0.019	0.088	0.066	0.043	0.029	0.157	0.168	0.160	0.107
100	0.039	0.025	0.021	0.016	0.087	0.055	0.040	0.024	0.135	0.107	0.085	0.036	0.220	0.265	0.222	0.187
200	0.038	0.036	0.022	0.015	0.144	0.078	0.049	0.034	0.232	0.160	0.120	0.071	0.316	0.355	0.345	0.295
500	0.079	0.053	0.031	0.018	0.270	0.169	0.101	0.063	0.409	0.290	0.221	0.128	0.517	0.486	0.500	0.427
Power	$H = T^{0.2}$				$H = T^{0.4}$				$H = T^{0.5}$				$H = T^{0.7}$			
Mean group																
50	0.427	0.463	0.421	0.405	0.545	0.529	0.509	0.489	0.554	0.572	0.531	0.513	0.566	0.600	0.576	0.597
100	0.595	0.596	0.577	0.570	0.715	0.656	0.650	0.645	0.743	0.693	0.700	0.672	0.708	0.694	0.710	0.764
200	0.727	0.728	0.739	0.699	0.789	0.793	0.795	0.774	0.804	0.808	0.813	0.786	0.792	0.796	0.823	0.808
500	0.848	0.858	0.852	0.828	0.906	0.891	0.908	0.883	0.895	0.894	0.933	0.910	0.879	0.906	0.890	0.900
Pooled																
50	0.369	0.412	0.389	0.348	0.506	0.480	0.469	0.419	0.535	0.547	0.511	0.463	0.543	0.597	0.572	0.574
100	0.533	0.529	0.523	0.493	0.647	0.604	0.602	0.585	0.679	0.656	0.649	0.627	0.685	0.673	0.686	0.736
200	0.661	0.672	0.668	0.644	0.759	0.738	0.742	0.716	0.764	0.769	0.769	0.741	0.776	0.790	0.794	0.801
500	0.806	0.830	0.818	0.790	0.870	0.859	0.879	0.862	0.865	0.876	0.900	0.885	0.864	0.887	0.874	0.881

Note: $\hat{V}_H$ in the Hausman test statistic (2.23) is computed as in footnote 2. The bandwidth parameters are set to $H = L$.

Table 2.4: Size and power of time-varying Hausman H test for exogeneity

Appendix A: Statement and proof of Lemmas

In the proof, we shall write $b_{jt} = b_{j,t}(H)$ and

$$K_t = \sum_{j=1}^{T} b_{jt}, \quad K_{2,t} = \sum_{j=1}^{T} b_{jt}^2.$$

We will use repeatedly the following property of the weights [see Giraitis et al., 2014]: for $t = [\tau T]$ $(0 < \tau < 1)$, as $H \to \infty$,

$$K_t = O(H), \quad K_{2,t} = O(H).$$

Moreover, we use H for the bandwidth for simplicity of notation and only introduce L when two bandwidth parameters interact.

Appendix A.1: Auxiliary results

Let (ξ_t) be an univariate strong-mixing sequence with mixing coefficient α_k^{ξ} satisfying

$$\alpha_k^{\xi} \leqslant c_{\xi}\phi^k, \;\; k \geqslant 1 \tag{2.27}$$

for some $0 < \phi < 1$ and $c_{\xi} > 0$. Assume that

$$\max_{t=1,2,\cdots,T} E|\xi_t|^{\theta} \leqslant C < \infty \tag{2.28}$$

for some $0 < \theta < 4$. The condition above implies that ξ_t has a fat tail $\mathcal{H}(\theta)$,

$$\mathbb{P}(|\xi_t| \geqslant \omega) \leqslant c|\omega|^{-\theta}, \;\; \omega > 0,$$

for some $c > 0$ which does not depend on t. We shall write $(\nu_t) \in \mathcal{E}(\alpha)$, $\alpha > 0$ to denote that

$$\mathbb{P}(|\nu_t| \geqslant \omega) \leqslant c_0 \exp\big(-c_1|\omega|^{\alpha}\big), \;\; \omega > 0,$$

for some $c_0, c_1 > 0$ do not depend on t.

Lemma A1. *Let (ξ_t) be an univariate strong-mixing sequence satisfying* (2.28) *and mixing coefficient* *satisfying* (2.27). *Consider the sums*

$$S_{T,t} := \frac{1}{\sqrt{K_t}} \sum_{j=1}^{T} b_{jt}(\xi_j - E\xi_j)$$

$$v_{T,t}^{(1)} := \frac{1}{K_t} \sum_{j=1}^{T} b_{jt}\beta_j(\xi_j - E\xi_j)$$

$$v_{T,t}^{(2)} := \frac{1}{K_t} \sum_{j=1}^{T} b_{jt}\beta_j^2(\xi_j - E\xi_j)$$

$$\Delta_{T,t} := \frac{1}{K_t} \sum_{j=1}^{T} b_{jt}(\beta_j - \beta_t)\xi_j,$$

where (β_j) *is such that* $\left|\beta_j - \beta_t\right| \leqslant \left(\frac{|j-t|}{T}\right)^{\gamma} \nu_{jt}$, $0 < \gamma < 1$, $1 \leqslant t, j \leqslant T$ *and* $\beta_t \in \mathcal{E}(\alpha)$, $\nu_{jt} \in \mathcal{E}(\alpha)$ *for some* $\alpha > 0$. *Assume that the bandwidth parameter H satisfies*

$$cT^{1/(\theta/4-1)} \leqslant H \leqslant T$$

for some $c > 0$ *and* $\delta > 0$. *Then, for any* $\epsilon > 0$

$$\max_{t=1,2,\cdots,T} \left|S_{T,t}\right| = O_p\Big(\log^{1/2} T + (TH)^{1/\theta} H^{\epsilon-1/2}\Big) \tag{2.29}$$

$$\max_{t=1,2,\cdots,T} \left|v_{T,t}^{(1)}\right| = O_p\bigg(\Big(\log T\Big)^{1/\alpha}\Big((\frac{H}{T})^{\gamma} + H^{-1/2}\log^{1/2} T\Big)\bigg) \tag{2.30}$$

$$\max_{t=1,2,\cdots,T} \left|v_{T,t}^{(2)}\right| = O_p\bigg(\Big(\log T\Big)^{1/\alpha}\Big((\frac{H}{T})^{\gamma} + H^{-1/2}\log^{1/2} T\Big)\bigg) \tag{2.31}$$

$$\max_{t=1,2,\cdots,T} \left|\Delta_{T,t}\right| = O_p\Big((H/T)^{\gamma}(\log T)^{1/\alpha}\Big). \tag{2.32}$$

Moreover, for $g_t \in \mathcal{E}(\alpha)$, *we have*

$$\max_{t=1,2,\cdots,T} |g_t| = O_p(\log^{1/\alpha} T). \tag{2.33}$$

Furthermore, if we assume that

$$E\left|\xi_t - \xi_j\right| \leqslant \frac{|j-t|}{T}, \quad 1 \leqslant t, j \leqslant T,$$

$\frac{1}{\sqrt{K_t}} S_{T,t}$ *obeys the weak law of large numbers (WLLN):*

$$\frac{1}{K_t} \sum_{j=1}^{T} b_{jt}\xi_j \xrightarrow{p} E(\xi_t). \tag{2.34}$$

Proof. (2.29), (2.32) and (2.33) are shown in Dendramis et al. [2020]. (2.29) is shown in (51), (2.33) is shown in (C.3) and (2.32) is shown in (70) of that paper. (2.30) is shown in (91) in Giraitis et al. [2020a]. We replace $(\frac{H}{T})^{1/2}$ with $(\frac{H}{T})^{\gamma}$ because we have a slightly different smoothness condition. We first show (2.31). Let us write:

$$\begin{aligned} v_{T,t}^{(2)} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\beta_j^2(\xi_j - E\xi_j) \\ &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\beta_j^2 - \beta_t^2)(\xi_j - E\xi_j) + \beta_t^2\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\xi_j - E\xi_j). \end{aligned}$$

Recall that $\beta_j^2 - \beta_t^2 = (\beta_j - \beta_t)(\beta_j + \beta_t) = O_p\Big((\frac{H}{T})^{\gamma}\Big)$, because by (2.4) $\big|\beta_j - \beta_t\big| = O_p\Big((\frac{H}{T})^{\gamma}\Big)$ and $|\beta_t| = O_p(1)$ for $j = 1, 2, ..., T$. In addition, we can always make ϵ in (2.29) small enough to have $(TH)^{1/\theta}H^{\epsilon-1} \leqslant H^{-\frac{1}{2}}$; then by (2.29), (2.32) and (2.33), we obtain:

$$\begin{aligned} \max_{t=1,2,\cdots,T}\big|v_{T,t}^{(2)}\big| &\leqslant \max_{t=1,2,\cdots,T}\left|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\beta_j^2 - \beta_t^2)(\xi_j - E\xi_j)\right| + \max_{t=1,2,\cdots,T}\big|\beta_t^2\big| \max_{t=1,2,\cdots,T}\left|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\xi_j - E\xi_j)\right| \\ &= O_p\bigg(\Big(\log T\Big)^{1/\alpha}\Big((\frac{H}{T})^{\gamma} + H^{-1/2}\log^{1/2} T\Big)\bigg). \end{aligned}$$

We now show (2.34). Write

$$\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\xi_j - E(\xi_t) = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\xi_j - \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}E(\xi_j) + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}E(\xi_j) - E(\xi_t)$$

By triangular inequality, we have

$$\left|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\xi_j - E(\xi_t)\right| \leqslant \left|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\xi_j - \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}E(\xi_j)\right| + \left|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}E(\xi_j) - E(\xi_t)\right|.$$

It is shown in (48) in Dendramis et al. [2020] that

$$\left|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\xi_j - \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}E(\xi_j)\right| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

In addition, we have

$$\left|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}E(\xi_j) - E(\xi_t)\right| \leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\, E\big|\xi_t - \xi_j\big| = O\Big(\frac{H}{T}\Big) = o(1),$$

which completes the proof. □

Appendix A.2: Useful lemmas

Lemma A2. *Consider model* (2.1) *and the following sums*

$$S_{xx,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}$$

$$\Delta_{x,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}(\beta_j - \beta_t)$$

$$S_{xu,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x_{ij} u_{ij}$$

$$S_{xe,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} e_{ij}$$

Then under Assumptions 2.1-2.4,

$$\left\|S_{xx,t}\right\|_{sp} = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{xx,t}\right\|_{sp} = O_p(1); \tag{2.35}$$

$$\left\|\Delta_{x,t}\right\| = O_p\Big((H/T)^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{x,t}\right\| = O_p\Big((H/T)^{\gamma_1}(\log T)^{1/\alpha}\Big); \tag{2.36}$$

$$\left\|S_{xu,t}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{xu,t}\right\| = O_p\Big(H^{-\frac{1}{2}}\log^{1/2} T\Big); \tag{2.37}$$

$$\left\|S_{xe,t}\right\| = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{xe,t}\right\| = O_p\Big(\log^{1/\alpha} T(1 + (H/T)^{\gamma_1})\Big). \tag{2.38}$$

Proof. *Proof of* (2.35). Notice that Assumption 2.4(i) implies that $\max_{t=1,2,\cdots,T}\left\|x_{it}x'_{it}\right\|_{sp} = O_p(1)$, $\forall i$. Therefore,

$$\begin{aligned}
\left\|S_{xx,t}\right\|_{sp} &\leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|x_{ij}x'_{ij}\right\|_{sp} \\
&\leqslant \max_{t=1,2,\cdots,T}\left\|x_{it}x'_{it}\right\|_{sp}\cdot\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} = O_p(1).
\end{aligned}$$

For the second part, write

$$\begin{aligned}
S_{xx,t} &= \frac{1}{H}\sum_{j=1}^{T} b_{jt}E(x_{ij}x'_{ij}) + \frac{1}{H}\sum_{j=1}^{T} b_{jt}(x_{ij}x'_{ij} - E(x_{ij}x'_{ij})) \\
&= S^{(1)}_{xx,it} + S^{(2)}_{xx,it} = S^{(1)}_{xx,it}(1 + \tilde{\Delta}_t),
\end{aligned} \tag{2.39}$$

where $\tilde{\Delta}_t = (S^{(1)}_{xx,t})^{-1}(S_{xx,t} - S^{(1)}_{xx,t})$. For $S^{(1)}_{xx,t}$, by Assumption 2.4(i),

$$\max_{t=1,2,\cdots,T}\left\|S^{(1)}_{xx,t}\right\|_{sp} \leqslant \max_{t=1,2,\cdots,T}\left\|\Sigma_{xx,t}\right\|_{sp} \cdot \frac{1}{K_t}\sum_{j=1}^{T} b_{jt} = O_p(1).$$

Next, consider

$$\max_{t=1,2,\cdots,T}\left\|S_{xx,t} - S^{(1)}_{xx,t}\right\|_{sp} = \max_{t=1,2,\cdots,T}\left\|\frac{1}{H}\sum_{j=1}^{T} b_{jt}(x_{ij}x'_{ij} - E(x_{ij}x'_{ij}))\right\|_{sp}.$$

By Assumption 2.1(i)–(ii), the (ℓ, k)-th component $\omega_{\ell k,ij} = x_{\ell,ij}x_{k,ij} - Ex_{\ell,ij}x_{k,ij}$ is strong-mixing. Moreover, let $\theta' = \theta/2$, $E\left|x_{\ell,ij}x_{k,ij}\right|^{\theta'} \leqslant C < \infty$. Together with the assumption of bandwidth in (2.8), we apply (2.29) to obtain

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{H}\sum_{j=1}^{T} b_{jt}(x_{ij}x'_{ij} - E(x_{ij}x'_{ij}))\right\|_{sp} = O_p(H^{-\frac{1}{2}}\log^{\frac{1}{2}} T + (TH)^{\frac{1}{\theta}}H^{\varepsilon-1})$$

for any $\varepsilon > 0$. We can always make ε small enough to have $(TH)^{\frac{1}{\theta}}H^{\varepsilon-1} \leqslant H^{-\frac{1}{2}}$. This then implies that

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{H}\sum_{j=1}^{T} b_{jt}(x_{ij}x'_{ij} - E(x_{ij}x'_{ij}))\right\|_{sp} = O_p(H^{-\frac{1}{2}}\log^{1/2} T) = o_p(1).$$

From (2.39), we have

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|S_{xx,t}\right\|_{sp} &\leqslant \max_{t=1,2,\cdots,T}\left\|S^{(1)}_{xx,t}\right\|_{sp}\max_{t=1,2,\cdots,T}\left\|I + \tilde{\Delta}_t\right\|_{sp}\\ &\leqslant \max_{t=1,2,\cdots,T}\left\|S^{(1)}_{xx,t}\right\|_{sp}(1 + \max_{t=1,2,\cdots,T}\left\|\tilde{\Delta}_t\right\|_{sp}) = O_p(1).\end{aligned}$$

Proof of (2.36). Notice that, (2.4) implies that $\left\|\beta_j - \beta_t\right\| = O_p\big((H/T)^{\gamma_1}\big)$. Then,

$$\left\|\Delta_{x,t}\right\| \leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|x_{ij}x'_{ij}(\beta_j - \beta_t)\right\| \leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|x_{ij}x'_{ij}\right\|_{sp}\left\|\beta_j - \beta_t\right\| = O_p\big((H/T)^{\gamma_1}\big).$$

For the second part, elements in $\Delta_{x,it}$ involve a finite number of linear combinations of sums

$$s_t := \frac{1}{H}\sum_{j=1}^{T} b_{jt}\omega_{\ell k,it}(\beta_{m,j} - \beta_{m,t}),$$

where $\omega_{\ell k,ij} = x_{\ell,ij}x_{k,ij} - Ex_{\ell,ij}x_{k,ij}$ is strong-mixing and $E\left|x_{\ell,ij}x_{k,ij}\right|^{\theta'} \leqslant C < \infty$. By Assumption 2.3, we can apply (2.32) to obtain $\max_{t=1,2,\cdots,T}|s_t| = O_p\big((H/T)^{\gamma_1}(\log T)^{1/\alpha}\big)$, which implies that

$$\max_{t=1,2,\cdots,T}\left\|\Delta_{x,t}\right\| = O_p\big((H/T)^{\gamma_1}(\log T)^{1/\alpha}\big).$$

Proof of (2.37). By the arguments used in Lemma A.4 in Giraitis et al. [2014] or Lemma 6.2 in Giraitis et al. [2018], we could show that

$$\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij} \xrightarrow{d} \mathcal{N},$$

with additional mixing or martingale difference type of assumptions on the process $(x_{it}u_{it})$, where $\mathcal{N}$ denotes Normal distribution. This implies that

$$\frac{K_t}{K_{2,t}^{1/2}} S_{xu,t} = O_p(1) \Longrightarrow S_{xu,t} = O_p\Big(\frac{1}{\sqrt{H}}\Big),$$

since $\frac{K_t}{K_{2,t}^{1/2}} = O(\sqrt{H})$. For the second part, elements in $S_{xu,t}$ involve finite a number of linear combinations of sums

$$\bar{s}_t := \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{\ell,ij}u_{ij}.$$

Because $Ex_{\ell,ij}u_{ij} = 0$, $(x_{\ell,ij}u_{ij})$ is strong-mixing with $E\left|x_{\ell,ij}u_{ij}\right|^{\theta'} \leqslant C$. We can then apply (2.29) to obtain $\max_{t=1,2,\cdots,T}|\bar{s}_t| = O_p(H^{-\frac{1}{2}}\log^{\frac{1}{2}} T + (TH)^{\frac{1}{\theta}}H^{\varepsilon-1}) = O_p(H^{-\frac{1}{2}}\log^{\frac{1}{2}} T)$, by the same reasoning used in the second part of (2.35), which implies that

$$\max_{t=1,2,\cdots,T}\left\|S_{xu,t}\right\| = O_p(H^{-\frac{1}{2}}\log^{\frac{1}{2}} T).$$

Proof of (2.38). Write

$$\begin{aligned} S_{xe,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij} \\ &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'(e_{ij}-e_{it}) + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{it} \\ &= S_{xe_1,t} + S_{xe_2,t}. \end{aligned} \tag{2.40}$$

From (2.4), same assumptions are imposed on (e_{it}) and (β_t). By similar arguments as in the proof of (2.36), we obtain

$$\left\|S_{xe_1,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big) = o_p(1)$$

and

$$\max_{t=1,2,\cdots,T}\left\|S_{xe_1,t}\right\| = O_p\Big((H/T)^{\gamma_1}(\log T)^{1/\alpha}\Big).$$

For $S_{xe_2,t}$, notice that

$$\left\|S_{xe_2,t}\right\| \leqslant \|e_{it}\| \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} \left\|x_{ij}x'_{ij}\right\| \leqslant \|e_{it}\| \max_{t=1,2,\cdots,T} \left\|x_{it}x'_{it}\right\|_{sp} \frac{1}{K_t} \sum_{j=1}^{T} b_{jt} = O_p(1),$$

since $\|e_{it}\| = O_p(1)$ by Assumption 2.3. Moreover, elements in $S_{xe_2,it}$ involve linear combinations of elements in e_{it} multiplying the sums $\frac{1}{H}\sum_{j=1}^{T} b_{jt}x_{\ell,ij}x_{k,ij}$, by (2.33), $\max_{t=1,2,\cdots,T}|e_{it}| = O_p\Big(\log^{1/\alpha} T\Big)$ and by (2.35), $\max_{t=1,2,\cdots,T}\Big|\frac{1}{H}\sum_{j=1}^{T} b_{jt}x_{\ell,ij}x_{k,ij}\Big| = O_p(1)$. Then, we obtain

$$\max_{t=1,2,\cdots,T}\left\|S_{xe_2,t}\right\| = O_p\Big(\log^{1/\alpha} T\Big).$$

So, by continuing from (2.40), we have

$$\begin{aligned}
&\left\|S_{xe,t}\right\| \leqslant \left\|S_{xe_1,t}\right\| + \left\|S_{xe_2,t}\right\| = O_p(1) \\
&\max_{t=1,2,\cdots,T}\left\|S_{xe,t}\right\| \leqslant \max_{t=1,2,\cdots,T}\left\|S_{xe_1,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xe_2,t}\right\| = O_p\Big(\log^{1/\alpha} T(1 + (H/T)^{\gamma_1})\Big).
\end{aligned}$$

□

Lemma A3. *Define*

$$r_{N,T,H,\gamma,\alpha} = (\frac{H}{T})^{\gamma} \log^{1/\alpha} T + \frac{\log^{1/\alpha} T}{\sqrt{N}} + \sqrt{\frac{\log T}{NH}}.$$

Consider model (2.11) *and the following sums*

$$\begin{aligned}
S_{z\psi,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j \\
\Delta_{z,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}(\beta_j - \beta_t) \\
S_{zu,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} u_{ij} \\
S_{ze,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij}
\end{aligned}$$

Then under Assumptions 2.1–2.3 (except Assumption 2.2(iii)), Assumptions 3.1-3.4, and $L = o\Big(L/\log^{\frac{\gamma_2}{\alpha}} T\Big)$,

$\log^{1/\alpha} T = o(N^{-1/2})$,

$$\left\|S_{z\psi,t}\right\|_{sp} = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{z\psi,t}\right\|_{sp} = O_p(1); \tag{2.41}$$

$$\left\|\Delta_{z,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{z,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\log^{2/\alpha} T\Big); \tag{2.42}$$

$$\left\|S_{zu,t}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| = O_p\bigg(\Big(\log T\Big)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\bigg); \tag{2.43}$$

$$\left\|S_{ze,t}\right\| = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| = O_p\bigg(\Big(\log T\Big)^{2/\alpha}\Big(1 + (\frac{H}{T})^{\min(\gamma_1,\gamma_2)}\Big)\bigg); \tag{2.44}$$

Proof. *Proof of* (2.41). Write

$$\begin{aligned}
&\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j \\
=& \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_j + \Psi'_j - \Psi'_t + \Psi'_t)z_{ij}z'_{ij}(\hat{\Psi}_j - \Psi_j + \Psi_j - \Psi_t + \Psi_t) \\
=& \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_j)z_{ij}z'_{ij}(\hat{\Psi}_j - \Psi_j) + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t)z_{ij}z'_{ij}(\hat{\Psi}_j - \Psi_j) + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_j)z_{ij}z'_{ij}(\Psi_j - \Psi_t) \\
&+ \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t)z_{ij}z'_{ij}(\Psi_j - \Psi_t) + \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t)z_{ij}z'_{ij}\Big)\Psi_t + \Psi'_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}z'_{ij}(\Psi_j - \Psi_t)\Big) \\
&+ \Psi'_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}z'_{ij}\Big)\Psi_t
\end{aligned}$$

From Lemma 1, $\max_{t=1,2,\cdots,T}\left\|\hat{\Psi}'_t - \Psi'_t\right\|_{sp} = o_p(1)$. Consider

$$\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_j)z_{ij}z'_{ij}(\hat{\Psi}_j - \Psi_j)\right\|_{sp} \leqslant \left(\max_{t=1,2,\cdots,T}\left\|\hat{\Psi}'_t - \Psi'_t\right\|_{sp}\right)^2 \frac{1}{K_t}\sum_{j=1}^{T} b_{jt} \max_{t=1,2,\cdots,T}\left\|z_{it}z'_{it}\right\|_{sp} = o_p(1).$$

Similarly, terms above involving $\hat{\Psi}'_j - \Psi'_j$ are also $o_P(1)$. Next, consider

$$\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t)z_{ij}z'_{ij}(\Psi_j - \Psi_t)\right\|_{sp} \leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|\Psi'_j - \Psi'_t\right\| \max_{t=1,2,\cdots,T}\left\|z_{it}z'_{it}\right\|_{sp} = O_p\Big((\frac{H}{T})^{\gamma_2}\Big) = o_p(1).$$

Similarly, terms above involving $\Psi'_j - \Psi'_t$ are also $o_P(1)$. Then,

$$\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j = \Psi'_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}z'_{ij}\Big)\Psi_t + o_p(1) = \Psi'_t\Sigma_{zz,i,t}\Psi_t + o_p(1)$$

which implies that

$$\left\|S_{z\psi,t}\right\|_{sp} = O_p(1),$$

since by Assumption 3.4(i), $\Psi_t'\Sigma_{zz,i,t}\Psi_t$ is positive definite. For the second part, since by Assumption 3.1(iii), $\max_{t=1,2,\cdots,T}\left\|\Sigma_{\Psi zz\Psi,it}\right\|_{sp} < \infty$, $\forall i$. Then, it follows from the expansion above and similar derivation of $S_{xx,t}^{(1)}$ in (2.35) that

$$\max_{t=1,2,\cdots,T}\left\|S_{z\psi,t}\right\|_{sp} = \max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\Psi_j' z_{ij} z_{ij}'\Psi_j\right\|_{sp} + o_p(1) = O_p(1).$$

Proof of (2.42). Write

$$\begin{aligned}
\left\|\Delta_{z,t}\right\| &\leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}_j' - \Psi_j')z_{ij}x_{ij}'(\beta_j - \beta_t)\right\| + \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\Psi_j' z_{ij}x_{ij}'(\beta_j - \beta_t)\right\| \\
&\leqslant \max_{t=1,2,\cdots,T}\left\|\hat{\Psi}_t' - \Psi_t'\right\|_{sp}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|z_{ij}x_{ij}'\right\|_{sp}\left\|\beta_j - \beta_t\right\| + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|\Psi_j'\right\|_{sp}\left\|z_{ij}x_{ij}'\right\|_{sp}\left\|\beta_j - \beta_t\right\| \\
&\leqslant \max_{t=1,2,\cdots,T}\left\|\hat{\Psi}_t' - \Psi_t'\right\|_{sp}\max_{t=1,2,\cdots,T}\left\|z_{it}x_{it}'\right\|_{sp}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|\beta_j - \beta_t\right\| + \max_{t=1,2,\cdots,T}\left\|z_{it}x_{it}'\right\|_{sp}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\left\|\Psi_j'\right\|_{sp}\left\|\beta_j - \beta_t\right\|
\end{aligned}$$

By (2.4), $\left\|\beta_j - \beta_t\right\| = O_p\left((\frac{H}{T})^{\gamma_1}\right)$. Together with Lemma 1, Assumptions 3.1(iii) and 3.3, we have

$$\left\|\frac{1}{H}\sum_{j=1}^{T} b_{jt}\hat{\Psi}_j' z_{ij}x_{ij}'(\beta_j - \beta_t)\right\| = O_p\left((\frac{H}{T})^{\gamma_1}\right).$$

Second part follows directly from (57) in Giraitis et al. [2020a]:

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{H}\sum_{j=1}^{T} b_{jt}\hat{\Psi}_j' z_{ij}x_{ij}'(\beta_j - \beta_t)\right\| = O_p\left((\frac{H}{T})^{\gamma_1}\log^{2/\alpha} T\right).$$

Proof of (2.43). (26) in Giraitis et al. [2020a] implies that

$$\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt}z_{ij}u_{ij} = O_p(1) \Longrightarrow \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}u_{ij} = O_p\left(\frac{1}{\sqrt{H}}\right),$$

since $\frac{K_t}{K_{2,t}^{1/2}} = O(\sqrt{H})$. Write

$$\begin{aligned}
S_{zu,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}\hat{\Psi}'_j z_{ij}u_{ij} \\
&= \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}(\hat{\Psi}'_j - \Psi'_j + \Psi'_j - \Psi'_t + \Psi'_t) z_{ij}u_{ij} \\
&= \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}(\hat{\Psi}'_j - \Psi'_j) z_{ij}u_{ij} + \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t}(\Psi'_j - \Psi'_t) z_{ij}u_{ij} + \Psi'_t \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t} z_{ij}u_{ij} \\
&= S_{zu,t,1} + S_{zu,t,2} + S_{zu,t,3}.
\end{aligned}$$

Since $S_{zu,t,1}$ involves $\hat{\Psi}'_j - \Psi'_j$, $\hat{\Psi}_t$ is uniformly consistent, $S_{zu,t,1}$ is asymptotically negligible. For $S_{zu,t,2}$, we have

$$\|S_{zu,t,2}\| \leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t} \|\Psi'_j - \Psi'_t\|_{sp} \|z_{ij}u_{ij}\| = O_p\Big((\frac{H}{T})^{\gamma_2}\Big) = o_p(1).$$

For $S_{zu,t,3}$, we have

$$\|S_{zu,t,3}\| \leqslant \|\Psi'_t\|_{sp} \left\| \frac{1}{K_t}\sum_{j=1}^{T} b_{j,t} z_{ij}u_{ij} \right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

It follows immediately from triangular inequality that

$$\|S_{zu,t}\| \leqslant \|S_{zu,t,1}\| + \|S_{zu,t,2}\| + \|S_{zu,t,3}\| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Second part follows similarly from (58) in Giraitis et al. [2020a]:

$$\max_{t=1,2,\cdots,T} \|S_{zu,t}\| = O_p\bigg(\Big(\log T\Big)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\bigg),$$

where L is the bandwidth parameter used to obtain $\hat{\Psi}_t$ and H is used to obtain $\hat{\beta}_t$. Notice that, there are two differences from (58) in Giraitis et al. [2020a]. First, both terms have possibly different smoothness parameters: γ_1, γ_2. Second, $r_{N,T,L,\gamma_2,\alpha}$ involves N, since first stage estimates are obtained from the panels.

