Time-varying coefficient estimation in SURE models. Application to portfolio management.

Isabel Casas [*]   Eva Ferreira [†]   Susan Orbe [‡]

September 20, 2017

Abstract

This paper provides a detailed analysis of the asymptotic properties of a kernel estimator for a Seemingly Unrelated Regression Equations model with time-varying coefficients (tv-SURE) under very general conditions. Theoretical results together with a simulation study differentiates the cases for which the estimation of a tv-SURE outperforms the estimation of a Single Regression Equations model with time-varying coefficients (tv-SRE). The study shows that Zellner's results cannot be straightforwardly extended to the time-varying case. The tv-SURE is applied to the Fama and French five-factor model using data from four different international markets. Finally, we provide the estimation under cross-restriction and discuss a testing procedure.

Keywords: Time-varying; Nonparametric; SURE; Five factor model; Asset pricing.

JEL: C01, C03, C14.

1   Introduction

Systems of multivariate regression equations are very common in the economic literature. The household consumption problem, Grunfeld's investment data problem, asset pricing, portfolio management and monetary policy problems are all suitable for multi-equation models. In fact, due to the financial globalization, strong and complex relationships among different markets have likely been established which lead to multi-equation specifications of asset/portfolio returns. For example, recently Ando and Bai (2015) analysed multifactor models in the presence of a large number of potential observable risk factors and

---

[*]Basque Center for Applied Mathematics, Avenida Mazarredo 14, 48009, Bilbao, Spain; and The University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark. *Corresponding author*: icasas@bcamath.org.

[†]Departamento de Economía Aplicada III & BETS. University of the Basque Country, Avenida Lehendakari Aguirre 83, 48015 Bilbao, Spain. *e-mail*: eva.ferreira@ehu.eus.

[‡]Departamento de Economía Aplicada III & BETS. University of the Basque Country, Avenida Lehendakari Aguirre 83, 48015 Bilbao, Spain. *e-mail*: susan.orbe@ehu.eus.

unobservable common and group-specific factors, where cross-correlation among equations is exploited. In summary, asset pricing models can be estimated in a seemingly unrelated regression equation (SURE) setting, exploiting the advantage from the cross-relation among equation errors.

In general, parameters of a system of equations may be estimated independently for each equation as if all equations were self-sufficient. We refer to this methodology as the Single Regression Equation (SRE, from now on). However, it is possible to consider that all equations in the system are bound by the variance-covariance matrix of the errors, the SURE methodology. Although, under the usual hypotheses of lack of causality and absence of lags, the SRE and SURE estimators are both consistent, Zellner (1962) shows that the SURE estimator may have a smaller variance than the SRE estimator, especially when there is a strong correlation among the errors of each equation and when the regressors among equations are orthogonal. Moreover, if the regressors are the same across equations, the SRE and SURE estimators are equivalent under stationary errors. Thus, there is no gain in estimating the classic CAPM as a SURE model rather than as a SRE model when the market return is the same for all equations. However, if the stocks are from different markets or a more complex model is used such as the Fama and French three-factor or the Fama and French five-factor models (Fama and French (1993), Fama and French (2015)), where regressors among equations are different, then the SURE might provide a more efficient estimator than the SRE, with the lowest variance.

Other examples where a SURE may be advantageous are: 1) international asset pricing models where instead of global market factors, local or domestic factors are used as regressors, as in Fama and French (2017); 2) models whose regressors include firm, industry and/or country macroeconomic variables, as in Aretz et al. (2010); and 3) models that relate coefficients with state variables as in Cai et al. (2015).

Additionally, economies and financial markets evolve over time and investment strategies that worked in the past might not be viable at the present time, thus models must be flexible to account for these changes. Recently, research in time-varying coefficients in nonparametric asset pricing models has been popular lately, see Ang and Kristensen (2012), Ferreira et al. (2011), Esteban et al. (2015) and Cai et al. (2015) for further details. Time-varying coefficients account for the variations in the asset/portfolio sensitivity to the risk factors, along time. On the basis of this, this paper focuses on SURE models with time-varying coefficients (tv-SURE from now on) and on its comparison with the time-varying Single Regression Equation (tv-SRE). The objective of our research is to study the conditions in which the tv-SURE is advantageous over the tv-SRE, and therefore the conditions in which Zellner's result may be extended to the time-varying framework.

Methodologically, we propose a kernel smoothing estimator for the above models, within the family of Orbe et al. (2003, 2005); Cai (2007); Ferreira et al. (2011); Ang and Kristensen (2012); Phillips et al. (2017) among others. This approach enables us to relax the usual assumption of stationarity and consider locally stationary variables as defined in Dahlhaus (1997) and Dahlhaus (2000).