Proof of (2.44). Write

$$\begin{aligned}
S_{ze,t} &= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_j + \Psi'_j - \Psi'_t + \Psi'_t) z_{ij} x'_{ij}(e_{ij} - e_{it} + e_{it}) \\
&= \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_j) z_{ij} x'_{ij}(e_{ij} - e_{it}) + \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_j) z_{ij} x'_{ij}\Big) e_{it} \\
&\quad + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t) z_{ij} x'_{ij}(e_{ij} - e_{it}) + \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t) z_{ij} x'_{ij}\Big) e_{it} \\
&\quad + \Psi'_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} z_{ij} x'_{ij}(e_{ij} - e_{it})\Big) + \Psi'_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} z_{ij} x'_{ij}\Big) e_{it}.
\end{aligned}$$

Consider the first term:

$$\begin{aligned}
\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\hat{\Psi}'_j - \Psi'_j) z_{ij} x'_{ij}(e_{ij} - e_{it})\Big\| &\leqslant \max_{t=1,2,\cdots,T}\big\|\hat{\Psi}'_t - \Psi'_t\big\|_{sp}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} z_{ij} x'_{ij}(e_{ij} - e_{it})\Big\| \\
&\leqslant \max_{t=1,2,\cdots,T}\big\|\hat{\Psi}'_t - \Psi'_t\big\|_{sp} \max_{t=1,2,\cdots,T}\big\|z_{it} x'_{it}\big\|_{sp} \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\big\|e_{ij} - e_{it}\big\| \\
&= o_p(1).
\end{aligned}$$

Thus, terms involving $\hat{\Psi}'_j - \Psi'_j$ are also $o_p(1)$. Consider the third term:

$$\begin{aligned}
\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t) z_{ij} x'_{ij}(e_{ij} - e_{it})\Big\| &\leqslant \max_{t=1,2,\cdots,T}\big\|z_{it} x'_{it}\big\|_{sp} \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\big\|\Psi'_j - \Psi'_t\big\|_{sp}\big\|e_{ij} - e_{it}\big\| \\
&= O_p\Big(\big(\frac{H}{T}\big)^{2\gamma_2}\Big) = o_p(1).
\end{aligned}$$

Thus, terms involving $\Psi'_j - \Psi'_t$ are also $o_p(1)$. The dominating term is the last one. Then, we have

$$\begin{aligned}
\big\|S_{ze,t}\big\| &= \Big\|\Psi'_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{j,t} z_{ij} x'_{ij}\Big) e_{it}\Big\| + o_p(1) \\
&= \big\|\Psi'_t\big\|_{sp}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{j,t} z_{ij} x'_{ij}\Big\|_{sp} \|e_{it}\| + o_p(1) \\
&= O_p(1).
\end{aligned}$$

To derive the uniform rate, it follows the similar reasoning as above, except the fact that we have to use

(2.32) and (2.33). Since by Assumption imposed in Lemma 1, $\hat{\Psi}_t$ is uniformly consistent, we have

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| \leqslant & \max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t)z_{ij}x'_{ij}(e_{ij}-e_{it})\right\| + \max_{t=1,2,\cdots,T}\left\|\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}(\Psi'_j - \Psi'_t)z_{ij}x'_{ij}\Big)e_{it}\right\| \\ & + \max_{t=1,2,\cdots,T}\left\|\Psi'_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}x'_{ij}(e_{ij}-e_{it})\Big)\right\| + \max_{t=1,2,\cdots,T}\left\|\Psi'_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}z_{ij}x'_{ij}\Big)e_{it}\right\| \\ = & O_p\bigg(\Big(\log T\Big)^{2/\alpha}\Big(1+(\frac{H}{T})^{\min(\gamma_1,\gamma_2)}\Big)\bigg).\end{aligned}$$

□

Appendix B: Mathematical proofs

Appendix B.1: Proof of Theorem 1

Under (2.1) and (2.3), the TVP-OLS-MG estimator defined in (2.6) can be written as:

$$\begin{aligned}\hat{\beta}_{MG,t} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{i,t} \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\Big)^{-1}\sum_{j=1}^{T} b_{jt}x_{ij}y_{ij}\Big] \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\Big)^{-1}\sum_{j=1}^{T} b_{jt}x_{ij}(x'_{ij}\beta_{ij}+u_{ij})\Big] \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big[\Big(\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\Big)^{-1}\sum_{j=1}^{T} b_{jt}x_{ij}(x'_{ij}(\beta_j-\beta_t+\beta_t+e_{ij})+u_{ij})\Big].\end{aligned}$$

Then, we have

$$\hat{\beta}_{MG,t}-\beta_t = \Delta_{x,it} + S_{xu,it} + S_{xe,it}, \tag{2.45}$$

where:

$$\begin{aligned}\Delta_{x,it} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}(\beta_j-\beta_t))\Big] \\ S_{xu,it} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij})\Big] \\ S_{xe,it} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij})^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}e_{ij})\Big].\end{aligned}$$

We will show that

$$\left\|\Delta_{x,it}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{x,it}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\log^{1/\alpha}T\Big); \tag{2.46}$$

$$\left\|S_{xu,it}\right\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{xu,it}\right\| = O_p\Big((NH)^{-1/2}\log^{1/2}T\Big); \tag{2.47}$$

$$\left\|S_{xe,it}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{xe,it}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\log^{1/\alpha}T(1+(\frac{H}{T})^{\gamma_1})\Big); \tag{2.48}$$

These together establish (1) and (2)(i) in Theorem 1.

Proof of (2.46). Notice that

$$\begin{aligned}
\left\|\Delta_{x,it}\right\| &\leqslant \frac{1}{N}\sum_{i=1}^{N}\left\|(\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}')^{-1}(\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'(\beta_j-\beta_t))\right\| \\
&\leqslant \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'(\beta_j-\beta_t)\right\| \\
&= O_p\Big((\frac{H}{T})^{\gamma_1}\Big).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\max_{t=1,2,\cdots,T}\left\|\Delta_{x,it}\right\| &\leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\right\|_{sp}^{-1}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'(\beta_j-\beta_t)\right\| \\
&= O_p\Big((\frac{H}{T})^{\gamma_1}\log^{1/\alpha}T\Big).
\end{aligned}$$

The final inequalities above all follow from (2.35) and (2.36) in Lemma A2, and the fact that all terms are *i.i.d.* across i (Assumption 2.2(i)).

Proof of (2.47). Let us define

$$Z_i^{xu} = (\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}')^{-1}(\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}u_{ij}).$$

By (2.35) and (2.37), we have

$$\left\|Z_i^{xu}\right\| \leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}u_{ij}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Then, we have

$$E\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i^{xu}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z_i^{xu}\right\|^2 = O\Big(\frac{1}{NH}\Big),$$

where the first equality follows from the fact that (Z_i^{xu}) is *i.i.d.* over i. This implies that

$$\left\|S_{xu,it}\right\| = O\Big(\frac{1}{\sqrt{NH}}\Big).$$

To derive the uniform rate, it follows a similar reasoning as above, except for the fact that we have to use the uniform rate of (2.35) and (2.37) in Lemma A2. This implies that

$$\max_{t=1,2,\cdots,T}\left\|S_{xu,it}\right\| = O_p\Big((NH)^{-1/2}\log^{1/2} T\Big)$$

Proof of (2.48). Let us define

$$Z_i^{xe} = (\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}')^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}).$$

By (2.35) and (2.38), we have

$$\left\|Z_i^{xe}\right\| \leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x_{ij}'e_{ij}\right\| = O_p(1).$$

Then, we have

$$E\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i^{xe}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z_i^{xe}\right\|^2 = O\Big(\frac{1}{N}\Big),$$

where the first line follows from the fact that (Z_i^{xe}) is *i.i.d.* over i. This implies that

$$\left\|S_{xe,it}\right\| = O\Big(\frac{1}{\sqrt{N}}\Big).$$

To derive the uniform rate, it follows a similar reasoning as above, except for the fact that we have to use the uniform rate of (2.35) and (2.38) in Lemma A2. This implies that

$$\max_{t=1,2,\cdots,T}\left\|S_{xe,it}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\log^{1/\alpha} T(1 + (\frac{H}{T})^{\gamma_1})\Big).$$

Now by combining (2.46), (2.47) and (2.48), we have

$$\begin{aligned}\sqrt{N}(\hat{\beta}_{MG,t}-\beta_t) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big[\Big(\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\Big)^{-1}\Big(\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'e_{ij}\Big)\Big]+O_p\Big(\frac{1}{\sqrt{H}}\Big)+O_p\Big(\sqrt{N}(\frac{H}{T})^{\gamma_1}\Big)\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}+o_p(1),\end{aligned}$$

since it is assumed that $(\frac{H}{T})^{\gamma_1}=o(N^{-1/2})$ as $(N,T)\to\infty$ and according to (2.38)

$$\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'e_{ij}=\Big(\frac{1}{K_t}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\Big)e_{it}+o_p(1).$$

By Assumption 2.2(i), we can apply CLT for *i.i.d.* sequences to obtain

$$\sqrt{N}\left(\Omega_{e,t}\right)^{-1/2}\left(\hat{\beta}_{MG,t}-\beta_t\right)\xrightarrow{d}N(0,I_k),$$

where

$$\Omega_{e,t}=\lim_{N\to\infty}\ \mathrm{Var}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}\Big),$$

and by Assumption 2.2(iv) it is positive definite.

We now show that $\Sigma_{MG,t}$ can be consistently estimated by

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}_{it}-\hat{\beta}_{MG,t})(\hat{\beta}_{it}-\hat{\beta}_{MG,t})'. \tag{2.49}$$

Because

$$\begin{aligned}\hat{\beta}_{it}-\beta_t &= (\frac{1}{H}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}')^{-1}(\frac{1}{H}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'e_{ij})+O_p\Big(\frac{1}{\sqrt{H}}\Big)+O_p\Big((\frac{H}{T})^{\gamma_1}\Big)\\ &= e_{it}+o_p(1)\end{aligned}$$

and

$$\hat{\beta}_{MG,t}-\beta_t=\frac{1}{N}\sum_{i=1}^{N}e_{it}+o_p(1)$$

we have

$$\hat{\beta}_{it}-\hat{\beta}_{MG,t}=e_{it}-\frac{1}{N}\sum_{i=1}^{N}e_{it}+o_p(1)$$

Then, as $(N,T)\to\infty$,

$$\frac{1}{N}\sum_{i=1}^{N}(\hat\beta_{i,t}-\hat\beta_{MG,t})(\hat\beta_{i,t}-\hat\beta_{MG,t})' \stackrel{p}{\longrightarrow} \Sigma_{MG,t},$$

which implies that (2.49) is a consistent estimator for $\Omega_{e,t}$.

By (2.45), we see that

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat\beta_{MG,t}-\beta_t\right\| &\leqslant \max_{t=1,2,\cdots,T}\left\|\Delta_{x,it}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xu,it}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xe,it}\right\| \\ &= O_p\Big((\frac{H}{T})^{\gamma_1}\log^{1/\alpha}T + \frac{\log^{1/\alpha}T}{\sqrt{N}} + \sqrt{\frac{\log T}{NH}}\Big),\end{aligned}$$

which completes the proof of Theorem 1(1) for $\hat\beta_{MG,t}$ and Theorem 1(2)(i).

We now explore the pooled estimator (2.7). Notice that

$$\begin{aligned}\hat\beta_{P,t}-\beta_t &= \Big(\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}y_{ij}-\beta_t \\ &= \Big(\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}(x_{ij}'\beta_{ij}+u_{ij})-\beta_t \\ &= \Big(\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}(x_{ij}'(\beta_j-\beta_t+\beta_t+e_{ij})+u_{ij})-\beta_t \\ &= \Big(\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}u_{ij} + \Big(\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'e_{ij} \\ &\quad + \Big(\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'\Big)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{T}b_{jt}x_{ij}x_{ij}'(\beta_j-\beta_t) \\ &= S_{xx,p,t}^{-1}(S_{xu,p,t}+S_{xe,p,t}+\Delta_{x,p,t}),\end{aligned} \tag{2.50}$$

where

$$S_{xx,p,t} = \frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}$$

$$\Delta_{x,p,t} = \frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} (\beta_j - \beta_t)$$

$$S_{xu,p,t} = \frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} u_{ij}$$

$$S_{xe,p,t} = \frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} e_{ij}$$

We will show that

$$\left\|S_{xx,p,t}\right\|_{sp} = O_p(1), \quad \max_{t=1,2,\cdots,T} \left\|S_{xx,p,t}\right\|_{sp} = O_p(1); \tag{2.51}$$

$$\left\|\Delta_{x,p,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T} \left\|\Delta_{x,p,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1} \log^{1/\alpha} T\Big); \tag{2.52}$$

$$\left\|S_{xu,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big), \quad \max_{t=1,2,\cdots,T} \left\|S_{xu,p,t}\right\| = O_p\Big((NH)^{-1/2} \log^{1/2} T\Big); \tag{2.53}$$

$$\left\|S_{xe,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big), \quad \max_{t=1,2,\cdots,T} \left\|S_{xe,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}} \log^{1/\alpha} T(1 + (\frac{H}{T})^{\gamma_1})\Big); \tag{2.54}$$

These together establish (1) and (2)(ii) in Theorem 1.

Proof of (2.51). This follows immediately from (2.35), since by Assumption 2.2(i), all are *i.i.d*. across *i*:

$$\left\|\frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}\right\|_{sp} \leqslant \frac{1}{N} \left\|\frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}\right\|_{sp} = O_p(1)$$

$$\max_{t=1,2,\cdots,T} \left\|\frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}\right\|_{sp} \leqslant \frac{1}{N} \max_{t=1,2,\cdots,T} \left\|\frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}\right\|_{sp} = O_p(1)$$

Proof of (2.52). By Assumption 2.2(i) and (2.36), we have

$$\left\|\frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} (\beta_j - \beta_t)\right\| \leqslant \frac{1}{N} \sum_{i=1}^{N} \left\|\frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} (\beta_j - \beta_t)\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big)$$

$$\max_{t=1,2,\cdots,T} \left\|\frac{1}{NK_t} \sum_{i=1}^{N} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} (\beta_j - \beta_t)\right\| \leqslant \frac{1}{N} \sum_{i=1}^{N} \max_{t=1,2,\cdots,T} \left\|\frac{1}{K_t} \sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} (\beta_j - \beta_t)\right\| = O_p\Big((\frac{H}{T})^{\gamma_1} \log^{1/\alpha} T\Big)$$

Proof of (2.53). Define

$$Z^{xu}_{i,p,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}u_{ij}.$$

By Assumption 2.2(i) and (2.37), we have

$$E\left\|S_{xu,p,t}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z^{xu}_{i,p,t}\right\|^2 = O\Big(\frac{1}{NH}\Big).$$

This establishes the first part of (2.53). To derive uniform rate, it follows same reasoning as above, except that we need to use uniform rate in (2.37). Then, we have

$$\max_{t=1,2,\cdots,T}\left\|S_{xu,p,t}\right\| = O_p\Big((NH)^{-1/2}\log^{1/2} T\Big).$$

Proof of (2.54). Define

$$Z^{xe}_{i,p,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}e_{ij}.$$

By Assumption 2.2(i) and (2.38), we have

$$E\left\|S_{xe,p,t}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z^{xe}_{i,p,t}\right\|^2 = O\Big(\frac{1}{N}\Big).$$

This establishes the first part of (2.54). To derive uniform rate, it follows same reasoning as above, except that we need to use uniform rate in (2.38). Then, we have

$$\max_{t=1,2,\cdots,T}\left\|S_{xu,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\log^{1/\alpha} T(1 + (\frac{H}{T})^{\gamma_1})\Big).$$

Now by combining (2.51), (2.52), (2.53), and (2.54), we have

$$\begin{aligned}\sqrt{N}(\hat{\beta}_{P,t} - \beta_t) &= \Big(\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\Big)^{-1}\Big(\frac{1}{\sqrt{N}K_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}e_{ij}\Big) + O_p\Big(\frac{1}{\sqrt{H}}\Big) + O_p\Big(\sqrt{N}(\frac{H}{T})^{\gamma_1}\Big)\\ &= \left(\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\Big)e_{it}\right) + o_p(1)\end{aligned}$$

Because it is assumed that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ as $(N,T) \to \infty$, the last two terms in the first line are $o_p(1)$ and the dominating term is the first one. As in the case of Mean group estimator, we can apply CLT for *i.i.d.* sequences to obtain

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}x_{ij}x'_{ij}\Big)e_{it} \xrightarrow{d} N(0, R_t),$$

where

$$R_t = \lim_{N\to\infty} \text{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Sigma_{xx,it} e_{it}\right),$$

since by (2.34), for each i, we have

$$\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x'_{ij} x_{ij} \xrightarrow{p} \Sigma_{xx,it}.$$

Then, since

$$\lim_{(N,T)\to\infty} \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij} = \overline{\Sigma}_{xx,t} = O(1),$$

it follows from Slutsky's theorem that

$$\sqrt{N}\,\overline{\Sigma}_{xx,t}\; R_t^{-1/2}\left(\hat{\beta}_{P,t} - \beta_t\right) \xrightarrow{d} N(0, I_k).$$

Consider

$$\hat{R}_t = \frac{1}{N}\sum_{i=1}^{N}\left[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}\Big)(\hat{\beta}_{i,t} - \hat{\beta}_{MG,t})(\hat{\beta}_{i,t} - \hat{\beta}_{MG,t})'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}\Big)\right].$$

We now show that $\hat{R}_t$ is a consistent estimator of R_t. Write

$$\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}(\hat{\beta}_{i,t} - \hat{\beta}_{MG,t}) = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt} x_{ij} x'_{ij}\Big(e_{it} - \frac{1}{N}\sum_{i=1}^{N} e_{it}\Big)$$

Then, as $(N, T) \to \infty$,

$$\hat{R}_t \xrightarrow{p} R_t$$

which establishes the claim.

By (2.50), together with (2.51), (2.52), (2.53), and (2.54), we have

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat{\beta}_{P,t} - \beta_t\right\| &\leqslant \max_{t=1,2,\cdots,T}\left\|S_{xx,p,t}\right\|_{sp}^{-1}\left(\max_{t=1,2,\cdots,T}\left\|\Delta_{x,p,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xu,p,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xe,p,t}\right\|\right)\\ &= O_p\Big((\frac{H}{T})^{\gamma_1}\log^{1/\alpha} T + \frac{\log^{1/\alpha} T}{\sqrt{N}} + \sqrt{\frac{\log T}{NH}}\Big).\end{aligned}$$

Appendix B.2: Proof of Theorem 2

Let us first consider the TVP-IV-MG estimator,

$$\begin{aligned}\hat{\beta}^{IV}_{MG,t} &= \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{IV}_{i,t}\\ &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j)^{-1}(\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} y_{ij})\Big].\end{aligned}$$

A similar expansion to the TVP-OLS-MG case gives

$$\hat{\beta}^{IV}_{MG,t} - \beta_t = \Delta_{z,t} + S_{zu,t} + S_{ze,t},$$

where

$$\begin{aligned}\Delta_{z,t} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j)^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}(\beta_j - \beta_t))\Big]\\ S_{zu,t} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j)^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} u_{ij})\Big]\\ S_{ze,t} &= \frac{1}{N}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j)^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij})\Big].\end{aligned}$$

We will show that

$$\|\Delta_{z,t}\| = O_p\big((\frac{H}{T})^{\gamma_1}\big), \quad \max_{t=1,2,\cdots,T}\|\Delta_{z,t}\| = O_p\big((\frac{H}{T})^{\gamma_1}\log^{2/\alpha} T\big) \tag{2.55}$$

$$\|S_{zu,t}\| = O_p\big(\frac{1}{\sqrt{NH}}\big), \quad \max_{t=1,2,\cdots,T}\|S_{zu,t}\| = O_p\Big(\frac{1}{\sqrt{N}}\big((\log T)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\big)\Big) \tag{2.56}$$

$$\|S_{ze,t}\| = O_p\big(\frac{1}{\sqrt{N}}\big), \quad \max_{t=1,2,\cdots,T}\|S_{ze,t}\| = O_p\Big(\frac{(\log T)^{2/\alpha}}{\sqrt{N}}\big(1 + (\frac{H}{T})^{\min(\gamma_1,\gamma_2)}\big)\Big). \tag{2.57}$$

These together establishes Theorem 2(i) and Theorem 2(ii).

Proof of (2.55). Observe that

$$\begin{aligned}\|\Delta_{z,t}\| &\leqslant \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big\|^{-1}_{sp}\Big\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}(\beta_j - \beta_t)\Big\|\\ &= O_p\big((\frac{H}{T})^{\gamma_1}\big),\end{aligned}$$

where the second line follows from (2.41) and (2.42) and the fact that by Assumption 3.2(i) all terms are *i.i.d.* accross i. Again, by uniform rates derived in (2.41) and (2.42), we have

$$\max_{t=1,2,\cdots,T}\left\|\Delta_{z,t}\right\| \leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j\right\|_{sp}^{-1}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}(\beta_j-\beta_t)\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\log^{2/\alpha}T\Big).$$

Proof of (2.56). Let

$$Z_i^{zu} = (\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j)^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}u_{ij}).$$

Clearly, by (2.41) and (2.43):

$$\left\|Z_i^{zu}\right\| \leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}u_{ij}\right\| = O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Thus, we have

$$E\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i^{zu}\right\|^2 = \frac{1}{HN^2}\sum_{i=1}^{N} HE\left\|Z_i^{zu}\right\|^2 = O\Big(\frac{1}{NH}\Big),$$

which implies that

$$\left\|S_{zu,t}\right\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big).$$

To derive the uniform rate, it follows similar reasoning as above, except for the fact that we have to use uniform rates derived in (2.41) and (2.43) to obtain

$$\max_{t=1,2,\cdots,T}\left\|S_{zu,t}\right\| = O_p\bigg(\frac{1}{\sqrt{N}}\Big((\log T)^{1/\alpha}r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\Big)\bigg).$$

Proof of (2.57). Let

$$Z_i^{ze} = (\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j)^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}e_{ij}).$$

Clearly, by (2.41) and (2.44):

$$\left\|Z_i^{ze}\right\| \leqslant \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}z'_{ij}\hat{\Psi}_j\right\|_{sp}^{-1}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij}x'_{ij}e_{ij}\right\| = O_p(1).$$

Thus, we have

$$E\left\|\frac{1}{N}\sum_{i=1}^{N} Z_i^{ze}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z_i^{ze}\right\|^2 = O\Big(\frac{1}{N}\Big),$$

which implies that

$$\left\|S_{ze,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big).$$

To derive the uniform rate, it follows again similar reasoning as above, except for the fact that we have to use uniform rates derived in (2.41) and (2.44) to obtain

$$\max_{t=1,2,\cdots,T}\left\|S_{ze,t}\right\| = O_p\bigg(\frac{(\log T)^{2/\alpha}}{\sqrt{N}}\Big(1+(\frac{H}{T})^{\min(\gamma_1,\gamma_2)}\Big)\bigg).$$

Now, we sum them up. By (2.55), (2.56) and (2.57), we first obtain the uniform consistency rate

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat{\beta}^{IV}_{MG,t}-\beta_t\right\| &\leqslant \max_{t=1,2,\cdots,T}\|\Delta_{xt}\| + \max_{t=1,2,\cdots,T}\left\|S_{xu,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{xe,t}\right\| \\ &= O_p\bigg(\frac{1}{\sqrt{N}}\Big((\log T)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\Big)\bigg).\end{aligned}$$

Then, we obtain the expansion of the estimator

$$\sqrt{N}(\hat{\beta}^{IV}_{MG,t}-\beta_t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big[(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j)^{-1}(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij})\Big] + O_p\Big(\sqrt{N}(\frac{H}{T})^{\gamma_1}\Big) + O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Because we assume that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ as $(N,T)\to\infty$, the last two terms above are $o_p(1)$ and the dominating term is the first one. Recall from the derivation of (2.41) and (2.44), together with (2.34), we have that

$$\begin{aligned}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j &= \Psi'_t\Sigma_{zz,it}\Psi_t + o_p(1)\\ \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij} &= \Psi'_t\Sigma_{zx,it} e_{it} + o_p(1).\end{aligned}$$

If $k = p$ and Ψ_t is invertible, we have

$$\sqrt{N}\,\Psi_t\,(\hat{\beta}^{IV}_{MG,t}-\beta_t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma^{-1}_{zz,it}\Sigma_{ze,it} e_{it} + o_p(1).$$

By Assumptions 3.4(i) and 3.4(ii), $\Sigma^{-1}_{zz,it}$ and $\Sigma_{ze,it}$ are all positive definite. Since (e_{it}) is *i.i.d.* across i, we can apply CLT for *i.i.d.* sequence to obtain

$$\sqrt{N}\,\Psi_t\,(\Omega^{IV}_{e,t})^{-1/2}\,(\hat{\beta}^{IV}_{MG,t}-\beta_t) \xrightarrow{d} N(0, I_k),$$

where $\Omega^{IV}_{e,t}$ is given by

$$\Omega^{IV}_{e,t} = \lim_{N\to\infty} \mathrm{Var}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma^{-1}_{zz,it}\Sigma_{ze,it}e_{it}\Big).$$

Consider

$$\hat{\beta}^{IV}_{it} - \beta_t = \Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)^{-1}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij}\Big) + o_p(1).$$

Similar expansion as in TVP-OLS case gives

$$\Psi_t\,(\hat{\beta}^{IV}_{it} - \beta_t) = \Sigma^{-1}_{zz,it}\Sigma_{ze,it}e_{it} + o_p(1).$$

Then, it follows similarly as in the proof of TVP-OLS-MG case that

$$\Psi_t\Big(\sum_{i=1}^{N}(\hat{\beta}^{IV}_{it} - \beta_t)(\hat{\beta}^{IV}_{it} - \beta_t)'\Big)\Psi'_t$$

is a consistent estimator for $\Omega^{IV}_{MG,t}$, by replacing Ψ_t with its estimated counterpart.

In the next step, we consider the pooled estimator

$$\hat{\beta}^{IV}_{P,t} = \Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)^{-1}\Big(\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} y_{i,j}\Big).$$

Similarly to the TVP-OLS-P case, we have

$$\hat{\beta}^{IV}_{P,t} - \beta_t = S^{-1}_{z\Psi,p,t}\Big(\Delta_{z,p,t} + S_{zu,p,t} + S_{ze,p,t}\Big),$$

where

$$S_{z\Psi,p,t} = \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j$$

$$\Delta_{z,p,t} = \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_j - \beta_t)$$

$$S_{zu,p,t} = \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} u_{i,j}$$

$$S_{ze,p,t} = \frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j} e_{i,j}.$$

We will show that

$$\left\|S_{z\Psi,p,t}\right\|_{sp} = O_p(1), \quad \max_{t=1,2,\cdots,T}\left\|S_{z\Psi,p,t}\right\|_{sp} = O_p(1) \tag{2.58}$$

$$\left\|\Delta_{z,p,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big), \quad \max_{t=1,2,\cdots,T}\left\|\Delta_{z,p,t}\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\log^{2/\alpha}T\Big) \tag{2.59}$$

$$\left\|S_{zu,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{NH}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{zu,p,t}\right\| = O_p\bigg(\frac{1}{\sqrt{N}}\Big(\big(\log T\big)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\Big)\bigg) \tag{2.60}$$

$$\left\|S_{ze,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big), \quad \max_{t=1,2,\cdots,T}\left\|S_{ze,p,t}\right\| = O_p\bigg(\frac{\big(\log T\big)^{2/\alpha}}{\sqrt{N}}\Big(1 + (\frac{H}{T})^{\min(\gamma_1,\gamma_2)}\Big)\bigg). \tag{2.61}$$

These together establishes uniform consistency and asymptotic normality for the pooled estimator.

Proof of (2.58). This follows immediately from (2.41) and Assumption 3.2(i):

$$\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j\right\|_{sp} \leqslant \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j\right\|_{sp} = O_p(1)$$

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j\right\|_{sp} \leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} z'_{i,j}\hat{\Psi}_j\right\|_{sp} = O_p(1)$$

Proof of (2.59). By Assumption 3.2(1) and (2.42), we have

$$\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_j - \beta_t)\right\| \leqslant \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_j - \beta_t)\right\| = O_p\Big((\frac{H}{T})^{\gamma_1}\Big)$$

$$\max_{t=1,2,\cdots,T}\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_j - \beta_t)\right\| \leqslant \frac{1}{N}\sum_{i=1}^{N}\max_{t=1,2,\cdots,T}\left\|\frac{1}{NK_t}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j}(\beta_j - \beta_t)\right\|$$
$$= O_p\Big((\frac{H}{T})^{\gamma_1}\log^{2/\alpha}T\Big).$$

Proof of (2.60). Define

$$Z^{zu}_{i,p,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} u_{i,j}.$$

By Assumption 3.2(i) and (2.43), we have

$$E\left\|S_{zu,p,t}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z^{zu}_{i,p,t}\right\|^2 = O\Big(\frac{1}{NH}\Big),$$

which establishes the first part of (2.60). To derive uniform rate, it follows a similar reasoning as above, except that we need to use uniform rate in (2.43). Then, we have

$$\max_{t=1,2,\cdots,T}\left\|S_{zu,p,t}\right\| = O_p\Big(\frac{1}{\sqrt{N}}\Big((\log T)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\Big)\Big).$$

Proof of (2.61). Define

$$Z^{ze}_{i,p,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{i,j} x'_{i,j} e_{ij}.$$

By Assumption 3.2(1) and (2.44), we have

$$E\left\|S_{ze,p,t}\right\|^2 = \frac{1}{N^2}\sum_{i=1}^{N} E\left\|Z^{ze}_{i,p,t}\right\|^2 = O\Big(\frac{1}{N}\Big),$$

which establishes the first part of (2.61). To derive uniform rate, it follows a similar reasoning as above, except that we need to use uniform rate in (2.44). Then, we have

$$\max_{t=1,2,\cdots,T}\left\|S_{ze,p,t}\right\| = O_p\Big(\frac{(\log T)^{2/\alpha}}{\sqrt{N}}\Big(1 + (\frac{H}{T})^{\min(\gamma_1,\gamma_2)}\Big)\Big).$$

Now we combine the equations. First, by (2.58), (2.59), (2.60), and (2.61), we have

$$\begin{aligned}\max_{t=1,2,\cdots,T}\left\|\hat{\beta}^{IV}_{P,t} - \beta_t\right\| &\max_{t=1,2,\cdots,T}\left\|S_{z\Psi,p,t}\right\|^{-1}_{sp}\Big(\max_{t=1,2,\cdots,T}\left\|\Delta_{z,p,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{zu,p,t}\right\| + \max_{t=1,2,\cdots,T}\left\|S_{ze,p,t}\right\|\Big)\\ &= O_p\Big(\frac{1}{\sqrt{N}}\Big((\log T)^{1/\alpha} r_{1,T,H,\gamma_1,\alpha} + r_{N,T,L,\gamma_2,\alpha}\Big)\Big).\end{aligned}$$

Then, we obtain the following expansion for the pooled estimator

$$\sqrt{N}(\hat{\beta}^{IV}_{P,t} - \beta_t) = \Big(\frac{1}{NH}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j\Big)^{-1}\Big(\frac{1}{\sqrt{N}H}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij}\Big) + O_p\Big(\sqrt{N}(\frac{H}{T})^{\gamma_1}\Big) + O_p\Big(\frac{1}{\sqrt{H}}\Big).$$

Because it is assumed that $(\frac{H}{T})^{\gamma_1} = o(N^{-1/2})$ as $(N,T)\to\infty$, the last two terms are $o_p(1)$ and the dominating term is the first one. Recall from the derivations of (2.41) and (2.44), together with (2.34) and the fact that $\overline{\Sigma}_{\Psi zz\Psi,t} = \text{plim}_{(N,T)\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\Psi'_j z_{ij} z'_{ij}\Psi_j = O_p(1)$, we have that

$$\frac{1}{NH}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} z'_{ij}\hat{\Psi}_j = \overline{\Sigma}_{\Psi zz\Psi,t} + o_p(1)$$

$$\frac{1}{\sqrt{N}H}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij} e_{ij} = \Psi_t\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{zx,it} e_{it} + o_p(1).$$

Since both $\overline{\Sigma}_{\Psi zz\Psi,t}$ and $\Sigma_{zx,it}$ are positive definite and the fact that (e_{it}) are $i.i.d.$ across i, we apply CLT for $i.i.d.$ sequence to obtain:

$$\frac{1}{\sqrt{N}H}\sum_{i=1}^{N}\sum_{j=1}^{T} b_{jt}\Sigma_{zx,it}e_{it} \xrightarrow{d} N\Big(0, R_{P,t}^{IV}\Big),$$

where $R_{P,t}^{IV}$ is given by

$$R_t^{IV} = \lim_{N\to\infty} \text{Var}\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Sigma_{zx,it}e_{it}\Big).$$

By Slutsky's theorem, we have

$$\sqrt{N}\,\overline{\Sigma}_{\Psi zz\Psi,t}\,\Psi_t^{-1}\left(R_{P,t}^{IV}\right)^{-1/2}\Big(\hat{\beta}_{P,t}^{IV} - \beta_t\Big) \xrightarrow{d} N(0, I_k).$$

Consider

$$\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}\Big)\Big(\hat{\beta}_{i,t}^{IV} - \hat{\beta}_{MG,t}^{IV}\Big) = \Psi_t\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt} z_{ij} x'_{ij}\Big)(e_{it} - \frac{1}{N}\sum_{i=1}^{N} e_{it}) + o_p(1).$$

Then, we have

$$\frac{1}{N}\sum_{i=1}^{N}\left[\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}\Big)\Big(\hat{\beta}_{i,t}^{IV} - \hat{\beta}_{MG,t}^{IV}\Big)\Big(\hat{\beta}_{i,t}^{IV} - \hat{\beta}_{MG,t}^{IV}\Big)'\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt}\hat{\Psi}'_j z_{ij} x'_{ij}\Big)'\right] = \Psi_t R_{p,t}^{IV}\Psi_t + o_p(1),$$

which shows that above is a consistent estimator for $\Psi_t R_{p,t}^{IV}\Psi_t$.

Appendix C: Additional empirical results

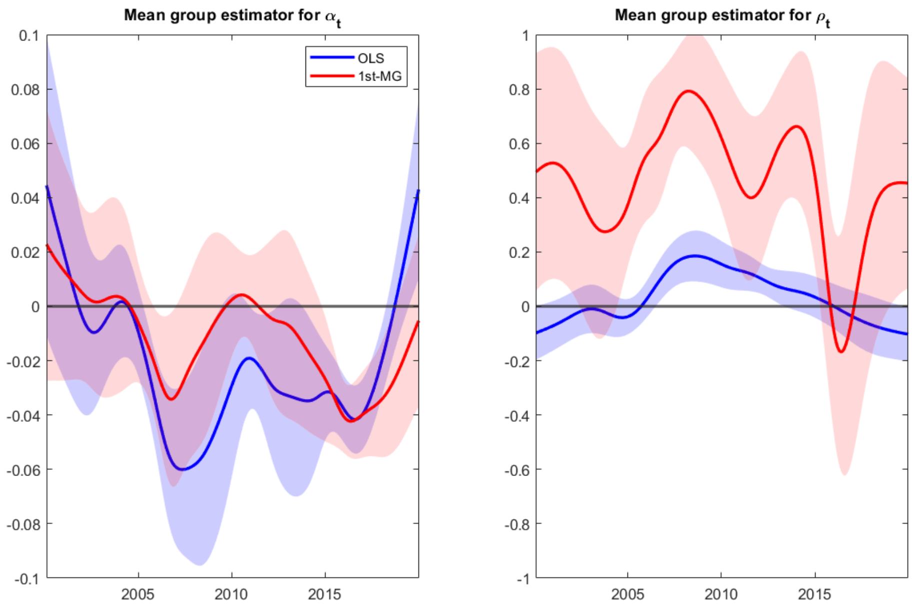

Figure 2.1: Empirical results for backward-looking Phillips curve. The solid blue lines show the point TVP-OLS estimates and the blue shaded areas show the 95% pointwise confidence intervals. The red solid lines show the point TVP-IV estimates and the red shaded areas show the 95% pointwise confidence intervals.

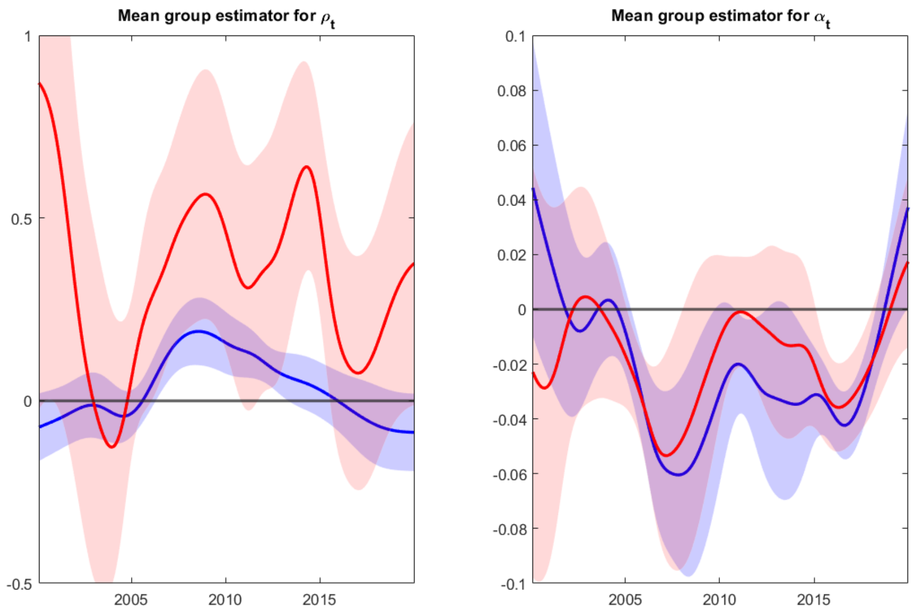

Figure 2.2: Empirical results for forward-looking Phillips curve. The solid blue lines show the point TVP-OLS estimates and the blue shaded areas show the 95% pointwise confidence intervals. The red solid lines show the point TVP-IV estimates and the red shaded areas show the 95% pointwise confidence intervals.

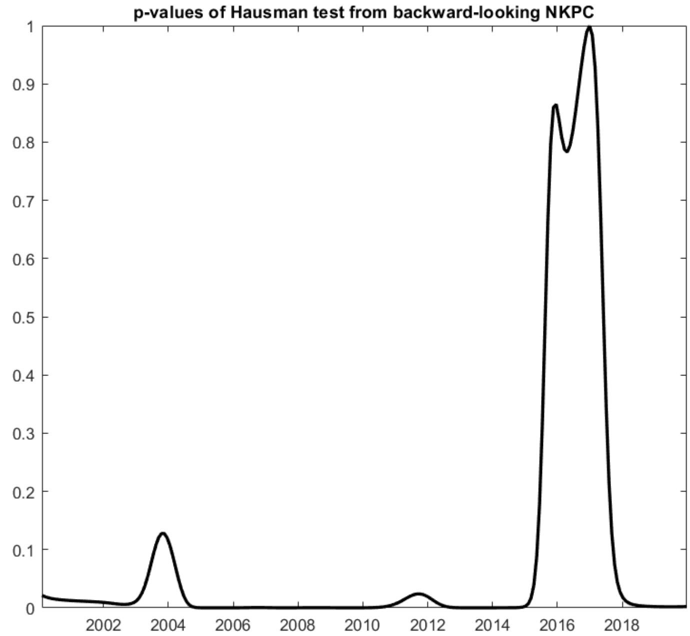

Figure 2.3: *p*-values of time-varying Hausman test for backward-looking Phillips curve.

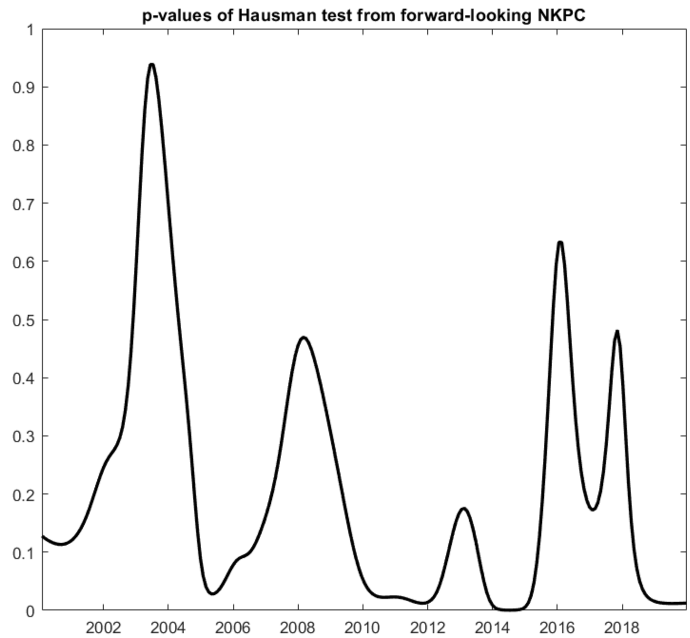

Figure 2.4: *p*-values of time-varying Hausman test for forward-looking Phillips curve.

Chapter 3

Time-Varying GMM Estimation

3.1 Introduction

Since the seminal work by Hansen [1982], the Generalized Method of Moments (GMM) has been widely used to analyze economic and finance data, see Hall et al. [2005], Hansen and West [2002] and Jagannathan et al. [2002] for an early review of the applications. Unlike likelihood-based inference, GMM only requires to specify a vector of moment conditions which often comes from economic and finance theory, without the need to specify the joint probability distribution of the data.

As parameter instability is pervasive (Stock and Watson [1996]), attempts have been made to handle structural change in the GMM framework. Most of the existing literature focuses on testing hypotheses about structural stability where under the alternative hypothesis, parameters are assumed to have break-points, possibly at unknown dates. However, as Hansen [2001] points out, "it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect." Indeed, all leading driving forces of structural change, such as technological improvements, climate change, and modifications in the institutional context, may take time to manifest their effects and thus change the parameters of econometric relationships. Moreover, the effects of the structural changes are hardly deterministic and much more likely stochastic due to the complex interactions in the economic system, so that the parameter changes in the econometric models are also better modeled as stochastic.

In this paper, I develop time-varying continuously updated GMM (TV CU-GMM) estimation and inferential theory for models whose parameters change smoothly and stochastically over time. The method extends the previous literature in various directions:

1. It extends the original CU-GMM estimator, which is first introduced in Hansen et al. [1996], to the setting in which there is time variation in model parameters.

2. The literature on nonparametric modeling of stochastic time-varying parameters is mostly on linear models in which estimators are available in closed form. Giraitis et al. [2016] develop quasi-maximum likelihood estimators which can handle nonlinear models, but the theory is restricted to the static Tobit

model. The estimator developed in this paper is fairly general to handle a broader class of models. In addition, the existing literature generally relies on martingale difference (MDS) or strictly stationarity assumptions to establish asymptotic normality for the estimator. These assumptions can be restrictive since many economic and financial series exhibit general unknown dependence and conditional heteroskedasticity patterns. The asymptotic theory developed in this paper is developed under weak dependence assumption of the data. Stationarity assumptions are also not needed. The assumptions required for consistency and asymptotic normality are fairly primitive. Standard α-mixing conditions are required with polynomial decay rate for mixing coefficients, but stationarity is not needed. The only required assumptions on the true process of time-varying parameters are smoothness and persistence. Moment conditions are also allowed to change over time, with a smoothness condition for time-varying moment functions.

3. I extend the Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix estimation in the constant coefficient setting to models with time-varying parameters. I prove that the time-varying HAC estimator is consistent. The bandwidth required is similar to the constant coefficient setting, except it is now of smaller order than quantity related to the bandwidth parameter used to construct sample average of moment functions, rather than the total number of time series observations.

4. I propose two new test statistics for the null of structural stability in this context. As in the structural break context, it can be shown that the null of structural stability can be decomposed into stability for identifying and overidentifying restrictions. The tests developed for the null of stability for identifying restrictions in the structural break context can be easily extended, but the test for stability for overidentifying restrictions is fairly new to the literature. I also derive the asymptotic distributions for these test statistics.

The finite-sample performance of the estimators and structural stability tests is evaluated by an extensive Monte-Carlo study. For the estimators, I assess the biases of point estimates and coverage probabilities of the true parameters. I also compute both the size and power of the proposed tests. Using a linear IV model with time-varying parameters as data generating process, I find that the time-varying CU-GMM estimates have satisfactory finite sample performance. They deliver higher coverage probabilities but since optimization is required, they are slightly less accurate than the linear time-varying IV estimates proposed in Giraitis et al. [2020b]. Both tests for structural stability are slightly oversized, but they have reasonable power. The strategy to identify the source of instability proposed by Hall and Sen [1999] in structural break context remains valid here. The Monte-Carlo results also provide some guidance on the choice of bandwidth. Overall, the choice of the bandwidth parameter for constructing sample averages of moment functions is far more important than the choice of the bandwidth parameter for the time-varying HAC covariance matrix. Setting the bandwidth parameter (for constructing sample averages of moment functions) equal to the square root of the number of observations leads to best performance in terms of bias and coverage probability considerations and the size and power of structural stability tests.

Finally, to illustrate in practice the use of time-varying CU-GMM estimator and structural stability tests, I provide an empirical application on dynamic asset pricing models with Stochastic Discount Factor (SDF) representation. I focus on the problem of joint pricing on the cross-section of equity and treasury portfolio returns. Using the 3-factor specification as in Adrian et al. [2015], I find that allowing time variation in conditional asset pricing models leads to substantial improvement in pricing performance. Time-varying CU-GMM estimation has the overall best pricing performance. Misspecification of price of risk dynamics may lead to large pricing errors. There is strong evidence of time variation in the price of risk estimates. The null of structural stability is also strongly rejected.

The paper is organized as follows. The rest of this section discusses related literature. Section 2 describes the framework and the time-varying CU-GMM estimators and presents the main theoretical results. Section 3 introduces two tests for structural stability and derives the limiting distributions of these tests. Section 4 evaluates the proposed estimators and tests in an extensive Monte Carlo study. Section 5 presents an empirical application on conditional asset pricing models with SDF representation. Section 6 summarizes the main results and concludes the paper. The proofs of all results are presented in the appendices.

NOTATION: The letter C stands for a generic finite positive constant, $\|A\|_{sp} = \sqrt{\lambda_{\max}(A'A)}$ is the spectrum norm of matrix A, where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of $\cdot$. $\|\cdot\|_p$ denotes the L^p norm, $\|\cdot\|$ is the Euclidean norm. $|\cdot|_p$ and $|\cdot|$ denote the associated norm when $\cdot$ is one dimensional. $a_n = O(b_n)$ states that the deterministic sequence $\{a_n\}$ is at most of order b_n. $a_n = o(b_n)$ states that the deterministic sequence $\{a_n\}$ is of smaller order than b_n. $x_n = O_p(y_n)$ states that the vector of random variables x_n is at most of order y_n in probability, and $x_n = o_p(y_n)$ is of smaller order than y_n in probability. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution.

Related literature

This paper naturally builds on research on continuously updated GMM estimation, which is first introduced in Hansen et al. [1996]. The estimator has exactly the same asymptotic distribution as two-step GMM but has better finite sample properties. Donald and Newey [2000] provide a Jackknife interpretation on why CU-GMM estimator performs better than two-step GMM. Smith [2011] provides a general treatment of generalized empirical likelihood (GEL) methods for moment condition models and shows that CU-GMM estimator is a special case of GEL methods. Peñaranda and Sentana [2015] show that, unlike two-step GMM estimator, CU-GMM estimator provides a unifying approach to the empirical evaluation of linear factor pricing models. The estimator developed in this paper is also related to, but different from kernel-weighted GMM estimators in Smith [2011], Gospodinov and Otsu [2012] and Kuersteiner [2012], since they all focus on models with stable moment conditions and stable parameters.

This paper also builds on a series of papers on how to handle structural instability in GMM framework. Ghysels and Hall [1990] propose the predictive test for stability of GMM estimates with one known breakpoint, which has been extended to the case of unknown breakpoints in Ghysels et al. [1998]. Andrews [1993] develops a supremum version of Wald, LM, and LR-type test for parameter stability with unknown breakpoints. Sowell [1996] presents optimal parameter stability tests for both one-sided and two-sided alternatives. Hall and Sen [1999] show that the null hypothesis of structural stability can be decomposed into one of parameter constancy and the other one for the validity of overidentifying restrictions in each sub-sample, and propose tests for both components. While the work mentioned above relies on conventional two-step GMM estimator, an extension for the test with one possibly unknown breakpoint to CU-GMM estimates has been provided in Hall et al. [2015]. Juhl and Xiao [2013] propose tests for stability of moment conditions, but parameters are assumed to be stable over time. Li and Müller [2009] develop time-varying two-step GMM estimation where a subset of the parameters are time-varying. Their focus is on deterministic and moderately large instabilities, in the sense that instability decreases as sample size increases. Creal et al. [2018] develop an observation-driven filtering approach of time-varying parameters with unknown dynamics based on moment conditions.

This paper also contributes to the literature on nonparametric modeling of smooth structural change in parameters. In linear models, Cai [2007] studies trending time-varying coefficient time series models with serially correlated errors. Chen and Hong [2012] develop testing procedures for smooth structural change in time series models. Chen [2015] provides inference procedures and tests for parameter stability in linear models with endogenous regressors. Both assume that parameter evolution is deterministic. Giraitis et al. [2014], Giraitis et al. [2018] and Giraitis et al. [2020b] develop a new framework for modeling smooth structural change in which time variation is assumed to be stochastic and smooth. Giraitis et al. [2014] and Giraitis et al. [2018] study autoregressive type models in which theory is developed under MDS assumption. Giraitis et al. [2020b] focus on linear IV models and consistency is established under weak dependent assumption. In the MLE framework, Chen and Hong [2016] propose an inference procedure and a test for parameter stability in Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models under deterministic time variation. Giraitis et al. [2016] provide a simple theoretical analysis on time-varying MLE estimation for static Tobit models under stochastic time variation.

This paper is closely related to Li et al. [2021] but differs in several important directions. First, they consider estimation and testing for deterministic smooth structural changes in conventional two-step GMM framework. As explained above, two-step GMM estimator has received some criticism in recent years, both theoretically and empirically[1]. Second, they assume that deterministic time-varying parameters are smooth functions of scaled time that can be approximated by a constant and develop a local constant estimator, which is more restrictive than the local linear estimator commonly employed in the literature. Finally, they only consider the test of parameter stability. As we shall see, the instability may not only come from parameters, and tests developed here can identify the source of instability.

[1]Hall et al. [2005] provides some surveys on finite sample issues related to two-step GMM estimators. Empirically, Hansen and Jagannathan [1997] argue that two-step GMM estimates cannot be used to evaluate asset pricing models.

3.2 Theoretical considerations

3.2.1 Model and the estimator

Suppose we have observed data for a sample of size T, $\{x_t\}_{t=1}^{T}$, $x_t = (y_t', z_t')' \in \mathcal{X}_t$, where $\mathcal{X}_t$ denotes sample space which is the subset of Euclidian space. $\theta_{0,t} \in \Theta$ is a $k \times 1$ vector of parameters of interests and Θ is the parameter space. All random elements are defined on a common probability space $(\Omega, \mathcal{F}, P)$. Consider the population conditional moment restrictions at time t implied by economic or finance theory,

$$E[g_t^c(w_t; \theta_{0,t})|\mathcal{F}_{-\infty}^t] = 0,$$

where $g_t^c : \mathcal{X}_t \times \Theta \to \mathbb{R}^{\bar{q}}$ is a vector-valued function which can be either linear or nonlinear in $\theta_{0,t}$, $w_t = (y_{t-\tau_T}, \cdots, y_{t+\tau_T})$, where $\tau_T = o(T)$ as $T \to \infty$, and $\mathcal{F}_{-\infty}^t = \sigma(x_s, \theta_s, s \leqslant t)$ is the information set. Note that, as in many rational expectation models (Fuhrer and Rudebusch [2004]), the framework allows both leads and lags of data to enter the model. g_t^c also has the subscript t, indicating that instability of moment conditions is allowed, as in Juhl and Xiao [2013]. The conditional moment restriction model defined above is typically estimated by the GMM estimator based on unconditional moment restrictions:

$$E[g_t(\theta_{0,t})] = 0, \tag{3.1}$$

where $g_t(\theta_{0,t}) = z_t \otimes g_t^c(w_t; \theta_{0,t}) : \mathcal{X}_t \times \Theta \to \mathbb{R}^q$.

As in constant coefficient GMM setting, the time-varying GMM estimator of $\theta_{0,t}$ is formed by choosing θ_t such that the sample average of $g_t(\theta_t)$ is close to its zero population value. Thus, we need an estimator for $E[g_t(\theta_{0,t})]$. Consider the kernel- weighted average of sample moment functions:

$$\overline{g}_t(\theta_t) = \frac{1}{K_t} \sum_{j=1}^{T} b_{jt,H} g_j(\theta_t), \tag{3.2}$$

where $K_t = \sum_{j=1}^{T} b_{jt,H}$. The kernel weights $b_{jt,H} = b\left(\frac{j-t}{H}\right)$ are computed with a kernel function $b(\cdot)$ and H is a bandwidth parameter. The corresponding GMM criterion function is given by

$$Q_{t,T} = \overline{g}_t'(\theta_t)\tilde{W}_t^{-1}(\theta_t)\overline{g}_t(\theta_t), \tag{3.3}$$

where $\tilde{W}_t(\theta_t)$ is a consistent estimator of the asymptotic long-run covariance matrix:

$$W_t(\theta_t) = \operatorname*{plim}_{T\to\infty} \operatorname{Var}\left(\frac{1}{K_{2,t}^{1/2}} \sum_{j=1}^{T} b_{jt,H} g_j(\theta_t)\right). \tag{3.4}$$

Then, the time-varying parameter continuous updated GMM (TVP CU-GMM) estimator is defined as the minimizer of (3.3):

$$\hat{\theta}_t = \arg\min_{\theta_t \in \Theta} Q_{t,T}. \tag{3.5}$$

I now provide three examples for the model (3.1) and their associated kernel- weighted average of sample moment functions.

Example 1. Asset pricing

Consider the dynamic asset pricing model

$$E[m_{t+1}(\theta_{0,t})r_{t+1,i}|\mathcal{F}_t] = 0,$$

where $i = 1, 2, \cdots, N$ and $\mathcal{F}_t$ is the investors' information set. $r_{t+1,i} = R_{t+1,i} - R_{t+1,f}$ is the excess return of asset (portfolio) i. m_{t+1} is a nonnegative stochastic discount factor implied by the absence of arbitrage opportunities, which is characterised by a $k \times 1$ vector of parameters $\theta_{0,t}$. m_{t+1} is commonly approximated by a linear function of candidate pricing factors f_{t+1}: $m_{t+1} = \theta_{00,t} + \theta'_{01,t} f_{t+1}$.

As explained in Peñaranda and Sentana [2015], if $k = 1$, $-\frac{\theta_{01,t}}{\theta_{00,t}}$ has a price of risk interpretation. Several papers have documented the fact that price of risk may be time-varying, see, for instance, Adrian et al. [2015] and Barroso et al. [2021].

Let R_{t+1} be a $N \times 1$ vector of stacked asset (portfolio) excess returns and select $z_t \in \mathcal{F}^t_{-\infty}$, where $\mathcal{F}^t_{-\infty}$ is the investors' information set, we obtain the unconditional moment restrictions:

$$E\Big[z_t \otimes ((\theta_{00,t} + \theta'_{01,t} f_{t+1})R_{t+1})\Big] = 0.$$

If the Gaussian kernel is used, (3.2) becomes

$$\overline{g}_t(\theta_t) = \frac{1}{K_t} \sum_{j=1}^{T} (1/\sqrt{2\pi}) e^{-(\frac{j-t}{H})^2} \Big(z_j \otimes ((\theta_{0,t} + \theta'_{1,t} f_{j+1})R_{j+1})\Big).$$

Example 2. New Keynesian Phillips curve (NKPC)

The NKPC is a rational expectation model capturing the trade-off between the rate of inflation and the level of real economic activity. A theoretical justification of the NKPC comes from the Calvo model. Consider the hybrid specification proposed by Galı and Gertler [1999]:

$$\pi_t = \gamma_{f,0t} E_t \pi_{t+1} + \gamma_{b,0t} \pi_{t-1} + \lambda_{0t} x_t + e_t,$$

where π_t is the inflation rate, $E_t \pi_{t+1}$ is the inflation expectation, π_{t-1} is the lagged inflation and x_t is a measure of real marginal cost. The GMM estimation is obtained by forming the moment conditions:

$$E\Big[z_t(\pi_t - \gamma_{f,0t} E_t \pi_{t+1} - \gamma_{b,0t} \pi_{t-1} - \lambda_{0t} x_t)\Big] = 0,$$

where z_t are any variables dated at t and earlier, which should satisfy both relevance and orthogonality conditions to be valid. Then, (3.2) becomes (with Gaussian kernel)

$$\overline{g}_t(\theta_t) = \frac{1}{K_t}\sum_{j=1}^{T}(1/\sqrt{2\pi})e^{-(\frac{j-t}{H})^2}\Big(z_j\big(\pi_j - \gamma_{f,t}\pi_{j+1} - \gamma_{b,t}\pi_{j-1} - \lambda_t x_j\big)\Big).$$

If we further assume that π_{t+1}, π_{t-1}, x_t are linearly related to z_t, we are back to time-varying linear IV models as in Chen [2015] and Giraitis et al. [2020b].

The original specification in Galı and Gertler [1999] assumes that parameters are constant over time. Several issues may arise which could make parameters unstable[2].

Example 3. Consumption-based Asset Pricing

The consumption-based asset pricing model relates real asset returns to consumption risk. Hansen and Singleton [1982] are first to exploit GMM estimation in this type of models. Consider a representative agent who maximizes her expected discounted utility

$$E\Big[\sum_{i=0}^{\infty}\beta_0^i U(c_{t+i})|\mathcal{F}_t\Big],$$

subject to her budget constraint, where c_t is consumption in period t and $\mathcal{F}_t$ is her information set. If the utility function is specified to have constant relative risk aversion (CRRA)

$$U(c_t) = \frac{c_t^{1-\gamma_0} - 1}{1-\gamma_0},$$

it can be shown that the Euler equation takes the form

$$E\Big[\beta_0(\frac{c_{t+1}}{c_t})^{-\gamma_0}R_{t+1}|\mathcal{F}_t\Big] = 1,$$

where R_{t+1} is a $N \times 1$ vector of real gross returns and γ is the risk aversion parameter.

The original specification in Hansen and Singleton [1982] assumes that both β and γ are constant over time. Malmendier and Nagel [2011], Cohn et al. [2015] and Guiso et al. [2018] all provide strong empirical and experimental evidence showing that risk aversion is time-varying among households, finance professionals, and investors respectively. If β_0 and γ_0 are all time varying, select $z_t \in \mathcal{F}_t$ and assume that the Gaussian kernel is used, (3.2) becomes

$$\overline{g}_t(\theta_t) = \frac{1}{K_t}\sum_{j=1}^{T}(1/\sqrt{2\pi})e^{-(\frac{j-t}{H})^2}\Big(z_j \otimes (\beta_t(\frac{c_{j+1}}{c_j})^{-\gamma_t}R_{j+1})\Big).$$

[2] Benati [2008] and Cogley et al. [2010] provide evidence of changing inflation persistence. With U.S. data, Galí and Gambetti [2019] find that Phillips curve is flattening over time. Bai et al. [2021] develop panel estimation method for NKPC model and also find instability with Eurozone data.

To make the approach operational, we need to specify $\tilde{W}_t(\theta_t)$. I propose the time-varying heteroskedasticity and autocorrelation consistent (HAC) estimates:

$$\tilde{W}_t(\theta_t) = \sum_{s=-T}^{T} k\Big(\frac{j}{L}\Big)\hat{\Gamma}(s), \tag{3.6}$$

where $k(\cdot)$ is another kernel function with bandwidth L which is different from $b(\cdot)$ and H for (3.2). $\hat{\Gamma}(s)$ is given by

$$\hat{\Gamma}(s) = \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H}\Big(g_j(\theta_t) - \frac{1}{T}\sum_{s=1}^{T} g_s(\theta_t)\Big)\Big(g'_{j+s}(\theta_t) - \frac{1}{T}\sum_{s=1}^{T} g_s(\theta_t)\Big), \tag{3.7}$$

where $K_{2,t} = \sum_{j=1}^{T} b_{jt,H}^2$ and $\hat{\Gamma}(s) = \hat{\Gamma}(-s)$ for $s < 0$.

3.2.2 Asymptotic theory

To establish the consistency and derive asymptotic distribution of (3.5), I impose the following regularity conditions.

Assumption 3.2.1. *$x_t = x_{T,t}$ is a triangular array of vectors whose elements $x_{\ell,t}$ is α-mixing, and its mixing coefficient,*

$$\alpha^x(h) = \sup_j \sup_{A\in\mathcal{F}_{-\infty}^{j}, B\in\mathcal{F}_{j+h}^{\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

satisfy $\alpha^x(h) = O(h^{-\gamma})$ for sone $\gamma > 1$.

Assumption 3.2.2. *$g_t(\cdot) : \mathcal{X}_t \times \Theta \to \mathbb{R}^q$, Θ is compact, $g_t(\cdot) = g_{T,t}(\cdot)$ is a triangular array of functions whose elements $g_{\ell,t}(\cdot)$ satisfy the following conditions*

(i) Given θ_t, $\{g_{\ell,t}(\theta_t)\}_t$ is a mixing process with mixing coefficient defined by

$$\alpha_T^g(h) = \sup_j \sup_{A\in\mathcal{F}_{-\infty}^{j}, B\in\mathcal{F}_{j+h}^{\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

which may depend on sample size T. $g_t(\theta_t) = g_t(w_t;\theta_t)$ is a measurable function with respect to $\mathcal{F}_{-\infty}^{t}$ and has twice continuously differentiable derivatives, $w_t = (y_{t-\tau_T}, \cdots, y_{t+\tau_T})$, $\tau_T = O(T^{\kappa})$ for some $\kappa \in [0, \frac{\nu^*}{4(1+\nu^*)})$ and $0 < \nu^* \leqslant 2$. Further more, the mixing coefficient defined in Assumption 2.1 satisfies

$$\sum_{h=1}^{\infty} h^{r^*} \alpha^x(h)^{\nu^*/(2+\nu^*)} < \infty,$$

for some $\frac{2\kappa(1+\nu^*)}{\nu^*-2\kappa(1+\nu^*)} < r^* < \frac{\gamma\nu^*}{1+\nu^*}$;

(ii) For some $r \in (2,4]$ and for some $p \geqslant r$, the following hold for all t:

$$\sup_{t\geqslant 1}\sup_{j\geqslant 1}\left|g_{\ell,j}(\theta_t)\right|_p < \infty$$
$$\forall(\ell,\ell_2),\quad \sup_{t\geqslant 1}\sup_{j\geqslant 1}\left|\frac{\partial g_{\ell,j}(\theta_t)}{\partial\theta_{\ell_2}}\right|_p < \infty,\ \sup_{t\geqslant 1}\sup_{j\geqslant 1}\left|\frac{\partial^2 g_{\ell,j}(\theta_t)}{\partial\theta_{\ell_2}^2}\right|_p < \infty;$$

(iii) Let $g_t^d(\theta_t) = \frac{\partial g_t(\theta_t)}{\partial\theta'}$ be the derivative matrix, $g_{\ell,t}(\theta_t)$ and elements in $g_t^d(\theta_t)$ satisfy the following smoothness condition:

$$\forall\ell,\quad \left|g_{\ell,t}(\theta_t) - g_{\ell,s}(\theta_s)\right| \leqslant C_1\left|\theta_{\ell,t} - \theta_{\ell,s}\right|$$
$$\forall(\ell,k),\quad \left|g_{\ell k,t}^d(\theta_t) - g_{\ell k,t}^d(\theta_s)\right| \leqslant C_2\left|\theta_{\ell,t} - \theta_{\ell,s}\right|,$$

for some fixed constants C_1, C_2 which do not depend on ℓ, k, t, s.

Assumption 3.2.3. *(i) Identification: $\forall t$, the matrix of derivatives:*

$$\frac{\partial E(g_t(\theta_t))}{\partial\theta'}$$

has full column rank at $\theta_{0,t}$;

(ii) Given $\theta_{0,t}$, the asymptotic long-run covariance matrix defined in (3.4) and the following matrix $W_{t,d,1}$ exist and are positive-definite:

$$W_{t,d_1} = \operatorname*{plim}_{T\to\infty}\frac{1}{K_{2,t}}\sum_{i=1}^{T}\sum_{j=1}^{T} b_{it,H}b_{jt,H}E\left[g_i(\theta_{0,t})(\frac{\partial g_j(\theta_{0,t})}{\partial\theta_{\ell_1}})\right],$$

for $\ell_1 = 1, 2, \cdots, k$, where $K_{2,t} = \sum_{j=1}^{T} b_{jt,H}^2$.

Assumption 3.2.4. *Define $\theta_t = \theta_{T,t}$ is a triangular array of vectors whose elements $(\theta_{\ell,t})$ are random processes that satisfy the smoothness condition*

$$\left|\theta_{\ell,t} - \theta_{\ell,s}\right| \leqslant (\frac{|t-s|}{T})^{\bar\gamma} r_{\ell,ts},\quad t, s = 1, 2, \cdots, T \tag{3.8}$$

for some $0 < \bar\gamma \leqslant 1$ and the distribution of variables $\Xi = \theta_{\ell,t}, r_{\ell,ts}$ has a thin tail

$$\mathbb{P}\left(|\Xi| \geqslant \omega\right) \leqslant \exp(-c|\omega|^{\alpha}),\quad \omega > 0$$

for some $c > 0$, $\alpha > 0$ that do not depend on ℓ, t, s and T.

Assumption 3.2.5. *(i) The kernel weights $b_{jt,H}$ are computed with a kernel function $b(\cdot) : \mathbb{R} \longrightarrow \mathbb{R}^+$, with $b_{jt,H} := b\left(\frac{j-t}{H}\right)$; (ii) $\int_{-\infty}^{\infty} b(a)da = 1$, $\int_{-\infty}^{\infty} b^{\frac{j}{2}}(a)da < \infty$ for $2 < j \leqslant 4$; (iii) $b(0) < \infty$ and $b(\cdot)$*

is continuous at 0 and almost everywhere; (iv) H is a bandwidth parameter such that $H = o(T)$ as $T \to \infty$; (iv) $|b(x)| \leqslant \bar{b}(x)$, where $\bar{b}(x)$ is nonincreasing function such that $\int_{-\infty}^{\infty} |x|^{-\gamma} \left|\bar{b}(x)\right| dx < \infty$, where $\gamma > 1$ is the size of the mixing coefficient defined in Assumption 2.1; (v) Define $b_2 = \int_{-\infty}^{\infty} b^2(a)da$ and let $b^*(d) = \frac{1}{b_2}\int_{-\infty}^{\infty} b(c)b(c+d)dc$, $\int_{-\infty}^{\infty} b^*(d) < \infty$.

Assumption 3.2.6. *(i) The kernel weights $k(\cdot)$ are computed with a kernel function $k(\cdot) : \mathbb{R} \longrightarrow [-1, 1]$, $k(0) = 1$, $k(x) = k(-x)$, $\forall x \in \mathbb{R}$; (ii) $\int_{-\infty}^{\infty} k(x)dx < \infty$, $k(\cdot)$ is continuous at zero and for almost all $x \in \mathbb{R}$; (iii) $|k(x)| \leqslant \bar{k}(x)$, where $\bar{k}(x)$ is a nonincreasing function such that $\int_{-\infty}^{\infty} |x| \left|\bar{k}(x)\right| dx < \infty$; (iv) L is another bandwidth parameter which may be different from but depend on H, such that $L = o(H^{1/2-1/r})$ as $T \to \infty$ for some $r \in (2, 4]$.*

Assumption 2.1 is a standard mixing condition to control the temporal dependence. The mixing coefficient is required to decay at a polynomial rate. Unlike Creal et al. [2018] and Li et al. [2021], $\{x_t\}_t$ are not required to be stationary or approximately stationary. Assumption 2.2 is a set of regularity conditions for the moment functions. As shown in Lemma B2, $\{g_{\ell,t}(\theta_t)\}_t$ is also a α-mixing process. The dependence of $\{g_{\ell,t}(\theta_t)\}_t$ is allowed to be possibly depend on sample size. Restrictions on τ_T also imply that the number of leads and lags in w_t has to be very small relative to sample size. As stated in the final piece of Assumption 2.2(i), some further restrictions are also needed for the mixing coefficients for $\{x_t\}_t$. Assumption 2.2(ii) is a standard condition to put some uniform bounds on the moments of $g_t(\cdot)$ and its associated first and second order derivatives. However, Assumption 2.2(iii) is high level and nonstandard. Although instability of moment conditions is allowed, the amount of time variation of both moment function and its associated derivatives has to be bounded. Assumption 2.3(i) is a standard condition to ensure that parameters in $\theta_{0,t}$ are locally identified at each t. Assumption 2.3(ii) is another high level condition to establish consistency of (3.6).

Assumption 2.4 is the one used in Giraitis et al. [2020b], which imposes conditions on time-varying structural parameters of interests. It implies that elements in θ_t are smoothly varying persistent stochastic processes. For example, one can specify an array of random processes as $\theta_{\ell,t} = \frac{1}{\sqrt{T}} u_{\ell,t}$, $t = 1, 2, \cdots, T$, where $u_{\ell,t}$ are random walk process such that $u_{\ell,t} - u_{\ell,t-1} = \nu_{\ell,t}$, where $\nu_{\ell,t}$ follows a weakly stationary ARMA(p, q) process. As explained in Giraitis et al. [2014], $\theta_{\ell,t}$ can include both deterministic and stochastic components, $\theta_{\ell,t} = \frac{1}{\sqrt{T}} u_{\ell,t} + m_\ell(t/T)$, where $m_\ell(t/T)$ is a deterministic function of scaled time t/T. $r_{\ell,ts}$ is a thin tailed component to introduce randomness in the time variation. More discussions and examples are provided in Giraitis et al. [2014], Giraitis et al. [2018] and Dendramis et al. [2021].

It is of some interest to compare Assumption 2.4 with local linear methods used in the literature on nonparametric modeling of deterministic time variation. Suppose that $\theta_{\ell,t} = \beta_\ell(t/T), \beta_\ell(\cdot)$ is a deterministic function of scaled time with continuous bounded first and second order derivative. First order Taylor expansion gives $\beta_\ell(t/T) \approx \beta_\ell(u) + \beta_\ell'(u)(t/T - u)$, $\ell = 1, 2, \cdots, k,$, where $u \in [0, 1]$ and $\beta'(\cdot)$ is the derivative of $\beta(\cdot)$[3]. By setting $u = s/T$, we obtain $\beta_\ell(t/T) \approx \beta_\ell(s/T) + \beta_\ell'(s'/T)(t/T - s/T)$, with s'

[3]In Li et al. [2021], local constant estimator is proposed, which implies that they assume $\beta_\ell(t/T) \approx \beta_\ell(u)$, which is a restrictive case of local linear methods.

between s and t, which implies that $\left|\theta_{\ell,t} - \theta_{\ell,s}\right| \leqslant C\left(\frac{|t-s|}{T}\right)$, $\forall C > 0$. This implies that local linear method is a special case of Assumption 2.4 obtained by setting $\bar{\gamma} = 1$ and $r_{\ell,ts} = C$, which is the upper bound of the derivative.

Assumptions 2.5 and 2.6 introduce conditions for kernel weights and bandwidth parameters used to construct (3.2) and time-varying HAC estimator (3.6). Assumption 2.5(ii) is similar to the one used in Giraitis et al. [2014], in which they require the integral to exist with order one and two. Assumptions 2.5(iv) and 2.5(v) are nonstandard. They are used in Lemma 1 to establish uniform weak law of large numbers and consistency of time-varying HAC estimator. Examples of $b(\cdot)$ satisfying Assumption 2.1 include the commonly used Gaussian kernel: $b(x) = (1/\sqrt{2\pi})e^{-x^2}$. Assumption 2.6 is essentially the one used in De Jong [2000] to establish weak consistency and ensure that the estimator is positive definite. Examples of allowable kernel functions include the commonly used Bartlett Kernel: $k(x) = 1 - |x|\, I_{|x|\leqslant 1}$. However, unlike De Jong [2000], since (3.7) is constructed by kernel-weighted average, L goes to infinite with a rate slower than $H^{1/2-1/r}$, rather than $T^{1/2-1/r}$, for some $2 < r \leqslant 4$.

The large sample properties of consistency and asymptotic normality of TV CU-GMM estimator defined in (3.5) rely on a uniform law of large numbers (UWLLN) and central limit theorem (CLT) defined in terms of kernel weighted sample average $\overline{g}_t(\theta_t)$, as well as the consistency of time-varying HAC estimator defined in (3.6). These are provided in the following lemma.

Lemma 3.2.1. *Under Assumptions 2.1-2.6, for each* $t = \lfloor \tau T \rfloor$, $0 < \tau < 1$, *as* $T \to \infty$,

(i) UWLLN:

$$\sup_{\theta_t \in \Theta} \left\| \overline{g}_t(\theta_t) - E(g_t(\theta_t)) \right\| = O_p\Big(H^{-1/2} + (\frac{H}{T})^{\bar{\gamma}}\Big)$$

(ii) CLT:

$$\frac{K_t}{K_{2,t}^{1/2}}\Big(\overline{g}_t(\theta_t) - E(\overline{g}_t(\theta_t))\Big) \xrightarrow{d} \mathcal{N}(0, W_t(\theta_t)),$$

where $W_t(\theta_t)$ *is defined in* (3.4).

(iii)

$$\tilde{W}_t \xrightarrow{p} W_t(\theta_t).$$

In Smith [2011], UWLLN and CLT are established for a similar kernel-weighted average of moment functions in the case of constant coefficients and stable moment conditions. He assumes that $\{x_t\}_t$ are stationary but Lemma 1 above shows that results still hold without the stationarity assumption. However, more conditions on memory (in terms of rate of decay of mixing coefficients) of $\{x_t\}_t$ and on processes $\{g_{\ell,t}(\theta_t)\}_t$ are required, see Assumptions 2.1 and 2.2(i). Since moment conditions may not be stable, another condition is needed to control for time variation for both moment function and its associated derivatives, see Assumption 2.2(iv). The proof of CLT in Lemma 1(ii) requires to verify a series of technical assumptions in Francq and Zakoïan [2005], which are mostly summarized in Assumption 2.2(i). These conditions are more intuitive and easier to verify than existing literature on CLT for mixing processes. To prove consistency of time-varying HAC estimator, as stated in Lemma 1(iii), we first need to extend the bounds provided in Hansen [1991] and Hansen [1992]. Then, the strategy employed in De Jong [2000] can be used to prove the results. The assumption of existence of only slightly more than second moments of $\{g_t(\cdot)\}_t$ is weaker than Smith [2011],which requires the restrictive assumption of moments of order higher than four. Unlike the constant coefficient case, as stated in Assumption 2.6(iv), consistency requires L to be scalable to H: $L = o(H^{1/2-1/r})$, for some $2 < r \leqslant 4$. As we shall see in Theorem 1, CLT requires that $(\frac{H}{T})^{\gamma} = o(H^{-\frac{1}{2}})$, which implies that we need to choose a relatively small value of L in practice.

Theorem 1 establishes the consistency and asymptotic normality for the TVP CU-GMM estimator $\hat{\theta}_t$.

Theorem 3.2.2. *Suppose Assumptions 2.1-2.6 are satisfied, for each* $t = \lfloor \tau T \rfloor$, $0 < \tau < 1$, *the TVP CU-GMM estimator defined in* (3.5) *has the following properties*

(i) Consistency:

$$\left\|\hat{\theta}_t - \theta_{0,t}\right\| = O_p\Big((\frac{H}{T})^{\bar{\gamma}} + H^{-1/2}\Big)$$

(ii) Asymptotic normality: if $(\frac{H}{T})^{\bar{\gamma}} = o(H^{-\frac{1}{2}})$, *then*

$$\frac{K_t}{K_{2,t}^{1/2}}(G_t' W_t^{-1} G_t)^{1/2}(\hat{\theta}_t - \theta_{0,t}) \xrightarrow{d} N(0, I_k).$$

where $K_t = \sum_{j=1}^{T} b_{jt,H}$, $K_{2,t} = \sum_{j=1}^{T} b_{jt,H}^2$ *and*

$$G_t = E\left[\frac{\partial g_t(\theta_{0,t})}{\partial \theta'}\right], \quad W_t = \operatorname*{plim}_{T\to\infty} Var\Big(\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt,H}\, g_j(\theta_{0,t})\Big).$$

By Lemma B1, $K_t = O(H)$, $K_{2,t} = O(H)$. Theorem 1(ii) implies that the convergence rate of the TVP CU-GMM estimator is proportional to $\sqrt{H}$. However, the exact rate depends on the choice of kernel, bandwidth and also on t. As explained in Giraitis et al. [2014], consistency is guaranteed by the persistency of the process $\theta_{0,t}$. If $\bar{\gamma}$ is close to 0, the consistency rate in Theorem 1(i) deteriorates. While consistency holds under minimal restrictions on H, asymptotic normality requires that $(\frac{H}{T})^{\bar{\gamma}} = o(H^{-\frac{1}{2}})$. If $\gamma = 1$, the condition implies that $H = o(T^{2/3})$, which is the maximum allowable value for H.

In the following corollary, we show that both G_t and W_t can be consistently estimated. This result can be used to construct pointwise confidence intervals for $\hat{\theta}_t$.

Corollary 1. *Under Assumptions 2.1-2.6, let*

$$\hat{G}_t = \frac{1}{K_t}\sum_{j=1}^{T}\frac{\partial g_j(\hat{\theta}_t)}{\partial \theta'}, \quad \hat{W}_t = \tilde{W}_t(\hat{\theta}_t),$$

where $K_t = \sum_{j=1}^{T} b_{jt,H}$ and $\tilde{W}_t$ is defined in (3.6). *Then, suppose that $(\frac{H}{T})^{\bar{\gamma}} = o(H^{-\frac{1}{2}})$, we have*

$$\hat{G}_t \xrightarrow{p} G_t, \quad \hat{W}_t \xrightarrow{p} W_t$$

3.3 Tests of structural stability

In this section, I propose two tests of structural stability for models given in (3.1). The null hypothesis of structural stability states that (3.1) holds for same θ_0 throughout the sample. In the GMM literature, attention has focused almost exclusively on the case where the instability involves discreet changes at some possibly unknown points in the sample known as the ”break points”. The only exception is Li et al. [2021], who propose a test for smooth structural change. They focus on the hypothesis of stability of parameters:

$$\mathbb{H}_0 : \quad \theta_{0,t} = \theta_0, \quad \forall t, \text{ for some } \theta \in \mathbb{R}^k,$$

and the alternative hypothesis is that $\mathbb{H}_0$ is false. However, in models with conditional moment restrictions as in (3.1), it is more interesting to discriminate different sources of instability.

As in Ghysels and Hall [1990], write the null hypothesis of structural stability as

$$\mathbb{H}_0^{\mathrm{SS}} : \quad E\Big[g_t(\theta_0)\Big] = 0, \;\; \forall t, \;\; \text{for some unique } \theta_0 \in \mathbb{R}^k.$$

In view of Sowell [1996] and Hall and Sen [1999], define $G_{w,t} = W_t^{-1/2} G_t$, where G_t and W_t are given in Theorem 1(ii), and rewrite (3.1) as

$$G'_{w,t} W_t^{-1/2} E\Big[g_t(\theta_{0,t})\Big] = 0. \tag{3.9}$$

(3.9) can be interpreted as the least square projection of $W_t^{-1/2} E\Big[g_t(\theta_{0,t})\Big]$ onto the column space of $G_{w,t}$ is zero. Then, by multiplying both sides in (3.9) with $G_{w,t}\Big(G'_{w,t}G_{w,t}\Big)^{-1}$, we obtain an alternative representation:

$$G_{w,t}\Big(G'_{w,t}G_{w,t}\Big)^{-1} G'_{w,t} W_t^{-1/2} E\Big[g_t(\theta_{0,t})\Big] = 0. \tag{3.10}$$

By Assumption 2.3(i), we know that $\operatorname{rank}(G_{w,t}\Big(G'_{w,t}G_{w,t}\Big)^{-1} G'_{w,t}) = \operatorname{rank}(G_{w,t}) = k$. Thus, the independent k moment conditions in (3.10) are used to identify the parameters $\theta_{0,t}$. Since the identifying restrictions are imposed in estimation, there are always parameter values which satisfy them in each t. Therefore, the identifying restrictions are said to be stable if they are satisfied by same θ_0, $\forall t$, which is formally stated as

$$\mathbb{H}_0^{\mathrm{I}} : \quad G_{w,t}\Big(G'_{w,t}G_{w,t}\Big)^{-1} G'_{w,t} W_t^{-1/2} E\Big[g_t(\theta_0)\Big] = 0, \;\; \text{for some } \theta_0 \in \mathbb{R}^k.$$

It can now be recognised that $\mathbb{H}_0^{\mathrm{I}}$ is equivalent to the null hypothesis of parameter stability. I consider to compare the differences between the following sample averages of moment conditions:

$$D_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}_t) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}), \tag{3.11}$$

where $\hat{\theta}_t$ is obtained from (3.5) and $\hat{\theta}$ is obtained from standard constant coefficient CU-GMM estimation:

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \Big(\frac{1}{T}\sum_{t=1}^{T} g_t(\theta)\Big)' \tilde{W}_T^{-1}(\theta)\Big(\frac{1}{T}\sum_{t=1}^{T} g_t(\theta)\Big),$$

where $\tilde{W}_T(\theta)$ is a consistent estimator of

$$W_T(\theta) = \operatorname*{plim}_{T\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_j(\theta)\right).$$

Define

$$\Omega_{1,T} = \operatorname*{plim}_{T\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}_t) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta})\right)$$

and let $\hat{\Omega}_{1,T}$ be a consistent estimator of $\Omega_{1,T}$[4], I propose the following Q test statistic

$$Q_T = \Big(\hat{\Omega}_{1,T}^{-1/2} D_T\Big)'\Big(\hat{\Omega}_{1,T}^{-1/2} D_T\Big). \tag{3.12}$$

Let us move to the stability of overidentifying restrictions. It follows immediately from (3.10) that the moment conditions not used in estimation are given by

$$\left(I_q - G_{w,t}\Big(G'_{w,t}G_{w,t}\Big)^{-1} G'_{w,t}\right) W_t^{-1/2} E\Big[g_t(\theta_{0,t})\Big] = 0. \tag{3.13}$$

As $\operatorname{rank}(I_q - G_{w,t}(G'_{w,t}G_{w,t})^{-1}G'_{w,t}) = q - k$, the independent $q - k$ moment conditions are often labeled as overidentifying restrictions. The stability of overidentifying restrictions implies that they hold for all t, which is formally stated as

$$\mathbb{H}_0^{\mathrm{O}} : \left(I_q - G_{w,t}\Big(G'_{w,t}G_{w,t}\Big)^{-1} G'_{w,t}\right) W_t^{-1/2} E\Big[g_t(\theta_{0,t})\Big] = 0, \quad \forall t.$$

Then, the alternative hypothesis $\mathbb{H}_1^{\mathrm{O}}$ implies that overidentifying restrictions are invalid for some t. Define

$$V_{T,t} = \hat{W}_t^{-1/2} \frac{1}{K_{2,t}^{1/2}} \sum_{j=1}^{T} b_{jt,H}\, g_j(\hat{\theta}_t),$$

[4]Detailed formulas of $\Omega_{1,T}$ and $\hat{\Omega}_{1,T}$ are given in the proof of Theorem 2.

it is clear that $V_{T,t}$ will be very close to zero under $\mathbb{H}_0^{\mathrm{O}}$. Consider to construct the test statistic by the squared version of the scaled average of $V_{T,t}$ for $t = 1, 2, \cdots, T$:

$$J_T = \Big(\hat{\Omega}_2^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_{T,t}\Big)'\Big(\hat{\Omega}_2^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_{T,t}\Big) \tag{3.14}$$

where $\hat{\Omega}_2$ is a consistent estimate of Ω_2[5]:

$$\Omega_2 = \operatorname*{plim}_{T\to\infty} \mathrm{Var}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_{T,t}\Big).$$

Therefore, the above analysis implies that any instability must be referred to as the violation of one of the null hypotheses $\mathbb{H}_0^{\mathrm{I}}$ and $\mathbb{H}_0^{\mathrm{I}}$. It follows that

$$\mathbb{H}_0^{\mathrm{SS}} = \mathbb{H}_0^{\mathrm{I}} \;\&\; \mathbb{H}_0^{\mathrm{O}}.$$

Both (3.12) and (3.14) can be used to identify the source of instability[6]. The asymptotic distributions of the two test statistics under the associated null hypothesis are presented in next theorem.

Theorem 3.3.1. *Under Assumptions 2.1-2.6*

(i) Given $\mathbb{H}_0^{O}$, under $\mathbb{H}_0^{I}$, suppose

- *$\hat{\theta}$ is root T consistent:*

$$\sqrt{T}\,(\hat{\theta} - \theta) = O_p(1)$$

- *$\{g_t(\theta)\}_t$ is a martingale difference sequence (MDS),*

we have that

$$Q_T \xrightarrow{d} \chi_q^2$$

(ii) Given $\mathbb{H}_1^{I}$, under $\mathbb{H}_0^{O}$, suppose

- *$\{g_t(\theta_t)\}_t$ is a martingale difference sequence (MDS),*

we have that

$$J_T \xrightarrow{d} \chi_{q-k}^2.$$

[5]Detailed formulas of $\Omega_{2,T}$ and $\hat{\Omega}_{2,T}$ are given in the proof of Theorem 2.

[6]The evaluation of the ability of the two test statistics to identify the source of instability requires to assess their power against alternatives. In the current version of the paper, this is done by Monte-Carlo simulations, which are presented in the next section

3.4 Monte Carlo study

In this section, I evaluate the finite sample performance of the time-varying CU-GMM estimator, and size and power of both Q-test and J-test. I consider the linear IV model with time-varying parameters as the data generating process (DGP):

$$y_t = x_t\beta_t + bz_{1,t} + u_t, \quad x_t = \psi_t' Z_t + v_t, \quad t = 1, 2, \cdots, T + R, \tag{3.15}$$

where $Z_t = (z_{1,t}, z_{2,t}, z_{3,t})'$ is a 3×1 vector of candidate instruments with associated parameters $\psi_t = (\psi_1, \psi_{2t}, \psi_{3t})'$. $R = 200$ is the number of observations used as burn-in. There is one endogenous variable x_t with 3 candidate instruments so the model is over identified. Instruments are generated according to

$$z_{j,t} = \rho_j z_{j,t-1} + \sqrt{1 - \bar{\rho}_j^2} e^z_{j,t}, \quad j = 1, 2, 3, \tag{3.16}$$

where $\bar{\rho}_j$ are obtained from uniform distribution: $\bar{\rho}_j \overset{i.i.d.}{\sim} \mathcal{U}[-1, 1]$ and $e^z_{j,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. I introduce time-varying correlation between u_t and v_t by specifying them as

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}\right)$$

where $\rho_t = \rho \frac{\xi_t}{\max_{1 \leqslant j \leqslant t} |\xi_j|}$, $t = 1, 2, \cdots, T+R$ is a bounded random walk process, such that $\xi_t - \xi_{t-1} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$.

The parameters $\psi_{\ell,t} = T^{-1/2}\xi_t^{(\ell)}$, $t = 1, 2, \cdots, T + R$, $\ell = 1, 2, 3$ are generated as three independent scaled random walk processes, where $\xi_t^{(\ell)} - \xi_{t-1}^{(\ell)} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $\ell = 1, 2, 3$. The parameters β_t are generated either according to $\beta_t = T^{-1/2}\xi_t$, where $\xi_t - \xi_{t-1} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, or $\beta_t = 1$, $t = 1, 2, \cdots, T + R$. For b, I set $b = 0$ (under $\mathbb{H}_0^{\mathrm{O}}$) or $b = 1 + \mathcal{N}(0, 1)$, for $t = \lfloor T/3 \rfloor + 1 + R, \cdots, \lfloor 2T/3 \rfloor + R$. Then, based on different generating mechanisms of β_t and b, there are four different DGPs:

- DGP 1: $\beta_t = T^{-1/2}\xi_t$, $b = 0$;
- DGP 2: $\beta_t = 1$, $b = 0$;
- DGP 3: $\beta_t = T^{-1/2}\xi_t$, $b = 1 + \mathcal{N}(0, 1)$;
- DGP 4: $\beta_t = 1$, $b = 1 + \mathcal{N}(0, 1)$.

DGP 1 is used to assess finite sample performance of the estimator. DGPs 1–4 are used to evaluate size and power of both Q-test and J-test in order to identify the source of possible instability.

The estimators are computed using the Gaussian kernel $b(x) = \exp(-x^2/2)$ with a variety of bandwidth values $H = T^\alpha$ for $\alpha = 0.4, 0.5, 0.7$. As explained in Giraitis et al. [2014], the DGP design of β_t implies that $\bar{\gamma} = 0.5$ and $H = o(T^{1/2})$ is required for CLT, but a higher bandwidth value may increase the efficiency, and it is interesting to evaluate the trade-off. For time-varying HAC estimator, since consistency requires a very small value of the bandwidth, following Assumption 2.6, I set $L = H^{0.3}$ and use the Bartlett kernel: $k(x) = 1 - |x|, |x| \leqslant 1$. I consider four sample sizes: $T = 750, 1000, 2000, 3000$. To compare the finite sample properties of the estimators, both the TV-OLS estimator in Giraitis et al. [2014] and the TV-IV estimator in Giraitis et al. [2020b] are used as competitors. The Monte Carlo analysis is based on 1,000 replications. An example of realized β_t, together with the estimates from TV-OLS, TV-IV and TV CU-GMM and the associated 95% confidence interval, with $T = 750$, are provided in Figure 1[7].

The performance of the estimators is evaluated by the root mean squared error (RMSE)

$$\text{RMSFE} = \sqrt{\frac{1}{T-2H}\sum_{t=H+1}^{T-H}(\hat{\beta}_t - \beta_t)^2},$$

the mean absolute deviation (MAD):

$$\text{MAD} = \frac{1}{T-2H}\sum_{t=H+1}^{T-H}\left|\hat{\beta}_t - \beta_t\right|,$$

and the 95% coverage rate, which is the estimated probability that the true β_t lies in the interval $(\hat{\beta}_t - 1.96\text{sd}(\hat{\beta}_t), \hat{\beta}_t + 1.96\text{sd}(\hat{\beta}_t))$, where $\text{sd}(\hat{\beta}_t)$ is the estimated variance of the estimator obtained from the associated asymptotic distributions.

Table 1 reports RMSE, MAD, and coverage rate of all estimators. Since x_t is endogenous, TV-OLS estimator is clearly not preferable. It delivers larger bias and lower coverage rates. For TV-IV and TV-CU GMM, both RMSE and MAD decrease with the increase of sample size. TV-IV estimator has smaller RMSE and MAD in all cases, particularly for smaller bandwidth value $H = T^{0.4}$, but has a lower coverage rate compared to TV CU-GMM. One explanation is that IV estimator has closed-form solutions, but numerical optimization is required for TV CU-GMM. Since all elements in Z_t follow an AR(1) process, moment conditions are serially correlated, so TV CU-GMM delivers a higher coverage rate. $H = T^{0.5}$ yields the lowest values of criteria (RMSE and MAD) but also has a slightly lower coverage rate compared to the case of $H = T^{0.4}$. Setting $H = T^{0.7}$ reduces the accuracy of the estimators. Coverage rate is also much lower since the condition for CLT (see, Theorem 1) is not satisfied.

Table 2 reports rejection frequencies for both Q-test and J-test under all four DGPs defined above. Overall, both tests have good power, and it increases with T. Power is also higher when we use larger bandwidth $H = T^{0.7}$, but differences compared to other choices of H are rather small. From DGP 2, it turns out that both tests are slightly oversized. DGP 1 indicates that when identifying restrictions are not stable, size distortion of J-test also gets larger for $H = T^{0.7}$. However, from DGP 4, we see that when overidentification restrictions are not stable, size distortion of Q-test is very sizeable.

Results in Tables 1 and 2 also provide some guidance on the choice of bandwidth. In some unreported

[7] As all of the estimators have slower convergence rate at the boundary points, performances of the estimators are also worse there compared to interior points.

MC results, it is found that the choice of L generally has little impact, so choice of H plays a more crucial role. Overall, setting $H = T^{0.5}$ leads to best performance in terms of bias and coverage probability considerations as well as size and power considerations. The results also provide a strategy to diagnose the source of instability. The strategy proposed by Hall and Sen [1999] in the structural change context remains valid here. If all tests fail to reject, then this is evidence that all aspects of the model are stable. If Q-test rejects but J-test fails to reject, then there is evidence of parameter variation. In all the other cases, any rejection would imply that instability could involve more than parameter time variation.

3.5 Empirical application

In this section, I present an empirical application on asset pricing model for equity and treasury returns. As example 1 in section 2.1, I consider the dynamic asset pricing model with stochastic discount factor (SDF) representation:

$$E[m_{t+1}(\theta_{0,t})R_{t+1}|\mathcal{F}_t] = 0,$$

where $m_{t+1}(\theta_t)$ is the SDF and R_{t+1} is a $N\times 1$ vector of excess returns. Following Cochrane [2009], I further normalize the SDF by setting $\theta_{00,t} = 1$, $\forall t$:

$$m_{t+1}(\theta_{0,t}) = 1 - \theta'_{01,t}f_t,$$

where the time-varying parameters $\theta_{01,t}$ are assumed to satisfy Assumption 2.4. According to Peñaranda and Sentana [2015], $\theta_{01,t}$ also has a price of risk interpretation.

With regard to testing portfolios and pricing factors, I follow Adrian et al. [2015] to obtain ten size-sorted portfolios for US equities from Ken French's online data library, combined with constant maturity Treasury portfolios with maturities one, two, five, seven, ten, 20, and 30 years from the Center for Research in Securities Prices(CRSP). Pricing factors used include the excess return on the value-weighted equity market portfolio (Mkt) from CRSP and the small minus big(SMB) portfolio from Fama and French [1993], and the ten-year Treasury yield(TSY10)[8]. I obtain the first two factors from French's website and the third from Federal Reserve Economic Data. The data is monthly and spans the period 1964:01-2020:12, for a total of 684 observations. Note that, Adrian et al. [2015] use beta representation and Fama-MacBeth two-pass procedure to obtain price of risk estimates and evaluate models by computing pricing errors. They are agnostic to the asset-specific beta dynamics but impose some structure on the dynamics of price of risk. Here, the estimates are obtained via SDF representation, and the approach is entirely agnostic since no specification of dynamics is required even for price of risk.

To implement the time-varying GMM estimation, I choose Gaussian kernel with bandwidth $H = T^{0.5}$ and Bartlett kernel with bandwidth $L = H^{0.3}$, as motivated by asymptotic theory and Monte-Carlo results. I also report p-values obtained from both Q-test and J-test, at 5% significance level. To evaluate asset pric-

[8]Since these factors are also traded, I also include them in R_{t+1} (in terms of excess returns).

ing models, I consider Hansen-Jagannathan (HJ) distance, pioneered in Hansen and Jagannathan [1997], which is commonly used to evaluate SDF. HJ distance is computed as

$$\text{HJ} = \sqrt{\Big(\frac{1}{T}\sum_{t=1}^{T}(1-\hat{\theta}_{1,t}'f_{t+1})\mathbf{r}_{t+1}\Big)'\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_{t+1}\mathbf{r}_{t+1}'\Big)^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}(1-\hat{\theta}_{1,t}'f_{t+1})\mathbf{r}_{t+1}\Big)},$$

where $\hat{\theta}_{1,t}$ is the time-varying estimates. Apart from continuously-updated time-varying GMM estimator, I also consider linear time-varying GMM estimator, obtained by setting the weighting matrix equal to $\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\mathbf{r}_{t+1}\mathbf{r}_{t+1}'$. In addition, I also follow Cochrane [1996] to consider $\hat{\theta}_{1,t}$ as an affine function of $\bar{k}\times 1$ vector of state variables z_t: $\theta_{01,t} = z_t\theta_1$. Then, I plug this into pricing equations to estimate θ_1 by standard constant coefficient CU-GMM method. As in Adrian et al. [2015], z_t includes term spread, which is measured by the differences between the yield on a ten-year Treasury note and the three-month Treasury bill; log dividend yield on S&P 500 index and TSY10.

Figure 3.2 provides plots of price of risk estimates. It is clear that there is substantial time variation in price of risk estimates, which is particularly evident for factors commonly used to price equity portfolios: Mkt and SMB. The estimates are centered around zero but have been significantly different from zero over various subperiods in the sample. The price of risk estimates from TSY10 factor, which is a good proxy of the level of the term structure of Treasury yields (Adrian et al. [2013]), is different from other factors. The estimates are smoother and significantly larger than zero in all cases. There is also an interesting pattern of the estimates since they first decline until the 1980s but then increase afterward. Table 3.3 reports p-values from both Q-test and J-test from time-varying parameter models. We see clearly that structural stability for both identifying and overidentifying restrictions is strongly rejected in all cases at 5% significance level.

Tables 3.4 reports HJ distance from different estimation methods. Gray shading indicates the best performing estimator. As a robustness check, I also consider estimation in the original sample period: 1964/1–2012/12 as in Adrian et al. [2015]. Time-varying CU-GMM estimator has the best pricing performance. Gains from time-varying parameters are substantial. HJ distance estimates obtained from constant coefficient models are at least three times larger than time-varying parameter models. However, gains are smaller from the extended sample period: 1964/1–2020/12. The pricing performance from time-varying CU-GMM and linear GMM are roughly similar, and gains from CU-GMM are sightly larger again in the original sample period. The pricing performance from the affine function approach is always the worst, indicating that the price of risk dynamics is likely to be misspecified.

3.6 Conclusions

Although attempts have been made to handle structural change in GMM framework, little has been done under smooth structural change, particularly for continuously updated GMM estimator. In this paper, I introduce time-varying continuously updated GMM estimator for models with stochastic time-varying parameters and unstable moment conditions, taking a non-parametric approach to remain as agnostic as possible regarding the type of parameter evolution and changes of moment conditions

I establish the asymptotic properties of the time-varying CU-GMM estimator. I also extend Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix estimation in constant coefficient setting to models with time-varying parameters and prove the consistency results under conventional increasing smoothing asymptotics. Then, I show that the null hypothesis of structural stability can be decomposed into stability for identifying and overidentifying restrictions. I propose two structural stability tests for these two components and derive the limiting distributions of these test statistics.

Next, I evaluate the finite sample properties of the estimator as well as the size and power of the proposed tests in an extensive Monte-Carlo study. The results show that performance from time-varying CU-GMM estimator, in terms of biases and coverage rates, is satisfactory. The proposed tests also have reasonable size and power. They can be used to identify the source of structural instability and strategy proposed in structural change literature can be directly applied. In terms of the choice of bandwidth, I find that the choice of bandwidth parameter for constructing sample averages of moment functions is far more important than the choice of bandwidth parameter for the time-varying HAC covariance matrix. By setting the value (for constructing sample averages of moment functions) to $T^{0.5}$ leads to the overall best performance for both estimators and the tests.

I illustrate the methods by an empirical application on dynamic asset pricing models with SDF representation for the joint cross-section of equity and treasury portfolios. I find substantial time variation in price of risk estimates. By allowing for time variation improves the pricing performance, but misspecification of dynamics of price of risk can lead to large pricing errors. Time-varying CU-GMM estimator has the overall best pricing performance.

The focus of this paper has been on consistent estimation and pointwise inference of the path coefficients β_t, under the assumption that elements in β_t are smoothly varying persistent stochastic processes. However, deterministic time variation is still a dominating approach in the literature. Thus, it is of great interest to develop tests for the null hypothesis of deterministic time variation, which is one of the current research topics by the author. Another interesting extension of the current paper is to provide a formal asymptotic power analysis of the two structural stability tests. In addition, I follow Chen and Hong [2012] to develop averaging statistics for structural stability. An advantage of this type of test statistics is that asymptotic distributions are standard, so there is no need to tabulate critical values. However, as Hoesch et al. [2020] point out, this may have low power if the null of stability is false everywhere, but the degree of instability is zero on average. It is also of great interest to develop a supremum version of the tests and compare the performance with those averaging statistics, which is also under investigation by the author.

Figures and Tables

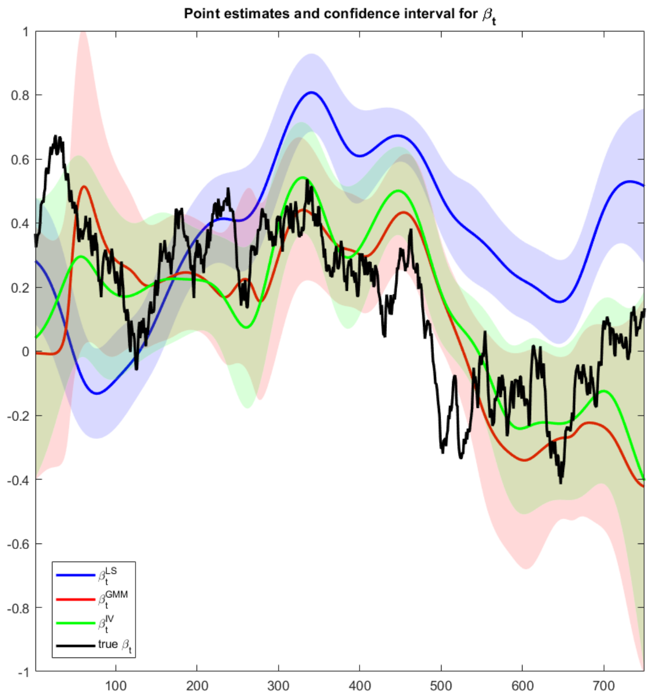

Figure 3.1: Realization of β_t, $\hat{\beta}_t$ and 95% confidence intervals for θ_t from OLS, linear IV and CU-GMM with a two-sided normal kernel and $H = T^{0.5}$ for $T = 500$. For TV CU-GMM estimator, Bartlett kernel is used to construct time-varying HAC estimator with bandwidth $L = H^{0.3}$. The solid black lines show the true realization of θ_t. The solid blue, green and line lines show the point estimates and the associated shaded areas show the 95% pointwise confidence intervals for OLS, linear IV and CU-GMM estimators, respectively.

		RMSFE			MAD			Coverage rate		
α	T	OLS	IV	GMM	OLS	IV	GMM	OLS	IV	GMM
0.4	750	0.289	0.157	0.220	0.248	0.122	0.150	0.36	0.86	0.87
	1000	0.287	0.150	0.217	0.245	0.116	0.143	0.34	0.86	0.88
	2000	0.303	0.141	0.209	0.257	0.107	0.131	0.29	0.87	0.89
	3000	0.309	0.137	0.196	0.262	0.102	0.124	0.26	0.88	0.90
0.5	750	0.283	0.140	0.178	0.245	0.110	0.130	0.28	0.80	0.82
	1000	0.289	0.133	0.169	0.250	0.105	0.122	0.25	0.80	0.82
	2000	0.299	0.120	0.150	0.257	0.093	0.106	0.20	0.80	0.83
	3000	0.309	0.114	0.147	0.265	0.088	0.101	0.18	0.81	0.84
0.7	750	0.291	0.179	0.183	0.251	0.145	0.147	0.17	0.44	0.52
	1000	0.291	0.174	0.178	0.250	0.141	0.142	0.16	0.42	0.50
	2000	0.291	0.163	0.162	0.249	0.132	0.130	0.12	0.36	0.46
	3000	0.300	0.156	0.157	0.258	0.125	0.125	0.10	0.35	0.44

Notes: The table reports RMSFE, MAD and 95% coverage rates from time-varying OLS, linear IV and CU-GMM estimators. Gray shading indicates the best performing estimator. Data generating process is from DGP 1. Details of DGP design, estimators and evaluation criteria are given in Section 4.

Table 3.1: Performance of estimators: RMSFE, MAD and 95% coverage rate

		DGP 1		DGP 2		DGP 3		DGP 4	
α	T	J-test	Q-test	J-test	Q-test	J-test	Q-test	J-test	Q-test
0.4	750	0.10	0.73	0.09	0.13	0.87	0.84	0.85	0.63
	1000	0.12	0.79	0.09	0.12	0.91	0.89	0.89	0.69
	2000	0.13	0.87	0.08	0.12	0.96	0.95	0.96	0.81
	3000	0.12	0.92	0.09	0.11	0.98	0.96	0.99	0.85
0.5	750	0.12	0.75	0.12	0.11	0.88	0.87	0.87	0.64
	1000	0.12	0.81	0.10	0.08	0.91	0.90	0.93	0.69
	2000	0.13	0.89	0.10	0.08	0.97	0.97	0.96	0.84
	3000	0.13	0.92	0.11	0.11	0.99	0.98	0.99	0.86
0.7	750	0.25	0.83	0.13	0.08	0.92	0.92	0.94	0.72
	1000	0.24	0.86	0.13	0.07	0.96	0.94	0.96	0.79
	2000	0.28	0.93	0.13	0.06	0.99	0.97	0.98	0.87
	3000	0.27	0.95	0.12	0.06	0.99	0.98	0.99	0.90

Notes: The table reports rejection frequencies from both J-test and Q-test under DGPS 1–4. Formulas of both tests are given in (3.14) and (3.12). Details of DGP designs are given in section 4.

Table 3.2: Rejection frequencies of structural stability tests

Sample period	J-test	Q-test
1/1964 - 12/2012	0.000	0.000
1/1964 - 12/2020	0.003	0.000

Notes: The table presents p-values from two structural stability test statistics presented in section 3. Test portfolios are joint cross section of size-sorted equities and treasuries. Pricing factors are market factor, Small Minus Big portfolio and 10 year treasury yield.

Table 3.3: Structural stability test: p-values

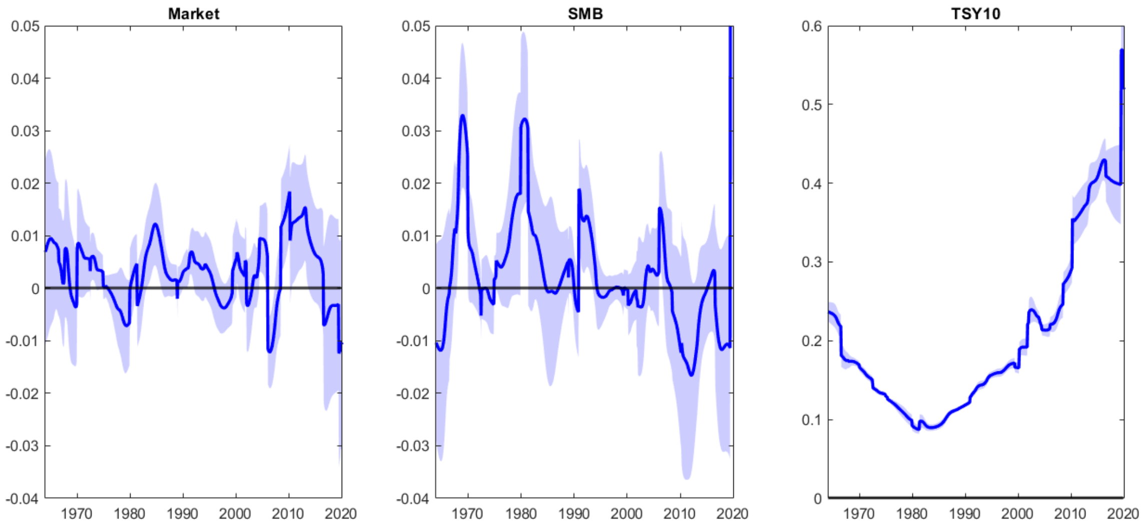

Figure 3.2: Time-varying price of risk estimates from 3 factor model. Testing portfolios are 10 size sorted equity portfolios and 7 constant maturity treasury portfolios.

CU-GMM		Linear GMM	
1/1964 - 12/2012	1/1964 - 12/2020	1/1964 - 12/2012	1/1964 - 12/2020
0.018	0.031	0.023	0.032
Constant		Affine	
1/1964 - 12/2012	1/1964 - 12/2020	1/1964 - 12/2012	1/1964 - 12/2020
0.094	0.104	2.338	0.834

Notes: The table presents Hansen-Jagannathan distance from both TVP models (estimated by different methods) and constant coefficient model. Test portfolios are joint cross section of size-sorted equities and treasuries. Pricing factors are market factor, Small Minus Big portfolio and 10 year treasury yield. Gray shading indicates the best performing estimator.

Table 3.4: Hansen-Jagannathan distance

Appendix A: Mathematical proofs

Appendix A.1: Proof of Lemma 1

By Triangular inequality,

$$\begin{aligned}\sup_{\theta_t\in\Theta}\Big\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_t)-E(g_t(\theta_t))\Big\| &\leqslant \sup_{\theta_t\in\Theta}\Big\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_t)-E\Big(\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_t)\Big)\Big\| \\ &\quad+\sup_{\theta_t\in\Theta}\Big\|E\Big(\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_t)\Big)-E(g_t(\theta_t))\Big\| \\ &=\sup_{\theta_t\in\Theta}M_{1,t}(\theta_t)+\sup_{\theta_t\in\Theta}M_{2,t}(\theta_t).\end{aligned}$$

We will show that

$$\sup_{\theta_t\in\Theta}M_{1,t}(\theta_t)=O_p(H^{-1/2}) \tag{3.17}$$

$$\sup_{\theta_t\in\Theta}M_{2,t}(\theta_t)=O\Big(\big(\frac{H}{T}\big)^{\bar{\gamma}}\Big). \tag{3.18}$$

These bounds prove the first part of the Lemma 1.
Proof of (3.17). Define $a_T=H^{-1/2}$, $\tau_T=(a_T)^{-\frac{1}{r-1}}$ where $2<r\leqslant 4$ and write

$$\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_t)=\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_t)\mathbb{1}\{g_j(\theta_t)\leqslant\tau_T\}+\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_t)\mathbb{1}\{g_j(\theta_t)>\tau_T\}$$

For the second term, notice that

$$\begin{aligned}E\Big\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_t)\mathbb{1}\{g_j(\theta_t)>\tau_T\}\Big\| &=\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}E\big\|g_j(\theta_t)\mathbb{1}\{g_j(\theta_t)>\tau_T\}\big\| \\ &=\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}E\Big\{\big\|g_j(\theta_t)\big\|^{r}\big\|g_j(\theta_t)\big\|^{1-r}\mathbb{1}\{g_j(\theta_t)>\tau_T\}\Big\} \\ &\leqslant\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}(\tau_T)^{-(r-1)}E\Big\{\big\|g_j(\theta_t)\big\|^{r}\mathbb{1}\{g_j(\theta_t)>\tau_T\}\Big\} \\ &\leqslant\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}(\tau_T)^{-(r-1)}E\big\|g_j(\theta_t)\big\|^{r} \\ &\leqslant\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}(\tau_T)^{-(r-1)}\cdot\sup_{j\geqslant 1}E\big\|g_j(\theta_t)\big\|^{r}=O(a_T)\end{aligned}$$

This implies that

$$\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t)\mathbb{1}\{g_j(\theta_t) > \tau_T\} = O_p(a_T) = o_p(1).$$

Then, we can focus on the first term. Let us move on to the first term. Write

$$\begin{aligned} Z_j &= b_{jt,H}g_j(\theta_t)\mathbb{1}\{g_j(\theta_t) \leqslant \tau_T\} - b_{jt,H}E[g_j(\theta_t)\mathbb{1}\{g_j(\theta_t) \leqslant \tau_T\}] \\ &= b_{jt,H}g_j^*(\theta_t) \end{aligned}$$

where $g_j^*(\theta_t) = g_j(\theta_t)\mathbb{1}\{g_j(\theta_t) \leqslant \tau_T\} - E[g_j(\theta_t)\mathbb{1}\{g_j(\theta_t) \leqslant \tau_T\}]$. By Lemma C2, $\forall t$, $\{g_j(\theta_t)\}$ is α-mixing with mixing coefficients satisfy $\alpha_m^g \leqslant Cm^{-\gamma}$. We will show that

$$E\left\|\frac{1}{K_t}\sum_{j=1}^{T} Z_j\right\| = o(1).$$

For notation simplicity, we assume that $k = q = 1$. The case $k > 1, q > 1$ reduces to the analysis of a finite number of similar sums of scalar variables.

Observe that, by Assumption 2.2(ii), $\left|g_j(\theta_t)\right| \leqslant 2B_T$ for some finite constant B_T which may depend on sample size and $g_j^*(\theta_t)$ have zero mean. Then, we can apply Theorem A.5 in Hall and Heyde [1980] to obtain

$$\left|E(g_k^*(\theta_t)g_j^*(\theta_t))\right| \leqslant 16B^2\alpha_{|k-j|}^g.$$

By Jensen's inequality,

$$\begin{aligned} \left(E\left|\frac{1}{K_t}\sum_{j=1}^{T} Z_j\right|\right)^2 &= \frac{1}{K_t^2}\sum_{k=1}^{T}\sum_{j=1}^{T} b_{kt,H}b_{jt,H}E\left|g_k^*(\theta_t)g_j^*(\theta_t)\right| \\ &\leqslant \frac{1}{K_t^2}\sum_{k=1}^{T}\sum_{j=1}^{T} b_{kt,H}b_{jt,H}\left|E[g_k^*(\theta_t)g_j^*(\theta_t)]\right| \\ &\leqslant \frac{16B^2}{K_t^2}\Big(\sum_{j=1}^{T} b_{jt,H}^2\alpha_0^g + \sum_{k,j}\sum_{k\neq j} b_{kt,H}b_{jt,H}\alpha_{|k-j|}^g\Big). \end{aligned} \tag{3.19}$$

Since $\alpha_0^g \leqslant \frac{1}{4}$ and we know from Lemma C1 that $K_t^2 = O(H^2)$, $\sum_{j=1}^{T} b_{jt,H}^2 = O(H)$, we have

$$\frac{16B^2}{K_t^2}\sum_{j=1}^{T} b_{jt,H}^2\alpha_0^g \leqslant \frac{16B^2}{4}\cdot\frac{1}{K_t^2}\sum_{j=1}^{T} b_{jt,H}^2 = O(H^{-1}) = o(1).$$

For the second term, observe that

$$\sum_{k,j}\sum_{k\neq j} b_{kt,H} b_{jt,H} \alpha^g_{T,|k-j|} = 2\sum_{j=1}^{T} b_{jt,H} \sum_{m=1}^{T-j} b_{j+m,t}\alpha^g_m,$$

and the fact that $K_t^2 = O(H^2)$, we have (ignore the constant term for simplicity):

$$\begin{aligned}
\left|\frac{1}{K_t^2}\sum_{j=1}^{T} b_{jt,H}\sum_{m=1}^{T-j} b_{j+m,t,H}\alpha^g_m\right| &= \left|\sum_{j=1}^{T}\frac{1}{H}b\Big(\frac{j-t}{H}\Big)\sum_{m=1}^{T-j}\frac{1}{H}b\Big(\frac{j+m-t}{H}\Big)\alpha^g_m\right| \\
&\leqslant \sum_{j=1}^{T}\frac{1}{H}b\Big(\frac{j-t}{H}\Big)\cdot\left|\sum_{m=1}^{T}\frac{1}{H}b\Big(\frac{j+m-t}{H}\Big)\alpha^g_m\right| \\
&\leqslant \sum_{j=1}^{T}\frac{1}{H}b\Big(\frac{j-t}{H}\Big)\cdot\sum_{m=1}^{T}\left|\frac{1}{H}b\Big(\frac{j+m-t}{H}\Big)\alpha^g_m\right| \\
&\leqslant C\cdot\sum_{j=1}^{T}\frac{1}{H}b\Big(\frac{j-t}{H}\Big)\cdot\sum_{m=1}^{T}\frac{1}{H}m^{-\gamma}b\Big(\frac{j+m-t}{H}\Big),
\end{aligned}$$

for some $0 < C, \gamma < \infty$ and the final inequality follows by plugging the fact that $\alpha^g_m \leqslant Cm^{-\gamma}$. Consider now the second sums:

$$\begin{aligned}
\left|\sum_{m=1}^{T}\frac{1}{H}m^{-\gamma}b\Big(\frac{j+m-t}{H}\Big)\right| &\leqslant \frac{1}{H^\gamma}\cdot\sum_{m=1}^{T}\frac{1}{H^{1-\gamma}}m^{-\gamma}\left|b\Big(\frac{j+m-t}{H}\Big)\right| \\
&\leqslant \frac{1}{H^\gamma}\cdot\lim_{T\to\infty}\frac{1}{H^{1-\gamma}}m^{-\gamma}\left|b\Big(\frac{j+m-t}{H}\Big)\right| \\
&= \frac{1}{H^\gamma}\cdot\left(\int_{-\infty}^{\infty}a^{-\gamma}b(a)da + o(1)\right) \\
&\leqslant \frac{1}{H^\gamma}\cdot\int_{-\infty}^{\infty}|a|^{-\gamma}\left|\bar{b}(a)\right|da = O(H^{-\gamma}) = o(1),
\end{aligned}$$

provided that $\int_{-\infty}^{\infty}|a|^{-\gamma}\left|\bar{b}(a)\right|da < \infty$, which is guaranteed by Assumption 2.5(iv) It follows from Lemma C1 that

$$\left|\sum_{j=1}^{T}\frac{1}{H}b\Big(\frac{j-t}{H}\Big)\right| \leqslant \int_{-\infty}^{\infty}b(a)da + o(1) = 1 + o(1).$$

By combining all above, we have that

$$\frac{1}{K_t^2}\sum_{k,j}\sum_{k\neq j} b_{kt,H} b_{jt,H}\alpha^g_{|k-j|} = o(1).$$

Continuing from (3.19), we see that

$$\Big(E\Big|\frac{1}{K_t}\sum_{j=1}^{T} Z_j\Big|\Big)^2 \leqslant \frac{16B^2}{K_t^2}\Big(\sum_{j=1}^{T} b_{jt,H}^2\alpha_0^g + \sum_{k,j}\sum_{k\neq j} b_{kt,H}b_{jt,H}\alpha_{|k-j|}^g\Big) = o(1),$$

since B is finite. Then, for any $M > 0$, by Markov inequality, we see that

$$\mathbb{P}\Big(\Big\|\frac{1}{K_t}\sum_{j=1}^{T} Z_{T,j}\Big\| > Ma_T\Big) \leqslant \frac{E\Big\|\frac{1}{K_t}\sum_{j=1}^{T} Z_{T,j}\Big\|^2}{(Ma_T)^2} = O(a_T) = o(1),$$

which implies that, for a given θ_t,

$$M_{1,t}(\theta_t) = O_p\big(H^{-1/2}\big).$$

Recall that, Θ is compact and by Assumption 2.4, $\|\theta_t\| = O_p(1)$. Consider the set $\{\theta_t : \|\theta_t\| \leqslant C\}$ for some very large constant C, an open cover $\cup_k A_k$ contains this set. By Heine–Borel theorem, every open cover of a compact set has a finite subcover. Then, $\exists M$, we have

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\theta_t\in\Theta}\Big\|\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t) - E[\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t)]\Big\| > Ma_T\Big)\\
&= \mathbb{P}\Big(\sup_{\|\theta_t\|\leqslant C}\Big\|\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t) - E[\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t)]\Big\| > Ma_T\Big)\\
&\leqslant \mathbb{P}\Big(\sup_{1\leqslant k\leqslant K_{max}}\sup_{\theta_t\in A_k}\Big\|\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t) - E[\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t)]\Big\| > Ma_T\Big)\\
&\leqslant \sum_{k=1}^{K_{max}}\mathbb{P}\Big(\sup_{\theta_t\in A_k}\Big\|\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t) - E[\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t)]\Big\| > Ma_T\Big) = o(1),
\end{aligned} \tag{3.20}$$

since K_{max} is finite. This completes the proof of (3.17).

Proof of (3.18). Observe that

$$\begin{aligned}
&\Big\|E\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t)\Big) - E(g_t(\theta_t))\Big\|\\
\leqslant\ &\Big\|E\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_t)\Big) - E\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_j)\Big)\Big\| + \Big\|E\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_j)\Big) - E(g_t(\theta_t))\Big\|\\
\leqslant\ &\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}E\left\|g_j(\theta_t) - g_j(\theta_j)\right\| + \frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}E\left\|g_j(\theta_j) - g_t(\theta_t)\right\|\\
=\ &M_{2,t,1} + M_{2,t,2}.
\end{aligned}$$

For $M_{2,t,1}$, by Taylor expansion for $g_j(\theta_t)$ and the Assumption 2.4, we have

$$\left\|M_{2,t,1}\right\| \leqslant \left\{\sup_j E\left\|\frac{\partial g_j(\bar{\theta}_t)}{\partial \theta'}\right\|_{sp}^2\right\}^{1/2} \frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\left\{E\left\|\theta_j - \theta_t\right\|^2\right\}^{1/2} = O_p\Big((\frac{H}{T})^{\bar{\gamma}}\Big),$$

where $\bar{\theta}_t$ lies between θ_t and θ_j. Similar arguments, together with Assumption 2.2(iii) yields

$$\left\|M_{2,t,2}\right\| = O\Big((\frac{H}{T})^{\bar{\gamma}}\Big).$$

Since the above holds uniformly in θ_t, this concludes the proof of both (3.18) and UWLLN.

We now move to CLT, which states that

$$\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt,H}\Big(g_j(\theta_t) - E(g_j(\theta_t))\Big) \xrightarrow{d} \mathcal{N}(0, W_t),$$

where

$$W_t(\theta_t) = \operatorname*{plim}_{T\to\infty} \operatorname{Var}\left(\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt,H} g_j(\theta_t)\right).$$

Define

$$z_j = b_{jt,H}\Big(g_j(\theta_t) - E(g_j(\theta_t))\Big).$$

The proof relies on checking the conditions (1)-(5) in Francq and Zakoïan [2005]. For the first two conditions, they require that

$$\sup_{1\leqslant t\leqslant T}\left\|z_j\right\|_{2+\nu^*} \leqslant \infty$$

for some $\nu^* \in (0, \infty]$ and W_t is positive definite. These are guaranteed by Assumption 2.2(ii) for $2 < \nu^* \leqslant 4$. Note that, Assumption 2.2(i) also implies that

$$\tau_T = O(T^{\kappa})$$

for some $\kappa \in [0, \frac{1}{6}]$. This verifies the third condition. By applying Lemma B2, we verify the forth condition:

$$\alpha_T^g(h) \leqslant \alpha^x(h - \tau_T),$$

for all $h > \tau_T$. For the fifth condition, notice that

$$\sum_{h=1}^{\infty} h^{r^*}\alpha^x(h)^{\nu^*/(\nu^*+2)} \leqslant \sum_{h=1}^{\infty} h^{r^* - \frac{\gamma\nu^*}{\nu^*+2}} < \infty$$

is satisfied if $r^* < \frac{\gamma\nu^*}{\nu^*+2}$, which is guaranteed by Assumption 2.2(i). This completes the proof.

We now prove the consistency of time-varying HAC estimator. Since both are evaluated at the true parameter vector $\theta_{0,t}$, we drop it for notation simplicity. We also assume that $\{g_t\}_t$ has been demeaned. Let $A^{(a,b)}$ be the (a,b)th element of matrix A, by Minkowski's inequality, for $2 < r \leqslant 4$,

$$\begin{aligned}\left\|\tilde{W}_t^{(a,b)} - W_t^{(a,b)}\right\|_{\frac{r}{2}} &= \left\|\tilde{W}_t^{(a,b)} - E(\tilde{W}_t^{(a,b)}) + E(\tilde{W}_t^{(a,b)}) - W_t^{(a,b)}\right\|_{\frac{r}{2}} \\ &\leqslant \left\|\tilde{W}_t^{(a,b)} - E(\tilde{W}_t^{(a,b)})\right\|_{\frac{r}{2}} + \left\|E(\tilde{W}_t^{(a,b)}) - W_t^{(a,b)}\right\|_{\frac{r}{2}} = W_{1,t} + W_{2,t}. \qquad (3.21)\end{aligned}$$

Then, the proof is completed if we could provide bounds for $W_{1,t}$ and $W_{2,t}$.

Bounds for $W_{1,t}$. Consider

$$\begin{aligned}\left\|\tilde{W}_t^{(a,b)} - E(\tilde{W}_t^{(a,b)})\right\|_{\frac{r}{2}} &\leqslant \max_{a,b}\left\|\frac{2}{K_{2,t}}\sum_{s=1}^{T}\left|k\Big(\frac{s}{L}\Big)\right|\left|\sum_{j=1}^{T-s} b_{jt,H}b_{j+s,t,H}\Big(g_j^a g_{j+s}^b - E(g_j^a g_{j+s}^b)\Big)\right|\right\|_{\frac{r}{2}} \\ &\leqslant 2\max_{a,b}\frac{1}{K_{2,t}}\sum_{s=1}^{T}\left|k\Big(\frac{s}{L}\Big)\right|\Big(\sum_{j=1}^{T-s} b_{jt,H}^{\frac{r}{2}}b_{j+s,t,H}^{\frac{r}{2}}\Big)^{\frac{2}{r}} \times 36\Big(\frac{r}{r-2}\Big)^{\frac{3}{2}}(A+2s)\sup_j\left\|g_j\right\|_p^2, \qquad (3.22)\end{aligned}$$

where the last inequality follows from Lemma B4 with $2 < r \leqslant 4$, $p > r$ and $A = 12\sum_{m=0}^{\infty}\alpha_m^{2(1/r-1/p)}$. Observe the fact that

$$\begin{aligned}\left|\frac{1}{L^2}\sum_{s=-T}^{T} sk\Big(\frac{s}{L}\Big)\right| &\leqslant \frac{1}{L^2}\sum_{s=-T}^{T}|s|\left|\bar{k}\Big(\frac{s}{L}\Big)\right| \\ &\leqslant \lim_{T\to\infty}\frac{1}{L^2}\sum_{s=-T}^{T}|s|\left|\bar{k}\Big(\frac{s}{L}\Big)\right| \\ &= \int_{-\infty}^{\infty}|x|\left|\bar{k}(x)\right|dx + o(1) = O(1),\end{aligned}$$

since by Assumption 2.6, $\int_{-\infty}^{\infty}|x|\left|\bar{k}(x)\right|dx < \infty$. Then, we continue from (3.22):

$$\begin{aligned}\left\|\tilde{W}_t^{(a,b)} - E(\tilde{W}_t^{(a,b)})\right\|_{\frac{r}{2}} &\leqslant C\int_{-\infty}^{\infty}|x|\left|\bar{k}(x)\right|dx\, 36A\Big(\frac{r}{r-2}\Big)^{\frac{3}{2}} L^2\frac{\Big(\sum_{j=1}^{T-s} b_{jt,H}^{\frac{r}{2}}b_{j+s,t,H}^{\frac{r}{2}}\Big)^{\frac{2}{r}}}{K_{2,t}} \\ &\leqslant C\int_{-\infty}^{\infty}|x|\left|\bar{k}(x)\right|dx\, 36A\Big(\frac{r}{r-2}\Big)^{\frac{3}{2}} L^2\frac{\Big(\sum_{j=1}^{T} b_{jt,H}^{\frac{r}{2}}b_{j+s,t,H}^{\frac{r}{2}}\Big)^{\frac{2}{r}}}{K_{2,t}}\end{aligned}$$

where $C = \sup_j\left\|g_j\right\|_p^2$. By Lemma B1, we know that $K_{2,t} = O(H)$ and $\Big(\sum_{j=1}^{T} b_{jt,H}^{\frac{r}{2}}b_{j+s,t,H}^{\frac{r}{2}}\Big)^{\frac{2}{r}} = O(H^{\frac{2}{r}})$. Clearly, consistency requires that

$$L = o(H^{\frac{1}{2}-\frac{1}{r}}),$$

which is guaranteed by Assumption 2.6.

Bounds for $W_{2,t}$. We proceed as in Lemma 6.6 in Gallant and White [1988]. Notice that,

$$W_t = \operatorname*{plim}_{T\to\infty} \mathrm{Var}\left(\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt,H} g_j\right) = \operatorname*{plim}_{T\to\infty} \frac{1}{K_{2,t}}\sum_{i=1}^{T}\sum_{j=1}^{T} b_{it,H} b_{jt,H} E\left(g_i g_j'\right).$$

Let $k_{\tau T} = k\left(\frac{\tau}{L}\right)$, since

$$\begin{aligned} W_t - \tilde{W}_t &= (1-k_{0T})\frac{1}{K_{2,t}}\sum_{j=1}^{T} b_{jt,H}^2 E(g_j g_j') + 2\sum_{s=1}^{L}(1-k_{sT})\frac{1}{K_{2,t}}\sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} E(g_j g_{j+s}') \\ &\quad + 2\sum_{s=L+1}^{T}\frac{1}{K_{2,t}}\sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} E(g_j g_{j+s}') \\ &= \Delta_{w1,t} + \Delta_{w2,t} + \Delta_{w3,t}, \end{aligned}$$

we need to show that $\Delta_{w1,t}, \Delta_{w2,t}, \Delta_{w3,t}$ vanish as $T \to \infty$.

Bounds for $\Delta_{w1,t}$. Notice that

$$\left\|\frac{1}{K_{2,t}}\sum_{j=1}^{T} b_{jt,H}^2 E(g_j g_j')\right\| \leqslant \frac{1}{K_{2,t}}\sum_{j=1}^{T} b_{jt,H}^2 \sup_j E\left\|g_j\right\|^2 = O(1),$$

and the fact that $k_{0T} = 1$. Then, $\Delta_{w1,t} \to 0$ as $T \to \infty$.

Bounds for $\Delta_{w2,t}$. Consider

$$\begin{aligned} \left\|\sum_{s=1}^{L}(1-k_{sT})\frac{1}{K_{2,t}}\sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} E(g_j g_{j+s}')\right\| &\leqslant \sum_{s=1}^{L}|1-k_{sT}|\frac{1}{K_{2,t}}\sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H}\left\|g_j g_{j+s}'\right\| \\ &\leqslant 4C^2\sum_{s=1}^{L}|1-k_{sT}|\frac{1}{K_{2,t}}\sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H}\alpha_T^g(s) \\ &\leqslant 4C^2\sum_{s=1}^{L}|1-k_{sT}| b^*\left(\frac{s}{H}\right)\alpha_T^g(s), \end{aligned}$$

where the second inequality follows from the mixing inequality in Theorem A.5 in Hall and Heyde [1980] and the final equality follows from Lemma B1(iii). Define

$$f_T(s) = 4C^2\,\mathbb{I}(s \leqslant L)|1-k_{sT}| b^*\left(\frac{s}{H}\right)\alpha_T^g(s)$$

and $\mu(s)$ be the counting measure, we apply the dominant convergence theorem to show that $\int_0^\infty f_T(s)d\mu(s) \to 0$. For each s, the requirement that $k_{sT} \to 1$ ensures that $f_T(s) \to 0$ as $T \to \infty$. Further, because $|k_{sT}| \leqslant c$ and $|f_T(S)| \leqslant \left|\bar{f}(s)\right|$, where $\bar{f}(s) = 4C^2(1+c)\bar{b}^*(\frac{s}{H})\alpha_T^g(s)$, since $\sup_s \alpha_T^g(s) \to 0$ as $s \to \infty$, and $\bar{b}^*(\frac{s}{H})$, which implies that $\bar{f}(s)$ is bounded and

$$\int_0^\infty \bar{f}(s)d\mu(s) = \sum_{s=1}^{\infty} 4C^2(1+c)\bar{b}^*(\frac{s}{H})\alpha_T^g(s) < \infty,$$

which further implies that $\int_0^\infty f_T(s)d\mu(s) \to 0$ and we conclude that $\Delta_{w2,t} \to 0$ as $T \to \infty$.

Bounds for $\Delta_{w3,t}$. Consider

$$\begin{aligned}\left\| \sum_{s=L+1}^{T} \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} E(g_j g'_{j+s}) \right\| &\leqslant \sum_{s=L+1}^{T} \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \left\| g_j g'_{j+s} \right\|_1 \\ &\leqslant 4C^2 \sum_{s=L+1}^{T} b^*(\frac{s}{H})\alpha_T^g(s) \\ &= 4C^2 \left(\sum_{s=1}^{T} b^*(\frac{s}{H})\alpha_T^g(s) - \sum_{s=1}^{L} b^*(\frac{s}{H})\alpha_T^g(s) \right).\end{aligned}$$

Because the requirements of $\alpha_T^g(s)$ ensure the convergence of the two sums in the parentheses to the same limits provided that $L \to \infty$ as $T \to \infty$. This guarantees that $\Delta_{w3,t} \to 0$ as $T \to \infty$.

Then, by combining all results above, we continue with (3.21):

$$\left\| \tilde{W}_t^{(a,b)} - W_t^{(a,b)} \right\|_{\frac{r}{2}} \leqslant W_{1,t} + W_{2,t} = o(1),$$

which completes the proof.

Appendix A.2: Proof of Theorem 1

The estimator is defined by (3.3) and (3.5):

$$\hat{\theta}_t = \arg\min_{\theta_t \in \Theta} Q_{t,T},$$

where the criteria function $Q_{t,T}$ is given by

$$Q_{t,T} = \overline{g}'_t(\theta_t)\tilde{W}_t^{-1}(\theta_t)\overline{g}_t(\theta_t).$$

To establish consistency and asymptotic normality of the estimator, we rely on Taylor series expansion. Write $\theta_{0,t}$ as the true value and consider a first-order Taylor series expansion of $\frac{\partial Q_{t,T}(\hat{\theta}_t)}{\partial \theta} = 0$ around $\theta_{0,t}$,

$$\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial \theta} + \frac{\partial^2 Q_{t,T}(\bar{\theta}_t)}{\partial \theta \partial \theta'}(\hat{\theta}_t - \theta_{0,t}) = 0,$$

where $\bar{\theta}_t$ lies between $\hat{\theta}_t$ and $\theta_{0,t}$. By rearranging terms, we have

$$\begin{aligned}\hat{\theta}_t - \theta_{0,t} &= -\Big(\frac{\partial^2 Q_{t,T}(\bar{\theta}_t)}{\partial\theta\partial\theta'}\Big)^{-1}\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta}\\ &= -\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big)^{-1}\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta} + \Bigg[\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big)^{-1} - \Big(\frac{\partial^2 Q_{t,T}(\bar{\theta}_t)}{\partial\theta\partial\theta'}\Big)^{-1}\Bigg]\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta}. \qquad (3.23)\end{aligned}$$

We need to show that

$$\Bigg\|\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big)^{-1} - \Big(\frac{\partial^2 Q_{t,T}(\bar{\theta}_t)}{\partial\theta\partial\theta'}\Big)^{-1}\Bigg\|_{sp} = o_p(1) \qquad (3.24)$$

$$\Bigg\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Bigg\|_{sp} = O_p(1), \qquad (3.25)$$

Then, consistency and asymptotic normality are obtained from the first term in (3.23). Thus, we need a detailed expansion for the first and second order derivatives for $Q_{t,T}(\theta_t)$.

Let us first compute the score:

$$\begin{aligned}\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta} &= 2\Bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial g_j(\theta_{0,t})}{\partial\theta'}\Bigg]'\tilde{W}_t^{-1}(\theta_{0,t})\Bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t})\Bigg]\\ &\quad + \Big(A_{2,1}(\theta_{0,t}),\cdots,A_{2,k}(\theta_{0,t})\Big)'\\ &= A_1(\theta_{0,t}) + A_2(\theta_{0,t}).\end{aligned}$$

The ℓ_1th elements in $A_2(\theta_{0,t})$ is given by

$$A_{2,\ell_1}(\theta_{0,t}) = \Bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t})\Bigg]'\frac{\partial\tilde{W}_t^{-1}(\theta_{0,t})}{\partial\theta_{\ell_1}}\Bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t})\Bigg],$$

where

$$\frac{\partial\tilde{W}_t^{-1}(\theta_{0,t})}{\partial\theta_{\ell_1}} = -\tilde{W}_t^{-1}(\theta_{0,t})\frac{\partial\tilde{W}_t(\theta_{0,t})}{\partial\theta_{\ell_1}}\tilde{W}_t^{-1}(\theta_{0,t}).$$

We need to show that

$$\Big\|A_1(\theta_{0,t})\Big\| = O_p\Big(\big(\frac{H}{T}\big)^{\bar{\gamma}} + H^{-1/2}\Big) \qquad (3.26)$$

$$\Big|A_{2,\ell_1}(\theta_{0,t})\Big| = o_p(1), \quad \text{for } \ell_1 = 1,2,\cdots,k. \qquad (3.27)$$

Proof of (3.27). First, notice that

$$\Big\|\tilde{W}_t^{-1}(\theta_{0,t}) - W_t^{-1}\Big\|_{sp} \leqslant \|W_t\|_{sp}^{-1}\Big\|W_t - \tilde{W}_t(\theta_{0,t})\Big\|_{sp}\Big\|\tilde{W}_t(\theta_{0,t})\Big\|_{sp}^{-1} = O_p(1),$$

by the fact that W_t is positive definite and Lemma 1(iii). Then, consider

$$\frac{\partial \tilde{W}_t(\theta_{0,t})}{\partial \theta_{\ell_1}} = \sum_{s=-T}^{T} k\Big(\frac{s}{L}\Big)\frac{1}{K_{2,t}}\bigg(\sum_{j=1}^{T-s} b_{jt,H}b_{j+s,t,H}g_j(\theta_{0,t})(\frac{\partial g_{j+s}(\theta_{0,t})}{\partial \theta_{\ell_1}})' + \sum_{j=1}^{T-s} b_{jt,H}b_{j+s,t,H}(\frac{\partial g_{j+s}(\theta_{0,t})}{\partial \theta_{\ell_1}})g'_{j+s}(\theta_{0,t})\bigg).$$

By similar arguments used in the proof of Lemma 1(iii), we can show that

$$\sum_{s=-T}^{T} k\Big(\frac{s}{L}\Big)\frac{1}{K_{2,t}}\bigg(\sum_{j=1}^{T-s} b_{jt,H}b_{j+s,t,H}g_j(\theta_{0,t})(\frac{\partial g_{j+s}(\theta_{0,t})}{\partial \theta_{\ell_1}})'\bigg) \xrightarrow{p} W_{t,d_1},$$

where

$$W_{t,d_1} = \operatorname*{plim}_{T\to\infty} \frac{1}{K_{2,t}}\sum_{i=1}^{T}\sum_{j=1}^{T} b_{it,H}b_{jt,H}E\Big[g_i(\theta_{0,t})(\frac{\partial g_j(\theta_{0,t})}{\partial \theta_{\ell_1}})\Big]$$

is positive definite. Then,

$$\bigg\|\frac{\partial \tilde{W}_t^{-1}(\theta_{0,t})}{\partial \theta_{\ell_1}}\bigg\|_{sp} \leqslant \|W_t\|_{sp}^{-2}\big\|W_{t,d_1}\big\| + o_p(1) = O_p(1).$$

By Lemma 1(i), since $E(g_t(\theta_{0,t})) = 0$, we have

$$\bigg\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t})\bigg\| = O_p\Big((\frac{H}{T})^{\bar{\gamma}} + H^{-1/2}\Big) = o_p(1),$$

which implies that

$$\big|A_{2,\ell_1}(\theta_{0,t})\big| \leqslant \bigg\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t})\bigg\|^2\bigg\|\frac{\partial \tilde{W}_t^{-1}(\theta_{0,t})}{\partial \theta_{\ell_1}}\bigg\|_{sp} + o_p(1) = o_p(1).$$

Proof of (3.26). First, by similar arguments as in the proof of Lemma 1(i), we can show that

$$\sup_{\theta_t\in\Theta}\bigg\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial g_j(\theta_t)}{\partial \theta'} - G_t(\theta_t)\bigg\|_{sp} = O_p\Big((\frac{H}{T})^{\bar{\gamma}} + H^{-1/2}\Big).$$

Define

$$G_{D,t} = \frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial g_j(\theta_t)}{\partial \theta'} - G_t$$
$$W_{D,t} = \tilde{W}_t^{-1}(\theta_{0,t}) - W_t^{-1},$$

and G_t, W_t are defined in Theorem 1. Let us rewrite $A_1(\theta_{0,t})$:

$$A_1(\theta_{0,t}) = G_t' W_t^{-1}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} g_j(\theta_{0,t})\Big) + G_{D,t}' W_{D,t}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} g_j(\theta_{0,t})\Big)$$
$$+ G_{D,t}' W_t^{-1}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} g_j(\theta_{0,t})\Big) + G_t' W_{D,t}\Big(\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} g_j(\theta_{0,t})\Big).$$

Then,

$$\big\|A_1(\theta_{0,t})\big\| \leqslant \|G_t\|_{sp}\|W_t\|_{sp}^{-1}\bigg\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} g_j(\theta_{0,t})\bigg\| + o_p(1) = O_p\Big((\frac{H}{T})^{\bar{\gamma}} + H^{-1/2}\Big).$$

Consider now the second order derivatives of the objective function:

$$\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'} = \begin{bmatrix} \frac{\partial A_1(\theta_{0,t})}{\partial\theta_1} & \cdots & \frac{\partial A_1(\theta_{0,t})}{\partial\theta_k} \end{bmatrix}_{k\times k} + \begin{bmatrix} \frac{\partial A_{2,1}(\theta_{0,t})}{\partial\theta'} \\ \vdots \\ \frac{\partial A_{2,k}(\theta_{0,t})}{\partial\theta'} \end{bmatrix}_{k\times k}$$
$$= B_1(\theta_{0,t}) + B_2(\theta_{0,t})$$

We will show that

$$\bigg\|\frac{\partial A_1(\theta_{0,t})}{\partial\theta_{\ell_2}}\bigg\| = O_p(1) \tag{3.28}$$
$$\bigg\|\frac{\partial A_{2,\ell_2}(\theta_{0,t})}{\partial\theta'}\bigg\| = o_p(1), \tag{3.29}$$

for $\ell_2 = 1, \cdots, k$.

Proof of (3.28). Consider

$$\bigg\|\frac{\partial A_1(\theta_{0,t})}{\partial\theta_{\ell_2}}\bigg\| = 2\bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\bigg]'\tilde{W}_t^{-1}(\theta_{0,t})\bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} g_j(\theta_{0,t})\bigg]$$
$$+ 2\bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\bigg]'\frac{\partial\tilde{W}_t^{-1}(\theta_{0,t})}{\partial\theta_{\ell_2}}\bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H} g_j(\theta_{0,t})\bigg]$$
$$+ 2\bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\bigg]'\tilde{W}_t^{-1}(\theta_{0,t})\bigg[\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\bigg]$$
$$= B_{11}(\theta_{0,t}) + B_{12}(\theta_{0,t}) + B_{13}(\theta_{0,t}).$$

We need to find bounds for the above three terms.

Bounds for $B_{11}(\theta_{0,t})$. Notice that, by similar arguments as used in Lemma 1(i), we obtain

$$\sup_{\theta_t\in\Theta}\Big\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}\frac{\partial^2 g_j(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}-E\Big[\frac{\partial^2 g_t(\theta_t)}{\partial\theta_{\ell_2}\partial\theta'}\Big]\Big\|_{sp}=O_p\Big((\frac{H}{T})^{\bar{\gamma}}+H^{-1/2}\Big).$$

Then,

$$\big\|B_{11}(\theta_{0,t})\big\|\leqslant\Big\|E\Big[\frac{\partial^2 g_t(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big]\Big\|_{sp}\|W_t\|_{sp}^{-1}\Big\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})\Big\|+o_p(1)=o_p(1).$$

Bounds for $B_{12}(\theta_{0,t})$. Again, by similar arguments as in Lemma 1(i), we obtain

$$\big\|B_{12}(\theta_{0,t})\big\|\leqslant\Big\|E\Big[\frac{\partial^2 g_t(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big]\Big\|_{sp}\Big\|\frac{\partial\tilde{W}_t^{-1}(\theta_{0,t})}{\partial\theta_{\ell_2}}\Big\|_{sp}\Big\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})\Big\|+o_p(1)=o_p(1).$$

Bounds for $B_{13}(\theta_{0,t})$:

$$\big\|B_{13}(\theta_{0,t})\big\|\leqslant\Big\|E\Big[\frac{\partial^2 g_t(\theta_{0,t})}{\partial\theta_{\ell_2}\partial\theta'}\Big]\Big\|_{sp}^2\|W_t\|_{sp}^{-1}+o_p(1)=O_p(1).$$

Summing up, by triangular inequality, we get:

$$\Big\|\frac{\partial A_1(\theta_{0,t})}{\partial\theta_{\ell_2}}\Big\|\leqslant\big\|B_{11}(\theta_{0,t})\big\|+\big\|B_{12}(\theta_{0,t})\big\|+\big\|B_{13}(\theta_{0,t})\big\|=O_p(1).$$

Proof of (3.29). Consider

$$\frac{\partial A_{2,\ell_2}(\theta_{0,t})}{\partial\theta'}=2\Big[\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})\Big]'\frac{\partial\tilde{W}_t^{-1}(\theta_{0,t})}{\partial\theta_{\ell_2}}\Big[\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}\frac{\partial g_j(\theta_{0,t})}{\partial\theta'}\Big]$$
$$+\Big[A_{2,1,1}(\theta_{0,t})\cdots A_{2,k,1}(\theta_{0,t})\Big]_{1\times k},$$

where a typical element $A_{2,\ell_4,1}(\theta_{0,t})$, $\ell_4=1,2,\cdots,k$ is given by

$$A_{2,\ell_4,1}(\theta_{0,t})=\Big[\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})\Big]'\frac{\partial^2\tilde{W}_t(\theta_{0,t})}{\partial\theta_{\ell_1}\partial\theta_{\ell_4}}\Big[\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})\Big].$$

Since both elements above involves $\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})$, which converges to 0 at the rate $(\frac{H}{T})^{\bar{\gamma}}+H^{-1/2}$. Similar arguments as above leads to (3.29), which concludes the claim. Again, by triangular inequality, we obtain

$$\Big\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big\|_{sp}\leqslant\big\|B_1(\theta_{0,t})\big\|_{sp}+\big\|B_2(\theta_{0,t})\big\|_{sp}=O_p(1)$$

We have established (3.25), but we still need to show (3.24). Notice that

$$\Big\|\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big)^{-1}-\Big(\frac{\partial^2 Q_{t,T}(\bar\theta_t)}{\partial\theta\partial\theta'}\Big)^{-1}\Big\|_{sp}\leqslant\Big\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big\|_{sp}^{-1}\Big\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}-\frac{\partial^2 Q_{t,T}(\bar\theta_t)}{\partial\theta\partial\theta'}\Big\|_{sp}\Big\|\frac{\partial^2 Q_{t,T}(\bar\theta_t)}{\partial\theta\partial\theta'}\Big\|_{sp}^{-1}.$$

Then, we need to show that

$$\Big\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}-\frac{\partial^2 Q_{t,T}(\bar\theta_t)}{\partial\theta\partial\theta'}\Big\|_{sp}=o_p(1),\quad\Big\|\frac{\partial^2 Q_{t,T}(\bar\theta_t)}{\partial\theta\partial\theta'}\Big\|_{sp}^{-1}=O_p(1).$$

These are straightforward, since we could let $\bar\theta_t\xrightarrow{p}\theta_{0,t}$. For example, by first order Taylor expansion,

$$\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\bar\theta_t)=\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})+\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}\frac{\partial g_j(\tilde\theta_t)}{\partial\theta'}(\bar\theta_t-\theta_{0,t}).$$

Clearly, by the fact that $\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}\frac{\partial g_j(\tilde\theta_t)}{\partial\theta'}$ is bounded, letting $\bar\theta_t\xrightarrow{p}\theta_{0,t}$ we conclude that

$$\Big\|\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\bar\theta_t)-\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})\Big\|=o_p(1)$$

It follows similar, but lengthy arguments, we obtain same results for other terms involving derivatives.

By continuing from (3.23), we obtain the consistency results:

$$\begin{aligned}\|\hat\theta_t-\theta_{0,t}\|&\leqslant\Big\|\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big\|_{sp}^{-1}\Big\|\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta}\Big\|+\Big\|\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta\partial\theta'}\Big)^{-1}-\Big(\frac{\partial^2 Q_{t,T}(\bar\theta_t)}{\partial\theta\partial\theta'}\Big)^{-1}\Big\|_{sp}\Big\|\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta}\Big\|\\&=O_p\Big((\frac{H}{T})^{\bar\gamma}+H^{-1/2}\Big)\end{aligned}$$

For CLT, we assume that $q=k=1$, since for the case of $q,k>1$ it follows immediately from Cramer-Wold device theorem. Clearly, from (3.23), we can rewrite the estimator as

$$\begin{aligned}\hat\theta_t-\theta_{0,t}&=-\Big(\frac{\partial^2 Q_{t,T}(\theta_{0,t})}{\partial\theta^2}\Big)^{-1}\frac{\partial Q_{t,T}(\theta_{0,t})}{\partial\theta}+o_p(1)\\&=-(G_t'W_t^{-1}G_t)^{-1}G_t'W_t^{-1}\frac{1}{K_t}\sum_{j=1}^{T}b_{jt,H}g_j(\theta_{0,t})+o_p(1).\end{aligned}$$

By Lemma 1(ii), we have

$$\frac{K_t}{K_{2,t}^{1/2}}(\hat{\theta}_t - \theta_{0,t}) = -(G_t'W_t^{-1}G_t)^{-1}G_t'W_t^{-1}\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t}) + o_p(1)$$
$$\xrightarrow{d} -(G_t'W_t^{-1}G_t)^{-1}G_t'W_t^{-1}W_t^{1/2}\mathcal{N}(0,1) + o_p(1),$$

since

$$\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t}) \xrightarrow{d} \mathcal{N}(0,1).$$

Then, for the general case $q, k > 1$, we have

$$\frac{K_t}{K_{2,t}^{1/2}}(G_t'W_t^{-1}G_t)^{1/2}(\hat{\theta}_t - \theta_{0,t}) \xrightarrow{d} N(0, I_k),$$

which completes the proof.

Appendix A.3: Proof of Corollary 1

By triangular inequality,

$$\left\|\hat{G}_t - G_t\right\|_{sp} \leqslant \left\|\hat{G}_t - \frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial g_j(\theta_{0,t})}{\partial\theta'}\right\|_{sp} + \left\|\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\frac{\partial g_j(\theta_{0,t})}{\partial\theta'} - G_t\right\|_{sp}$$
$$= G_{t,1} + G_{t,2}.$$

Again, we have stated that by similar arguments used in the proof of (3.18), $G_{t,2} = o_p(1)$. For $G_{t,1}$, notice that

$$G_{t,1} \leqslant \frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}\left\|\frac{\partial g_j(\hat{\theta}_t)}{\partial\theta'} - \frac{\partial g_j(\theta_{0,t})}{\partial\theta'}\right\|_{sp} \leqslant \sum_{j=1}^{T} b_{jt,H}\left\|\frac{\partial^2 g_j(\bar{\theta}_t)}{\partial\theta_\ell\partial\theta'}\right\|_{sp}\left\|\hat{\theta}_t - \theta_{0,t}\right\|,$$

where the second inequality follows from first order Taylor expansion of $\frac{\partial g_j(\hat{\theta}_t)}{\partial\theta'}$ and it holds uniformly for all θ_ℓ, $\ell = 1, 2, \cdots, k$. Since $\left\|\frac{\partial^2 g_j(\bar{\theta}_t)}{\partial\theta_\ell\partial\theta'}\right\|_{sp} < \infty$ and by consistency we have that $\left\|\hat{\theta}_t - \theta_{0,t}\right\|_{sp} = o_p(1)$, this completes the proof.

Notice that

$$\hat{W}_t = \sum_{s=-T}^{T} k\Big(\frac{j}{L}\Big)\frac{1}{K_{2,t}}\sum_{j=1}^{T-s} b_{jt,H}b_{j+s,t,H}\Big(g_j^*(\hat{\theta}_t)\Big)\Big(g_{j+s}^*(\hat{\theta}_t)\Big)',$$

where $g_{j+s}^*(\hat{\theta}_t) = g_{j+s}'(\hat{\theta}_t) - \frac{1}{T}\sum_{s=1}^{T} g_s(\hat{\theta}_t)$. By triangular inequality,

$$\left\|\hat{W}_t - W_t\right\|_{sp} \leqslant \left\|\hat{W}_t - \tilde{W}_t\right\|_{sp} + \left\|\tilde{W}_t - W_t\right\|_{sp} = W_{e,t,1} + W_{e,t,2}.$$

In Lemma 1(iii), we have shown that $W_{e,t,2} = o_p(1)$. What remains is to show that $W_{e,t,1} = o_p(1)$. Consider

$$
\begin{aligned}
&\frac{K_t}{K_{2,t}^{1/2} L} \Big\| \sum_{s=-T}^{T} k\Big(\frac{j}{L}\Big) \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \Big(g_j^*(\hat{\theta}_t)\Big)\Big(g_{j+s}^*(\hat{\theta}_t)\Big)' - \sum_{s=-T}^{T} k\Big(\frac{j}{L}\Big) \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \Big(g_j^*(\theta_{0,t})\Big)\Big(g_{j+s}^*(\theta_{0,t})\Big)' \Big\|_{sp} \\
&\leqslant \sum_{s=-T}^{T} \frac{1}{L} k\Big(\frac{j}{L}\Big) \frac{K_t}{K_{2,t}^{1/2}} \Big\| \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \Big(g_j^*(\hat{\theta}_t)\Big)\Big(g_{j+s}^*(\hat{\theta}_t)\Big)' - \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \Big(g_j^*(\theta_{0,t})\Big)\Big(g_{j+s}^*(\theta_{0,t})\Big)' \Big\|_{sp} \\
&\leqslant \sum_{s=-T}^{T} \frac{1}{L} k\Big(\frac{j}{L}\Big) \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \frac{K_t}{K_{2,t}^{1/2}} \Big\| \Big(g_j^*(\hat{\theta}_t)\Big)\Big(g_{j+s}^*(\hat{\theta}_t)\Big)' - \Big(g_j^*(\theta_{0,t})\Big)\Big(g_{j+s}^*(\theta_{0,t})\Big)' \Big\|_{sp} \\
&\leqslant \sum_{s=-T}^{T} \frac{1}{L} k\Big(\frac{j}{L}\Big) \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \Big\| \frac{\partial g_j^*(\bar{\theta}_t)}{\partial \theta_\ell} \Big(g_{j+s}^*(\bar{\theta}_t)\Big)' + g_{j+s}^*(\bar{\theta}_t) \Big(\frac{\partial g_j^*(\bar{\theta}_t)}{\partial \theta_\ell}\Big)' \Big\|_{sp} \frac{K_t}{K_{2,t}^{1/2}} \|\hat{\theta}_t - \theta_{0,t}\|, \qquad (3.30)
\end{aligned}
$$

where the last line follows from a first order Taylor expansion of $\Big(g_j^*(\hat{\theta}_t)\Big)\Big(g_{j+s}^*(\hat{\theta}_t)\Big)'$ around $\theta_{0,t}$. Observe that

$$
\begin{aligned}
&\lim_{T\to\infty} \sum_{s=-T}^{T} \frac{1}{L} k\Big(\frac{j}{L}\Big) = \int_{-\infty}^{\infty} k(a)da < \infty \\
&\lim_{T\to\infty} \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} = \frac{1}{K_{2,t}} \int_{-\infty}^{\infty} b(c)b(c+d)dc < \infty \\
&\Big\| \frac{\partial g_j^*(\bar{\theta}_t)}{\partial \theta_\ell} \Big(g_{j+s}^*(\bar{\theta}_t)\Big)' \Big\|_{sp} < \Big\{ \Big\| \frac{\partial g_j^*(\bar{\theta}_t)}{\partial \theta_\ell} \Big\|^2 \Big\}^{1/2} \Big\{ \|g_{j+s}^*(\bar{\theta}_t)\|^2 \Big\}^{1/2} < \infty \\
&\frac{K_t}{K_{2,t}^{1/2}} \|\hat{\theta}_t - \theta_{0,t}\| = O_p(1).
\end{aligned}
$$

Then, continuing from (3.30), we have

$$
\frac{K_t}{K_{2,t}^{1/2} L} \Big\| \sum_{s=-T}^{T} k\Big(\frac{j}{L}\Big) \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \Big(g_j^*(\hat{\theta}_t)\Big)\Big(g_{j+s}^*(\hat{\theta}_t)\Big)' - \sum_{s=-T}^{T} k\Big(\frac{j}{L}\Big) \frac{1}{K_{2,t}} \sum_{j=1}^{T-s} b_{jt,H} b_{j+s,t,H} \Big(g_j^*(\theta_{0,t})\Big)\Big(g_{j+s}^*(\theta_{0,t})\Big)' \Big\|_{sp} = O_p
$$

which implies that

$$
W_{e,t,2} = O_p\Big(L \frac{K_{2,t}^{1/2}}{K_t}\Big) = o_p(1).
$$

Appendix A.4: Proof of Theorem 2

To derive the limiting distribution of J_T, consider again the decomposition of $V_{T,t}$:

$$
\begin{aligned}
V_{T,t} &= \hat{W}_t^{-1/2}\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt,H}g_j(\hat{\theta}_t)\\
&= W_t^{-1/2}\frac{1}{K_{2,t}^{1/2}}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t}) + W_t^{-1/2}G_t\frac{K_t}{K_{2,t}^{1/2}}\Big(-(G_t'W_t^{-1}G_t)^{-1}G_t'W_t^{-1}\frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t})\Big) + o_p(1)\\
&= W_t^{-1/2}\frac{1}{K_{2,t}^{1/2}}\Big(I_q - G_t(G_t'W_t^{-1}G_t)^{-1}G_t'W_t^{-1}\Big)\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,t})\\
&= W_t^{-1/2}\frac{1}{K_{2,t}^{1/2}}\Big(I_q - G_t(G_t'W_t^{-1}G_t)^{-1}G_t'W_t^{-1}\Big)\sum_{j=1}^{T} b_{jt,H}\Big(g_j(\theta_{0,j}) + \frac{\partial g_j(\bar{\theta}_t)}{\partial\theta'}(\theta_{0,t}-\theta_{0,j})\Big) + o_p(1)\\
&= W_t^{-1/2}\frac{1}{K_{2,t}^{1/2}}\Big(I_q - G_t(G_t'W_t^{-1}G_t)^{-1}G_t'W_t^{-1}\Big)\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,j}) + o_p(1),
\end{aligned}
$$

where the final line follows from the fact that, $\forall j$,

$$
\left\|g_j(\theta_{0,t}) - g_j(\theta_{0,j})\right\| \leqslant \left\|\frac{\partial g_j(\bar{\theta}_t)}{\partial\theta'}\right\|_{sp}\left\|\theta_{0,t}-\theta_{0,j}\right\| = O_p((\frac{H}{T})^{\bar{\gamma}}) = o_p(1).
$$

Then, consider the scaled sum of $V_{T,t}$:

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_{T,t} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Bigg(W_t^{-1/2}\frac{1}{K_{2,t}^{1/2}}\Big(I_q - G_t(G_t'W_t^{-1}G_t)^{-1}G_t'W_t^{-1}\Big)\sum_{j=1}^{T} b_{jt,H}g_j(\theta_{0,j})\Bigg) + o_p(1)\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Bigg(\sum_{j=1}^{T}\frac{b_{tj,H}}{K_{2,t}^{1/2}}W_j^{-1/2}(I_q - G_j(G_j'W_j^{-1}G_j)^{-1}G_j'W_j^{-1})\Bigg)g_t(\theta_{0,t}) + o_p(1)\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Xi_t g_t(\theta_{0,t}) + o_p(1),
\end{aligned}
$$

where $\Xi_t = \sum_{j=1}^{T}\frac{b_{tj,H}}{K_{2,t}^{1/2}}W_j^{-1/2}(I_q - G_j(G_j'W_j^{-1}G_j)^{-1}G_j'W_j^{-1})\big)$. In view of Cramer-Wold device, it is sufficient to show that, for $b \in \mathbb{R}^q$, where $\|b\| = 1$, the following holds

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T} b'\Xi_t g_t(\theta_{0,t}) \xrightarrow{d} \mathcal{N}(0, b'\Omega_1 b),
$$

where

$$
\Omega_2 = \operatorname*{plim}_{T\to\infty}\ \mathrm{Var}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Xi_t g_t(\theta_{0,t})\Big) < \infty
$$

is positive definite. It can be easily shown that, $\hat{\Omega}_2$ is given by

$$\hat{\Omega}_2 = \frac{1}{T}\sum_{t=1}^{T}\Big(\big(\sum_{j=1}^{T}\frac{b_{tj,H}}{K_{2,t}^{1/2}}\hat{W}_j^{-1/2}(I_q - \hat{G}_j(\hat{G}_j'\hat{W}_j^{-1}\hat{G}_j)^{-1}\hat{G}_j'\hat{W}_j^{-1})\big)g_t(\hat{\theta}_t)g_t'(\hat{\theta}_t)\big(\sum_{j=1}^{T}\frac{b_{tj,H}}{K_{2,t}^{1/2}}\hat{W}_j^{-1/2}(I_q - \hat{G}_j(\hat{G}_j'\hat{W}_j^{-1}\hat{G}_j)^{-1}\hat{G}_j'\hat{W}_j^{-1})\big)'$$

By Theorem 3.2 in Hall and Heyde [1980], we need to verify the following two conditions:

$$\frac{1}{T}\sum_{t=1}^{T} b'\Xi_t g_t(\theta_{0,t})g_t'(\theta_{0,t})\Xi_t' b \xrightarrow{p} b'\Omega_2 b \tag{3.31}$$

$$\max_{1\leqslant t\leqslant T}\left|\frac{1}{\sqrt{T}}b'\Xi_t g_t(\theta_{0,t})\right| \xrightarrow{p} 0. \tag{3.32}$$

Proof of (3.31). Observe that

$$E\Big[\frac{1}{T}\sum_{t=1}^{T} b'\Xi_t g_t(\theta_{0,t})g_t'(\theta_{0,t})\Xi_t' b\Big] = b'\frac{1}{T}\sum_{t=1}^{T}\Xi_t E(g_t(\theta_{0,t})g_t'(\theta_{0,t}))\Xi_t' b \longrightarrow b'\Omega_1 b$$

and

$$\mathrm{Var}\Big(\frac{1}{T}\sum_{t=1}^{T} b'\Xi_t g_t(\theta_{0,t})g_t'(\theta_{0,t})\Xi_t' b\Big) = \frac{1}{T^2}\sum_{t=1}^{T}\Big\{\Big[(\textstyle\sum_{s=1}^{q} b_s\xi_{s1})^2 \ \cdots \ (\sum_{s=1}^{q} b_s\xi_{sq})^2\Big]\begin{bmatrix}\mathrm{Var}(g_{t1}^2)\\ \vdots \\ \mathrm{Var}(g_{tq}^2)\end{bmatrix}$$
$$+ 2\sum_{j_1,j_2,j_1\neq j_2}(\sum_{s=1}^{q} b_s\xi_{sj_1})^2(\sum_{s=1}^{q} b_s\xi_{sj_2})^2\mathrm{Var}(g_{tj_1}g_{tj_2})\Big\}.$$

Since q is finite, $\forall j$, $I_q - G_j(G_j'W_j^{-1}G_j)^{-1}G_j'W_j^{-1}$ is idempotent and W_t is positive definite, we have $(\sum_{s=1}^{q} b_s\xi_{sj})^2 \leqslant \sum_{s=1}^{q} b_s^2\xi_{sj}^2 < \infty$. In addition, since $\sup_{1\leqslant t\leqslant T}\left\|\mathrm{Var}(g_t(\theta_{0,t})g_t'(\theta_{0,t}))\right\| < \infty$, we have, $\forall (j_1, j_2)$, $\sup_{1\leqslant t\leqslant T}\mathrm{Var}(g_{tj_1}g_{tj_2}') < \infty$. This implies that

$$\Big\{\Big[(\textstyle\sum_{s=1}^{q} b_s\xi_{s1})^2 \ \cdots \ (\sum_{s=1}^{q} b_s\xi_{sq})^2\Big]\begin{bmatrix}\mathrm{Var}(g_{t1}^2)\\ \vdots \\ \mathrm{Var}(g_{tq}^2)\end{bmatrix} + 2\sum_{j_1,j_2,j_1\neq j_2}(\sum_{s=1}^{q} b_s\xi_{sj_1})^2(\sum_{s=1}^{q} b_s\xi_{sj_2})^2\mathrm{Var}(g_{tj_1}g_{tj_2})\Big\} = O(1),$$

which further implies that

$$\mathrm{Var}\Big(\frac{1}{T}\sum_{t=1}^{T} b'\Xi_t g_t(\theta_{0,t})g_t'(\theta_{0,t})\Xi_t' b\Big) = O(\frac{1}{T}) = o(1).$$

(3.31) follows from Markov's inequality.

Proof of (3.32). Observe that, for any $\epsilon > 0$,

$$E\Big[\frac{1}{T}b'\Xi_t g_t(\theta_{0,t})g_t'(\theta_{0,t})\Xi_t' b\mathbb{I}(\big\|g_t(\theta_{0,t})\big\| > \sqrt{T}\epsilon)\Big] = \frac{1}{T}b'\Xi_t E\Big[g_t(\theta_{0,t})g_t'(\theta_{0,t})\mathbb{I}(\big\|g_t(\theta_{0,t})\big\| > \sqrt{T}\epsilon)\Big]\Xi_t' b.$$

By Theorem 12.10 in Davidson [1994], since $E\Big[g_t(\theta_{0,t})g_t'(\theta_{0,t})\Big] < \infty$, we have

$$E\Big[g_t(\theta_{0,t})g_t'(\theta_{0,t})\mathbb{I}(\big\|g_t(\theta_{0,t})\big\| > \sqrt{T}\epsilon)\Big] \longrightarrow 0.$$

Thus,

$$\mathbb{P}\Big(\max_{1\leqslant t\leqslant T}\Big|\frac{1}{\sqrt{T}}b'\Xi_t g_t(\theta_{0,t})\Big| > \epsilon\Big) \leqslant \frac{1}{\epsilon^2}\sum_{t=1}^{T} E\Big[\frac{1}{T}b'\Xi_t g_t(\theta_{0,t})g_t'(\theta_{0,t})\Xi_t' b\mathbb{I}(\big\|g_t(\theta_{0,t})\big\| > \sqrt{T}\epsilon)\Big] = o(1),$$

which completes the proof.

To derive the limiting distribution of Q_T, first order Taylor series expansion gives

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}_t) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\theta_0) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial g_t(\bar{\theta}_1)}{\partial\theta'}(\hat{\theta}_t - \theta)$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\theta_0) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial g_t(\bar{\theta}_2)}{\partial\theta'}(\hat{\theta} - \theta),$$

where $\bar{\theta}_1$ lies between $\hat{\theta}_t$ and θ_0, $\bar{\theta}_2$ lies between $\hat{\theta}$ and θ_0. Following similar procedure as in (3.23), we could expand both $\hat{\theta}_t - \theta$ and $\hat{\theta} - \theta$:

$$\hat{\theta}_t - \theta = -(\bar{G}_t'\bar{W}_t^{-1}\bar{G}_t)^{-1}\bar{G}_t'\bar{W}_t^{-1} \times \frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta)$$
$$\hat{\theta} - \theta = -(G_T'W_T^{-1}G_T)^{-1}G_T'W_T^{-1} \times \frac{1}{T}\sum_{j=1}^{T} g_j(\theta),$$

where $\bar{G}_t$ and $\bar{W}_t$ are defined similar as in Theorem 1, except for the fact that these quantities are evaluated at θ for all t, instead of $\theta_{0,t}$ for each t. G_T and W_T are standard score and covariance matrix in the case of fixed coefficients. Notice that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial g_t(\bar{\theta}_2)}{\partial\theta'}(\hat{\theta} - \theta) = \Big(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g_t(\bar{\theta}_2)}{\partial\theta'}\Big)\times\sqrt{T}(\hat{\theta} - \theta) = -G_T(G_T'W_T^{-1}G_T)^{-1}G_T'W_T^{-1} \times \frac{1}{T}\sum_{j=1}^{T} g_j(\theta) + o_p(1).$$

Then, we can write

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}_t) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \frac{\partial g_t(\bar{\theta}_1)}{\partial \theta'}\Big(-(\bar{G}_t'\bar{W}_t^{-1}\bar{G}_t)^{-1}\bar{G}_t'\bar{W}_t^{-1} \times \frac{1}{K_t}\sum_{j=1}^{T} b_{jt,H}g_j(\theta)\Big)$$
$$+ G_T(G_T'W_T^{-1}G_T)^{-1}G_T'W_T^{-1} \times \frac{1}{T}\sum_{j=1}^{T} g_j(\theta) + o_p(1)$$
$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Bigg(G_T(G_T'W_T^{-1}G_T)^{-1}G_T'W_T^{-1} - \sum_{j=1}^{T}(\frac{b_{tj,H}}{K_j})\frac{\partial g_j(\theta)}{\partial \theta'}(\bar{G}_j'\bar{W}_j^{-1}\bar{G}_j)^{-1}\bar{G}_j'\bar{W}_j^{-1}\Bigg)$$
$$g_t(\theta) + o_p(1)$$
$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Xi_{t,3}g_t(\theta) + o_p(1).$$

By similar procedures as used in deriving limiting distribution of J_T, we could show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}_t) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t(\hat{\theta}) \xrightarrow{d} \mathcal{N}(0, \Omega_1),$$

which completes the proof. Ω_1 is given by

$$\Omega_1 = \operatorname*{plim}_{T\to\infty} \operatorname{Var}\Bigg(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(G_T(G_T'W_T^{-1}G_T)^{-1}G_t'W_t^{-1} - \sum_{j=1}^{T}\frac{b_{tj,H}}{Kt}\frac{\partial g_j(\theta)}{\partial \theta'}(\bar{G}_j'\bar{W}_j^{-1}\bar{G}_j)^{-1}(\bar{G}_j'\bar{W}_j^{-1})\Big)g_t(\theta)\Bigg). \quad (3.33)$$

Both $\bar{G}_t$ and $\bar{W}_t$ are defined similar as in Theorem 1, except for the fact that these quantities are evaluated at θ for all t, instead of $\theta_{0,t}$ for each t. G_T and W_T are similarly defined as in the fixed coefficient case. It can be easily shown that, $\hat{\Omega}_1$ is given by

$$\hat{\Omega}_1 = \frac{1}{T}\sum_{t=1}^{T}\Bigg(\hat{G}_T(\hat{G}_T'\hat{W}_T^{-1}\hat{G}_T)^{-1}\hat{G}_T'\hat{W}_t^{-1} - \Big(\sum_{j=1}^{T}(\frac{b_{tj,H}}{K_t})\hat{\bar{G}}_j(\hat{\bar{G}}_j'\hat{\bar{W}}_j^{-1}\hat{\bar{G}}_j)^{-1}(\hat{\bar{G}}_j'\hat{\bar{W}}_j^{-1})\Big)$$
$$g_t(\hat{\theta})g_t'(\hat{\theta})\Big(\hat{G}_T(\hat{G}_T'\hat{W}_T^{-1}\hat{G}_T)^{-1}\hat{G}_T'\hat{W}_t^{-1} - \Big(\sum_{j=1}^{T}(\frac{b_{tj,H}}{K_t})\hat{\bar{G}}_j(\hat{\bar{G}}_j'\hat{\bar{W}}_j^{-1}\hat{\bar{G}}_j)^{-1}\Big)'\Bigg),$$

where $\hat{\bar{G}}_j$ and $\hat{\bar{W}}_j^{-1}$ are defined similarly as in Corollary 1, except that $\hat{\theta}_t$ is replaced by $\hat{\theta}$. $\hat{G}_T$ and $\hat{W}_T$ are obtained from standard constant coefficient CU-GMM estimation.

Appendix B: Auxiliary results

Lemma B1. *Consider the class of kernel weights $b_{jt,H} = b\Big(\frac{j-t}{H}\Big)$ computed with a kernel function $b(\cdot)$ and bandwidth H, both satisfying Assumption 2.5, we have that*

(i)

$$\sum_{j=1}^{T} b_{jt,H} = O(H), \quad \sum_{j=1}^{T} b_{tj,H} = O(H), \quad \sum_{j=1}^{T} b^{k}_{jt,H} = O(H), 1 < k \leqslant 2;$$

(ii)

$$\Big(\sum_{j=1}^{T-s} b^{\frac{r}{2}}_{jt,H} b^{\frac{r}{2}}_{j+s,t,H}\Big)^{\frac{2}{r}} = O(H^{\frac{2}{r}})$$

for some $2 < r \leqslant 4$, uniformly in s;

(iii)

$$\frac{\sum_{j=1}^{T} b\Big(\frac{j-t}{H}b\Big(\frac{j+s-t}{H}\Big)}{\sum_{j=1}^{T} b^2\Big(\frac{j-t}{H}\Big)} = b^*\Big(\frac{s}{H}\Big) + o(1),$$

uniformly in s, $b^*(\cdot)$ is the induced kernel $b^*(d) = \frac{1}{b_2}\int_{-\infty}^{\infty} b(c)b(c+d)dc$ and $b_2 = \int_{-\infty}^{\infty} b^2(a)da$.

Proof. Proof of (i). Notice that

$$\frac{1}{H}\sum_{j=1}^{T} b_{jt,H} = \frac{1}{H}\sum_{j=1}^{T} b\Big(\frac{j-t}{H}\Big) = \frac{1}{H}\sum_{s=1-t}^{T-t} b\Big(\frac{s}{H}\Big).$$

Observe that

$$\left|\frac{1}{H}\sum_{s=1-t}^{T-t} b\Big(\frac{s}{H}\Big)\right| \leqslant \frac{1}{H}\sum_{s=1-T}^{T-1}\left|b\Big(\frac{s}{H}\Big)\right|$$

Using the change of variable $s = \lfloor Ha \rfloor$, we have

$$\begin{aligned}
\frac{1}{H}\sum_{s=1-T}^{T-1}\left|b\Big(\frac{s}{H}\Big)\right| &\leqslant \lim_{T\to\infty}\frac{1}{H}\sum_{s=1-T}^{T-1}\left|b\Big(\frac{s}{H}\Big)\right| \\
&= \lim_{T\to\infty}\int_{(1-T)/H}^{(T-1)/H} b(a)da + \lim_{T\to\infty}\frac{1}{H}b(0) \\
&= \int_{-\infty}^{\infty} b(a)da + o(1) = 1 + o(1).
\end{aligned}$$

Thus, $\sum_{j=1}^{T} b_{jt,H} = O(H)$. Similarly, we obtain $\sum_{j=1}^{T} b_{tj,H} = O(H)$ and $\sum_{j=1}^{T} b^{k}_{jt,H} = O(H)$ for $1 < k \leqslant 2$.

Proof of (ii). Notice that

$$\sum_{j=1}^{T-s} b_{jt,H}^{\frac{r}{2}} b_{j+s,t,H}^{\frac{r}{2}} = \sum_{j=1}^{T-s} b^{\frac{r}{2}}\Big(\frac{j-t}{H}\Big) b^{\frac{r}{2}}\Big(\frac{j+s-t}{H}\Big) = \sum_{m=1-t}^{T-s-t} b^{\frac{r}{2}}\Big(\frac{m}{H}\Big) b^{\frac{r}{2}}\Big(\frac{m+s}{H}\Big)$$
$$\leqslant \sup_a \, b(a) \sum_{m=1-t}^{T-s-t} b^{\frac{r}{2}}\Big(\frac{m}{H}\Big).$$

Then,

$$\left| \frac{1}{H} \sum_{j=1}^{T-s} b_{jt,H}^{\frac{r}{2}} b_{j+s,t,H}^{\frac{r}{2}} \right| \leqslant \sup_a \, b(a) \, \frac{1}{H} \sum_{m=1-T}^{T-1} b^{\frac{r}{2}}\Big(\frac{m}{H}\Big)$$
$$\leqslant \sup_a \, b(a) \int_{-\infty}^{\infty} b^{\frac{r}{2}}(a) da + o(1),$$

which implies (ii), provided that $\int_{-\infty}^{\infty} b^{\frac{r}{2}}(a)da < \infty$. □

Proof of (iii). Notice that from (ii), it is straightforward to see that

$$\frac{1}{H} \sum_{j=1}^{T} b\Big(\frac{j-t}{H}\Big) b\Big(\frac{j+s-t}{H}\Big) = \frac{1}{H} \sum_{m=-\infty}^{\infty} b\Big(\frac{m}{H}\Big) b\Big(\frac{m+s}{H}\Big) + o(1)$$

Using the change of variables $m = \lfloor Hc \rfloor$ and $s = \lfloor Hd \rfloor$,

$$\frac{1}{H} \sum_{m=-\infty}^{\infty} b\Big(\frac{m}{H}\Big) b\Big(\frac{m+s}{H}\Big) = \lim_{T\to\infty} \frac{1}{H} \sum_{m=1-T}^{T-1} b\Big(\frac{m}{H}\Big) b\Big(\frac{m+s}{H}\Big)$$
$$= \lim_{T\to\infty} \int_{(1-T)/H}^{(T-1)/H} b(c)b(c+d)dc + \frac{1}{H}(b(0)b(-\frac{s}{H}) + b(\frac{s}{H})b(0))$$
$$= \int_{-\infty}^{\infty} b(c)b(c+d)dc + o(1).$$

From (1), we see that

$$\frac{1}{H} \sum_{j=1}^{T} b_{jt,H}^2 = \int_{-\infty}^{\infty} b^2(a)da + o(1).$$

Then, we obtain

$$\frac{\sum_{j=1}^{T} b\Big(\frac{j-t}{H}\Big) b\Big(\frac{j+s-t}{H}\Big)}{\sum_{j=1}^{T} b^2\Big(\frac{j-t}{H}\Big)} = \frac{\frac{1}{H}\sum_{j=1}^{T} b\Big(\frac{j-t}{H}\Big) b\Big(\frac{j+s-t}{H}\Big)}{\frac{1}{H}\sum_{j=1}^{T} b^2\Big(\frac{j-t}{H}\Big)} = b^*(d) + o(1).$$

Lemma B2. *Consider a triangular sequence* $\{g_{T,t}\}_t$ *of the form*

$$g_{T,t} = f_T(x_{t-\tau_T}, \cdots, x_{t+\tau_T}),$$

where $f_T(\cdot)$ is some measurable function with respect to the natural filtration $\mathcal{F}_{-\infty}^{t} = \sigma(x_s, s \leqslant t)$. $x_t = x_{T,t}$ is a α-mixing process with mixing coefficient satisfying $\alpha_h^x = O(h^{-\gamma})$ for some $\gamma > 1$. $\tau_T = o(T)$ as $T \to \infty$. Let $\alpha_T^g(h)$ be the mixing coefficient of $\{g_{T,t}\}_t$, which may depend on sample size, $\alpha_T^g(h) = \frac{1}{4}$ for $h \leqslant 0$ and $\alpha_T^g(h) = 0$ for $h \geqslant T$. We have $\sup_T \alpha_T^g(h) \to 0$ as $h \to \infty$, which implies that $\{g_{T,t}\}$ is also α-mixing.

Proof. Define $\mathcal{G}_{T,t} = \sigma(\cdots, g_{T,t-1}, g_{T,t})$, $\mathcal{G}_{T,t+h} = \sigma(g_{T,t+h}, g_{T,t+h+1}, \cdots)$ and the α-mixing coefficient

$$\alpha_T^g(h) = \sup_t \sup_{A \in \mathcal{G}_{T,t}, B \in \mathcal{G}_{T,t+h}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(b)|.$$

Observe that

$$\begin{aligned} \mathcal{G}_{T,t} &= \sigma(\cdots, g_{T,t-1}, g_{T,t}) \subseteq \sigma(\cdots, x_t, x_{t+1}, \cdots, x_{t+h_T}) = \mathcal{F}_{-\infty}^{t+h_T} \\ \mathcal{G}_{T,t+h} &= \sigma(g_{T,t+h}, g_{T,t+h+1}, \cdots) \subseteq \sigma(x_{t+h-h_T}, x_{t+h-h_T+1}, \cdots) = \mathcal{F}_{t+h-h_T}^{\infty}, \end{aligned}$$

which implies that

$$\alpha_T^g(h) \leqslant \alpha^x(h - 2\tau_T) \leqslant C(h - 2\tau_T)^{-\gamma},$$

for some $C > 0$ and $\gamma > 1$. This concludes that $\{g_{T,t}\}$ is also α-mixing. □

Lemma B3. *For a sequence of random variables $\{V_t\}_t$, set $\mathcal{F}_{-\infty}^{t} = \sigma(V_s, s \leqslant t)$. If for some $r > 1$, and all $m \geqslant 1$, $\|E(V_t|\mathcal{F}_{t-m})\|_r \leqslant c_t\psi_m$, $c_t \geqslant 0$ and $\Psi = \sum_{m=1}^{\infty} \psi_m$, then*

$$\Big\|\sum_{j=1}^{T} b_{jt,H}^2 V_j\Big\|_r \leqslant 36\Big(\frac{r}{r-1}\Big)^{\frac{3}{2}} \Psi \Big(\sum_{j=1}^{T} b_{jt,H}^{2r'} c_j^{r'}\Big)^{\frac{1}{r'}},$$

where $r' = \min(r, 2)$ and $b_{jt,H}$ are kernel weights computed with a kernel function $b(\cdot)$ and bandwidth which satisfy Assumption 2.5.

Proof. First, assumptions in this lemma imply that the sequence $\{V_t, \mathcal{F}_{-\infty}^{t}\}$ is an L^r mixingale. Then, following McLeish [1975], we have the following representation for V_t:

$$V_t = \sum_{k=-\infty}^{\infty} V_{kt}, \quad V_{kt} = E_{t-k}V_t - E_{t-k-1}V_t.$$

Note that, according to Lemma 1 in Hansen [1991], $\{V_{kt}, \mathcal{F}_{-\infty}^{t-k}\}$ is a martingale difference sequence (MDS), for all k. Now, following the proof of Lemma 1 in Hansen [1991], for $r > 1$, we have

$$\begin{aligned}
\Big\|\sum_{j=1}^{T} b_{jt,H}^2 V_j\Big\|_r &= \Big\|\sum_{j=1}^{T} b_{jt,H}^2 \sum_{k=-\infty}^{\infty} V_{kj}\Big\|_r = \Big\|\sum_{k}\sum_{j=1}^{T} b_{jt,H}^2 V_{kj}\Big\|_r \\
&\leqslant \sum_{k}\Big\|\sum_{j=1}^{T} b_{jt,H}^2 V_{kj}\Big\|_r \leqslant \sum_{k}\frac{r}{r-1}\Big\|\sum_{j=1}^{T} b_{jt,H}^2 V_{kj}\Big\|_r \\
&\leqslant 18\Big(\frac{r}{r-1}\Big)^{\frac{3}{2}}\sum_{k}\Big(E\Big[\sum_{j=1}^{T} b_{jt,H}^4 V_{kj}^2\Big]^{\frac{r}{2}}\Big)^{\frac{1}{r}}, \qquad (3.34)
\end{aligned}$$

where the final three inequalities are triangular inequality, Doob's inequality and Burkholder's inequality. Following the proof of Lemma 2 in Hansen [1991], we know that

$$\|V_{kt}\|_r \leqslant 2c_t\psi_t.$$

Then, we proceed by separately considering two cases: $1 < r \leqslant 2$ and $r > 2$. For $1 < r \leqslant 2$, observe that, for $x \geqslant 0$, $y \geqslant 0$, $(x+y)^{r/2} \leqslant x^{r/2} + y^{r/2}$. Then, by continuing from (3.34), we have

$$\begin{aligned}
\Big\|\sum_{j=1}^{T} b_{jt,H}^2 V_j\Big\|_r &\leqslant 18\Big(\frac{r}{r-1}\Big)^{\frac{3}{2}}\sum_{k}\Big(\sum_{j=1}^{T} b_{jt,H}^{2r}\|V_{kj}\|_r^r\Big)^{\frac{1}{r}} \\
&\leqslant 18\Big(\frac{r}{r-1}\Big)^{\frac{3}{2}}\sum_{k}\Big(\sum_{j=1}^{T} b_{jt,H}^{2r}(2c_j\psi_k)^r\Big)^{\frac{1}{r}} \\
&= 36\Big(\frac{r}{r-1}\Big)^{\frac{3}{2}}\sum_{k}\psi_k\Big(\sum_{j=1}^{T} b_{jt,H}^{2r}(c_j)^r\Big)^{\frac{1}{r}}.
\end{aligned}$$

For $r > 2$, we continue from (3.34) by applying Minkowski's inequality:

$$\begin{aligned}
\Big\|\sum_{j=1}^{T} b_{jt,H}^2 V_j\Big\|_r &\leqslant 18\Big(\frac{r}{r-1}\Big)^{\frac{3}{2}}\sum_{k}\Big(\Big(\sum_{j=1}^{T}\|b_{jt,H}^4 V_{kj}^2\|_{r/2}\Big)^{r/2}\Big)^{1/r} \\
&= 36\Big(\frac{r}{r-1}\Big)^{\frac{3}{2}}\sum_{k}\psi_k\Big(\sum_{j=1}^{T} b_{jt,H}^4(c_j)^2\Big)^{\frac{1}{2}}.
\end{aligned}$$

□

Lemma B4. *For a sequence of (possibly vector valued) random variables* $\{V_t\}_t$*, set* $\mathcal{F}_{-\infty}^t = \sigma(V_s, s \leqslant t)$*.* α_m *is the associated* α*-mixing coefficient for this process. Assume that for some* $r \in (2,4]$ *such that* $r > 2 + 1/q$ *and* $p > r$*,* $A = 12\sum_{m=0}^{\infty}\alpha_m^{2(1/r-1/p)} < \infty$ *and* $\sup_t\|V_t\|_p = C < \infty$*. Suppose that* $b_{jt,H}$ *are kernel weights computed with a kernel function* $b(\cdot)$ *and bandwidth which satisfy Assumption 2.4, let* V_t^a *be the* *ath elements in V_t, then we have*

$$\max_{a,b}\left\|\frac{1}{K_{2,t}}\sum_{j=1}^{T-s} b_{jt,H}b_{j+s,t,H}\Big(V_j^aV_{j+s}^b - E(V_j^aV_{j+s}^b)\Big)\right\|_{r/2} \leqslant 36(A+2s)(\frac{r}{r-2})^{3/2}\sup_{1\leqslant j\leqslant T-s}\left\|V_j\right\|_p^2\frac{\Big(\sum_{j=1}^{T-s} b_{jt,H}^{\frac{r}{2}}b_{j+s,t,H}^{\frac{r}{2}}\Big)^{\frac{2}{r}}}{K_{2,t}},$$

where $K_{2,t} = \sum_{j=1}^{T} b_{jt,H}^2$.

Proof. We proceed as in De Jong [2000]. Consider the adapted mixingale $\{V_tV_{t+j}, \mathcal{F}_{t+j}\}$, for some $\gamma > \beta \geqslant 1$, and $j < m$, Lemma 1 in Hansen [1992] can be directly applied:

$$\left\|E(V_t^aV_{t+j}^b|\mathcal{F}_{t+j-m}) - E(V_t^aV_{t+j}^b)\right\|_\beta \leqslant 12\alpha_{m-j}^{\frac{1}{\beta}-\frac{1}{\gamma}}\,\mathbb{I}(m > j)\,\|V_t\|_{2\gamma}\left\|V_{t+j}\right\|_{2\gamma}.$$

For $j \geqslant m$, we have

$$\begin{aligned}\left\|E(V_t^aV_{t+j}^b|\mathcal{F}_{t+j-m}) - E(V_t^aV_{t+j}^b)\right\|_\beta &= \left\|E(V_t^aV_{t+j}^b|\mathcal{F}_{t+j-m}) - V_t^aV_{t+j}^b + V_t^aV_{t+j}^b - E(V_t^aV_{t+j}^b)\right\|_\beta\\ &\leqslant \left\|E(V_t^aV_{t+j}^b|\mathcal{F}_{t+j-m}) - V_t^aV_{t+j}^b\right\|_\gamma + \left\|V_t^aV_{t+j}^b - E(V_t^aV_{t+j}^b)\right\|_\gamma\\ &\leqslant \left\|V_t^aV_{t+j}^b\right\|_\gamma + \left\|E(V_t^aV_{t+j}^b)\right\|_\gamma\\ &\leqslant 2\|V_t\|_{2\gamma}\left\|V_{t+j}\right\|_{2\gamma}\mathbb{I}(m \leqslant j),\end{aligned}$$

where the final two inequalities follow from the fact that $V_t^aV_{t+j}^b$ is adaptive to $\mathcal{F}_{t+j-m}$ and the Blackwell's Theorem. Then, over all, we have

$$\left\|E(V_t^aV_{t+j}^b|\mathcal{F}_{t+j-m}) - E(V_t^aV_{t+j}^b)\right\|_\beta \leqslant \Big(12\alpha_{m-j}^{\frac{1}{\beta}-\frac{1}{\gamma}}\,\mathbb{I}(m > j) + 2\mathbb{I}(m \leqslant j)\Big)\|V_t\|_{2\gamma}\left\|V_{t+j}\right\|_{2\gamma}.$$

This also implies that, for $2 < r \leqslant 4$, we have

$$\left\|E(b_{jt,H}b_{j+s,t,H}V_j^aV_{j+s}^b|\mathcal{F}_{j+s-m}) - E(b_{jt,H}b_{j+s,t,H}V_j^aV_{j+s}^b)\right\|_{\frac{r}{2}} \leqslant b_{jt,H}b_{j+s,t,H}\Big(12\alpha_{m-j}^{\frac{1}{\beta}-\frac{1}{\gamma}}\,\mathbb{I}(m > j)+2\mathbb{I}(m \leqslant j)\Big)\left\|V_j\right\|_p\left\|V_{j+s}\right\|_p.$$

Then, by combining the above and Lemma B3, for $2 < r \leqslant 4$ such that $r > 2 + 1/q$ and $p > r$, we obtain

$$\begin{aligned}\max_{a,b}\left\|\frac{1}{K_{2,t}}\sum_{j=1}^{T-s} b_{jt,H}b_{j+s,t,H}\Big(V_j^aV_{j+s}^b - E(V_j^aV_{j+s}^b)\Big)\right\|_{r/2} &\leqslant \frac{36}{K_{2,t}}\left(\sum_{m=0}^{\infty}(12\alpha_{m-j}^{\frac{1}{\beta}-\frac{1}{\gamma}}\,\mathbb{I}(m > j) + 2\mathbb{I}(m \leqslant j))\right)\Big(\frac{r}{r-2}\Big)^{\frac{3}{2}}\\ &\quad\times\left(\sum_{j=1}^{T-s} b_{jt,H}^{\frac{r}{2}}b_{j+s,t,H}^{\frac{r}{2}}\left\|V_j\right\|_p^{\frac{r}{2}}\left\|V_{j+s}\right\|_p^{\frac{r}{2}}\right)^{\frac{2}{r}}\\ &\leqslant 36(A+2s)(r/(r-2))^{\frac{3}{2}}\sup_j\left\|V_j\right\|_p^2\Big(\sum_{j=1}^{T-s} b_{jt,H}^{\frac{r}{2}}b_{j+s,t,H}^{\frac{r}{2}}\Big)^{\frac{2}{r}}/K_{2,t},\end{aligned}$$

where $A = 12\sum_{m=0}^{\infty}\alpha_m^{2(1/r-1/p)}$. □

Summary of the future work

In Chapter 1, we propose a hierarchical shrinkage approach to high-dimensional VARs and show empirically that it is successful to improve macroeconomic forecast accuracy in the multi-country context. Currently, we are investigating on how to extend this approach to the model with censored data. In Giraitis et al. [2016], they estimate a dynamic bivariate Tobit-type model for large financial networks subject to sparsity and persistency. They solve the problem of curse of dimensionality by constructing a few proxy variables to reduce dimensionality a priori. We believe that it would be possible to impose priors on unrestricted model parameters and estimate the model by Bayesian MCMC method.

The direct extension of Chapter 2 is to extend section 2 in this paper to dynamic context. It is well known that, in dynamic panel data model with heterogenous coefficients, pooled estimator is biased but mean group estimator remains consistent. We have obtained some MC results and found that in our setting MG estimator is indeed better than pooled estimator but unlike fixed coefficient case, the differences are rather small. We are currently investigating this theoretically.

There are many possible extensions of Chapter 3. A more detailed extension of Chapter 3 itself includes bootstrap version of the global tests and construct simultaneous confidence bands for estimated path coefficients. A more broader extension includes to relax the identification and correct specification assumptions and provide robust inference procedures. All these extensions are interesting but technically challenging, which certainly demand future studies.

Bibliography

Milton Abramowitz and Irene A Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical table. *Applied Mathematics Series*, 55, 1965.

Tobias Adrian, Richard K Crump, and Emanuel Moench. Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1):110–138, 2013.

Tobias Adrian, Richard K Crump, and Emanuel Moench. Regression-based estimation of dynamic asset pricing models. *Journal of Financial Economics*, 118(2):211–244, 2015.

Donald WK Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

Elena Angelini, Magdalena Lalik, Michele Lenza, and Joan Paredes. Mind the gap: a multi-country bvar benchmark for the eurosystem projections. *International Journal of Forecasting*, 35(4):1658–1668, 2019.

Artin Armagan, Merlise Clyde, and David B Dunson. Generalized beta mixtures of gaussians. In *Advances in neural information processing systems*, pages 523–531, 2011.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Yu Bai, Massimiliano Marcellino, and George Kapetanios. Estimation and inference in large heterogeneous panels with stochastic time-varying coefficients. *mimeo*, 2021.

Badi H Baltagi, Qu Feng, and Chihwa Kao. Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, 191(1):176–195, 2016.

Badi H Baltagi, Qu Feng, and Chihwa Kao. Structural changes in heterogeneous panels with endogenous regressors. *Journal of Applied Econometrics*, 34(6):883–892, 2019.

Pedro Barroso, Martijn Boons, and Paul Karehnke. Time-varying state variable risk premia in the icapm. *Journal of Financial Economics*, 139(2):428–451, 2021.

Luca Benati. Investigating inflation persistence across monetary regimes. *The Quarterly Journal of Economics*, 123(3):1005–1060, 2008.

Angela Bitto and Sylvia Frühwirth-Schnatter. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97, 2019.

Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

Annalisa Cadonna, Sylvia Frühwirth-Schnatter, and Peter Knaus. Triple the gamma - a unifying shrinkage prior for variance and variable selection in sparse state space and tvp models. *Econometrics*, 8(2):20, 2020.

Zongwu Cai. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188, 2007.

Fabio Canova and Matteo Ciccarelli. Estimating multicountry var models. *International Economic Review*, 50(3):929–959, 2009.

Fabio Canova and Matteo Ciccarelli. Panel vector autoregressive models: A survey. In *VAR Models in Macroeconomics - New Developments and Applications*, pages 205–246. Emerald Group Publishing Limited, 2013.

Fabio Canova, Matteo Ciccarelli, and Eva Ortega. Similarities and convergence in g-7 cycles. *Journal of Monetary economics*, 54(3):850–878, 2007.

Andrea Carriero, Todd E Clark, and Massimiliano Marcellino. Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 212(1):137–154, 2019.

Andrea Carriero, Joshua Chan, Todd E Clark, and Massimiliano Marcellino. Corrigendum to: Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Mimeo*, 2021.

Carlos M Carvalho, Nicholas G Polson, and James G Scott. Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics*, pages 73–80, 2009.

Carlos M Carvalho, Nicholas G Polson, and James G Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010.

Joshua CC Chan. Notes on bayesian macroeconometrics. `http://joshuachan.org/papers/BayesMacro.pdf` 2017.

Joshua CC Chan. Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews*, 37(8):807–823, 2018.

Joshua CC Chan. Minnesota-type adaptive hierarchical priors for large bayesian vars. *International Journal of Forecasting,* forthcoming, 2021.

Bin Chen. Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics*, 185(1):196–215, 2015.

Bin Chen and Yongmiao Hong. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183, 2012.

Bin Chen and Yongmiao Hong. Detecting for smooth structural changes in garch models. *Econometric Theory*, 32(3):740–791, 2016.

Bin Chen and Liquan Huang. Nonparametric testing for smooth structural changes in panel data models. *Journal of Econometrics*, 202(2):245–267, 2018.

Todd E Clark and Francesco Ravazzolo. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575, 2015.

John H Cochrane. A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, 104(3):572–621, 1996.

John H Cochrane. *Asset pricing: Revised edition*. Princeton university press, 2009.

Timothy Cogley and Thomas J Sargent. Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic Dynamics*, 8(2):262–302, 2005.

Timothy Cogley, Giorgio E Primiceri, and Thomas J Sargent. Inflation-gap persistence in the us. *American Economic Journal: Macroeconomics*, 2(1):43–69, 2010.

Alain Cohn, Jan Engelmann, Ernst Fehr, and Michel André Maréchal. Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review*, 105(2):860–85, 2015.

Laura Coroneo and Fabrizio Iacone. Comparing predictive accuracy in small samples using fixed-smoothing asymptotics. *Journal of Applied Econometrics*, 35(4):391–409, 2020.

Drew Creal, Siem Jan Koopman, André Lucas, and Marcin Zamojski. Observation-driven filtering of time-varying parameters based on moment conditions. *Mimeo*, 2018.

Jamie L Cross, Chenghan Hou, and Aubrey Poon. Macroeconomic forecasting with large bayesian vars: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, 36(3):899–915, 2020.

Jesús Crespo Cuaresma, Martin Feldkircher, and Florian Huber. Forecasting with global vector autoregressive models: A bayesian approach. *Journal of Applied Econometrics*, 31(7):1371–1391, 2016.

Antonello D'Agostino, Luca Gambetti, and Domenico Giannone. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1):82–101, 2013.

James Davidson. *Stochastic limit theory: An introduction for econometricians*. Oxford University Press, 1994.

Robert M De Jong. A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators. *Econometric Theory*, 16(2):262–268, 2000.

Stéphane Dées and Jochen Güntner. Forecasting inflation across euro area countries and sectors: A panel var approach. *Journal of Forecasting*, 36(4):431–453, 2017.

Marco Del Negro and Giorgio E Primiceri. Time varying structural vector autoregressions and monetary policy: a corrigendum. *The Review of Economic Studies*, 82(4):1342–1345, 2015.

Yiannis Dendramis, Liudas Giraitis, and George Kapetanios. Estimation of time-varying covariance matrices for large datasets. *Econometric Theory,* forthcoming, 2020.

Yiannis Dendramis, Liudas Giraitis, and George Kapetanios. Estimation of time-varying covariance matrices for large datasets. *Econometric Theory,* forthcoming, 2021.

Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 13(3):253–263, 1995.

Thomas Doan, Robert Litterman, and Christopher Sims. Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, 3(1):1–100, 1984.

Stephen G Donald and Whitney K Newey. A jackknife interpretation of the continuous updating estimator. *Economics Letters*, 67(3):239–243, 2000.

Jonas Dovern, Martin Feldkircher, and Florian Huber. Does joint modelling of the world economy pay off? evaluating global forecasts from a bayesian gvar. *Journal of Economic Dynamics and Control*, 70: 86–100, 2016.

Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

Martin Feldkircher, Florian Huber, Gary Koop, and Michael Pfarrhofer. Approximate bayesian inference and forecasting in huge-dimensional multi-country vars. *Mimeo*, 2021.

Lendie Follett and Cindy Yu. Achieving parsimony in bayesian vector autoregressions with the horseshoe prior. *Econometrics and Statistics*, 11:130–144, 2019.

Christian Francq and Jean-Michel Zakoïan. A central limit theorem for mixing triangular arrays of variables whose dependence is allowed to grow with the sample size. *Econometric Theory*, 21(6):1165–1171, 2005.

Jeffrey C Fuhrer and Glenn D Rudebusch. Estimating the euler equation for output. *Journal of Monetary Economics*, 51(6):1133–1153, 2004.

Jordi Galí and Luca Gambetti. Has the us wage phillips curve flattened? a semi-structural exploration. Technical report, National Bureau of Economic Research, 2019.

Jordi Galı and Mark Gertler. Inflation dynamics: A structural econometric analysis. *Journal of monetary Economics*, 44(2):195–222, 1999.

Jordi Galí and Mark Gertler. Inflation dynamics: A structural econometric approach. *Journal of Monetary Economics*, 2:195–222, 1999.

A Ronald Gallant and Halbert White. *A unified theory of estimation and inference for nonlinear dynamic models*. Blackwell, 1988.

Deborah Gefang, Gary Koop, and Aubrey Poon. Variational bayesian inference in large vector autoregressions with hierarchical shrinkage. *Mimeo*, 2019.

Edward I George, Dongchu Sun, and Shawn Ni. Bayesian stochastic search for var model restrictions. *Journal of Econometrics*, 142(1):553–580, 2008.

Eric Ghysels and Alastair Hall. A test for structural stability of euler conditions parameters estimated via the generalized method of moments estimator. *International Economic Review*, pages 355–364, 1990.

Eric Ghysels, Alain Guay, and Alastair Hall. Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics*, 82(2):209–233, 1998.

Domenico Giannone, Lucrezia Reichlin, et al. Comments on" forecasting economic and financial variables with global vars". *International Journal of Forecasting*, 25(4):684–686, 2009.

Domenico Giannone, Michele Lenza, and Giorgio E Primiceri. Economic predictions with big data: The illusion of sparsity. *Econometrica,* forthcoming, 2021.

L Giraitis, G Kapetanios, and T Yates. Inference on multivariate heteroscedastic stochastic time varying coefficient models. *Journal of Time Series Analysis*, 39(2):129–149, 2018.

Liudas Giraitis, George Kapetanios, and Tony Yates. Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, 179(1):46–65, 2014.

Liudas Giraitis, George Kapetanios, Anne Wetherilt, and Filip Žikeš. Estimating the dynamics and persistence of financial networks, with an application to the sterling money market. *Journal of Applied Econometrics*, 31(1):58–84, 2016.

Liudas Giraitis, George Kapetanios, and Massimiliano Marcellino. Time-varying instrumental variable estimation. *Journal of Econometrics,* forthcoming, 2020a.

Liudas Giraitis, George Kapetanios, and Massimiliano Marcellino. Time-varying instrumental variable estimation. *Journal of Econometrics,* forthcoming, 2020b.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

Nikolay Gospodinov and Taisuke Otsu. Local gmm estimation of time series models with conditional moment restrictions. *Journal of Econometrics*, 170(2):476–490, 2012.

Bryan S Graham and James L Powell. Identification and estimation of average partial effects in "irregular" correlated random coefficient panel data models. *Econometrica*, 80(5):2105–2152, 2012.

Jim Griffin and Phil Brown. Hierarchical shrinkage priors for regression models. *Bayesian Analysis*, 12 (1):135–159, 2017.

Jim Griffin and Philip Brown. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188, 2010.

Luigi Guiso, Paola Sapienza, and Luigi Zingales. Time varying risk aversion. *Journal of Financial Economics*, 128(3):403–421, 2018.

Alastair R Hall and Amit Sen. Structural stability testing in models estimated by generalized method of moments. *Journal of Business & Economic Statistics*, 17(3):335–348, 1999.

Alastair R Hall, Sanggohn Han, and Otilia Boldea. Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics*, 170(2):281–302, 2012.

Alastair R Hall, Yuyi Li, Chris D Orme, and Arthur Sinko. Testing for structural instability in moment restriction models: an info-metric approach. *Econometric Reviews*, 34(3):286–327, 2015.

Alastair R Hall et al. *Generalized method of moments*. Oxford university press, 2005.

Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press, 1980.

James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2):357–384, 1989.

Bruce E Hansen. Strong laws for dependent heterogeneous processes. *Econometric theory*, 7(2):213–221, 1991.

Bruce E Hansen. Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica: Journal of the Econometric Society*, pages 967–972, 1992.

Bruce E Hansen. The new econometrics of structural change: dating breaks in us labour productivity. *Journal of Economic perspectives*, 15(4):117–128, 2001.

Bruce E Hansen and Kenneth D West. Generalized method of moments and macroeconomics. *Journal of Business & Economic Statistics*, 20(4):460–469, 2002.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, pages 1029–1054, 1982.

Lars Peter Hansen and Ravi Jagannathan. Assessing specification errors in stochastic discount factor models. *The Journal of Finance*, 52(2):557–590, 1997.

Lars Peter Hansen and Kenneth J Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, pages 1269–1286, 1982.

Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280, 1996.

Peter R Hansen and Asger Lunde. Estimating the persistence and the autocorrelation function of a time series that is measured with error. *Econometric Theory*, 30(1):60–93, 2014.

Matthew Harding and Carlos Lamarche. Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. *Economics Letters*, 111(3):197–199, 2011.

Lukas Hoesch, Barbara Rossi, and Tatevik Sekhposyan. Has the information channel of monetary policy disappeared? revisiting the empirical evidence. *mimeo*, 2020.

Lajos Horváth and Lorenzo Trapani. Statistical inference in a random coefficient panel model. *Journal of econometrics*, 193(1):54–75, 2016.

Cheng Hsiao and M Hashem Pesaran. Random coefficient models. In Laszlo Matyas and Patrick Sevestre, editors, *The econometrics of panel data*, pages 185–213. Springer, 2008.

Cheng Hsiao, M Hashem Pesaran, A Kamil Tahmiscioglu, et al. Bayes estimation of short-run coefficients in dynamic panel data models, 1998.

Florian Huber. Density forecasting using bayesian global vector autoregressions with stochastic volatility. *International Journal of Forecasting*, 32(3):818–837, 2016.

Florian Huber and Martin Feldkircher. Adaptive shrinkage in bayesian vector autoregressive models. *Journal of Business & Economic Statistics*, 37(1):27–39, 2019.

Ravi Jagannathan, Georgios Skoulakis, and Zhenyu Wang. Generalized methods of moments: Applications in finance. *Journal of Business & Economic Statistics*, 20(4):470–481, 2002.

Ted Juhl and Zhijie Xiao. Nonparametric tests of moment condition stability. *Econometric Theory*, 29(1): 90–114, 2013.

George Kapetanios, Massimiliano Marcellino, and Fabrizio Venditti. Large time-varying parameter VARs: A nonparametric approach. *Journal of Applied Econometrics*, 34(7):1027–1049, 2019a.

George Kapetanios, Massimiliano Giuseppe Marcellino, and Fabrizio Venditti. Large time-varying parameter vars: a non-parametric approach. *Journal of Applied Econometrics*, 34(7):1027–1049, 2019b.

Sangjoon Kim, Neil Shephard, and Siddhartha Chib. Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, 65(3):361–393, 1998.

Gary Koop and Dimitris Korobilis. Model uncertainty in panel vector autoregressive models. *European Economic Review*, 81:115–131, 2016.

Gary Koop and Dimitris Korobilis. Forecasting with high dimensional panel vars. *Oxford Bulletin of Economics and Statistics*, 81(5):937–959, 2019.

Gary Koop, Roberto Leon-Gonzalez, and Rodney W Strachan. On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33(4):997–1017, 2009.

Gary Koop, Stuart McIntyre, James Mitchell, and Aubrey Poon. Regional output growth in the united kingdom: more timely and higher frequency estimates from 1970. *Journal of Applied Econometrics*, 35 (2):176–197, 2020.

Dimitris Korobilis. Prior selection for panel vector autoregressions. *Computational Statistics & Data Analysis*, 101:110–120, 2016.

Dimitris Korobilis and Davide Pettenuzzo. Adaptive hierarchical priors for high-dimensional vector autoregressions. *Journal of Econometrics*, 212(1):241–271, 2019.

Guido M Kuersteiner. Kernel-weighted gmm estimators for linear time series models. *Journal of Econometrics*, 170(2):399–421, 2012.

Haiqi Li, Jin Zhou, and Yongmiao Hong. Estimating and testing for smooth structural changes in moment condition models. *Mimeo*, 2021.

Hong Li and Ulrich K Müller. Valid inference in partially unstable generalized method of moments models. *The Review of Economic Studies*, 76(1):343–365, 2009.

Robert B Litterman. Forecasting with bayesian vector autoregressions—five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38, 1986.

Fei Liu, Jiti Gao, and Yanrong Yang. Nonparametric estimation in panel data models with heterogeneity and time-varyingness. *Available at SSRN 3214046*, 2018.

Fei Liu, Jiti Gao, and Yanrong Yang. Time-varying panel data models with an additive factor structure. *Available at SSRN 3729869*, 2020.

Enes Makalic and Daniel F Schmidt. A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182, 2015.

Ulrike Malmendier and Stefan Nagel. Depression babies: do macroeconomic experiences affect risk taking? *The quarterly journal of economics*, 126(1):373–416, 2011.

Don L McLeish. A maximal inequality and dependent strong laws. *The Annals of probability*, 3(5): 829–839, 1975.

Kamiar Mohaddes and Mehdi Raissi. Compilation, revision and updating of the global var (gvar) database, 1979q2-2016q4. 2018.

Jukka Nyblom. Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230, 1989.

Yasuhiro Omori, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449, 2007.

Trevor Park and George Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103 (482):681–686, 2008.

Francisco Peñaranda and Enrique Sentana. A unifying approach to the empirical evaluation of asset pricing models. *Review of Economics and Statistics*, 97(2):412–435, 2015.

M Hashem Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012, 2006.

M Hashem Pesaran. *Time series and panel data econometrics*. Oxford University Press, 2015.

M Hashem Pesaran and Ron Smith. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79–113, 1995.

M Hashem Pesaran and Allan Timmermann. A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4):461–465, 1992.

M Hashem Pesaran, Til Schuermann, and L Vanessa Smith. Forecasting economic and financial variables with global vars. *International Journal of Forecasting*, 25(4):642–675, 2009.

Giorgio E Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.

Gareth O Roberts and Jeffrey S Rosenthal. Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

Peter M Robinson. Time-varying nonlinear regression. In Peter HacklAnders and Holger Westlund, editors, *Economic Structural Change*, pages 179–190. Springer, 1991.

Michael S Smith and Shaun P Vahey. Asymmetric forecast densities for us macroeconomic variables from a gaussian copula model of cross-sectional and serial dependence. *Journal of Business & Economic Statistics*, 34(3):416–434, 2016.

Richard J Smith. Gel criteria for moment condition models. *Econometric Theory*, pages 1192–1235, 2011.

Fallaw Sowell. Optimal tests for parameter instability in the generalized method of moments framework. *Econometrica: Journal of the Econometric Society*, pages 1085–1107, 1996.

James H Stock and Mark W Watson. Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30, 1996.

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179, 2002a.

James H Stock and Mark W Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162, 2002b.

Timo Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218, 1994.

Nadeem Ul Haque, M Hashem Pesaran, and Sunil Sharma. Neglected heterogeneity and dynamics in cross-country savings regressions. IMF working paper*, International Monetary Fund, Washington, D.C.*, 1999.