A different approach has been considered in Henderson et al. (2015), where a SURE estimation is provided through coefficients depending on some extra stationary variables, in a similar manner to Cai et al. (2009) and Das (2005).

There are two main differences between the constant coefficients and time-varying coefficients frameworks. First, nonparametric estimators of the tv-SURE and tv-SRE are biased and this bias plays an important role in the asymptotic results. The best estimator between the tv-SURE and the tv-SRE is the one with the smallest mean square error. In Zellner's framework instead, the best estimator between the SURE and SRE is the one with the smallest variance, as both of them are unbiased. Second, the smoothing parameter or bandwidth also plays an important role in the choice of the best estimator. Even if all regressors are equal across equations, the tv-SRE and tv-SURE estimates might be different when each equation is estimated with a different bandwidth. An intuitive explanation of this is that, although a tv-SURE model is linear, the inclusion of time-varying coefficients permits nonlinear relationships between the regressors and the dependent variable. Recall that any non-linear model may be approximated by a time-varying coefficients linear model as shown in Granger (2008).

Therefore, Zellner's results cannot be straightforwardly extended to the time-varying coefficients context and a deeper analysis must be done before deciding to use a tv-SURE over a tv-SRE.

We propose a general tv-SURE estimator, that allows for different degree of smoothness across equations. The asymptotic results provide consistency and asymptotic normality, at the usual rates of converge of nonparametric methodologies. A careful study of the leading terms shows the role of the bandwidth in the bias and variance. Thus, the theoretical results and the simulation study support the selection of a unique bandwidth for all equations in this context.

Empirically, the tv-SURE for asset pricing models uses the information across markets that, in a globalized world, are likely to be highly correlated. Moreover, the time varying estimation of the coefficients might benefit dynamic investment strategies. We apply the estimation procedure to the Fama and French five-factors specification, Fama and French (2015) and Fama and French (2017), and a discussion of the results is provided. We have used portfolios for four different regions, North America, Europe, Asia Pacific and Japan. The estimation of the sensitivities over time allows us to make an interpretation of the betas for different types of portfolios not only through their long-run value as in Ang and Kristensen (2012), but also according to the stability over time and the variations over several financial periods. Since the coefficients measure the sensitivity of portfolio excess returns to risk factors in a certain region, a natural question is whether or not this sensitivity is equal for all international markets. With this question in mind, a tv-SURE estimator under cross-restrictions and a test of these restrictions are proposed and the asymptotic properties derived.

The rest of the paper is structured as follows. Section 2 presents the model and the estimators proposed, together with the main asymptotic properties. Section 3 presents a simulation study. Section 4 presents the results for the Fama and French five-factor application. Section 5 discusses the estimation

and testing when cross-restrictions are considered and Section 6 concludes. The proofs of the main results are relegated to the Appendix.

2 Tv-SURE estimation, consistency and asymptotic normality

A tv-SURE model is a system of Seemingly Unrelated Regression Equations with time-varying coefficients, as follows,

$$y_{mt} = x_{mt}\beta_{mt} + u_{mt} \quad m = 1,\ldots,M \quad t = 1,\ldots,T; \tag{1}$$

where y_{mt} denotes the t-th observation of the explained variable and the exogenous variables are $x_{mt} = (x_{m1t},\ldots,x_{mp_mt})$. Each equation might have a different number of exogenous variables, p_m. The p_m order vector $\beta_{mt} = (\beta_{m1t},\ldots,\beta_{mp_mt})^T$ is unknown and represents the coefficients at time t of equation m. The error term u_{mt} is a random process such that $\mathbb{E}(u_{mt}) = \mathbb{E}(u_{mt}|x_{mt}) = 0$ and $\mathbb{E}(u_{mt}u_{m't'}) = \delta_{tt'}\sigma_{mm't}$, where $\delta_{tt'} = 0$ if $t \neq t'$ and 1 if $t = t'$.

In summary, the system consists of M equations with T observations for each and $P = \sum_{m=1}^{M} p_m$ number of explanatory variables. Stacking all equations in (1), the compact matrix notation for the tv-SURE is,

$$Y_t = X_t\beta_t + u_t \quad m = 1,\ldots,M \quad t = 1,\ldots,T; \tag{2}$$

where $Y_t = (y_{1t}\ldots y_{Mt})^T$, $X_t = diag(x_{1t}\ldots x_{Mt})$ and $\beta_t = (\beta_{1t}^T,\ldots,\beta_{Mt}^T)^T$ is a vector of order P. The error vector $u_t = (u_{1t}\ldots u_{Mt})^T$ has zero mean and covariance matrix $\mathbb{E}(u_t u_t^T) = \Sigma_t$ with elements $\sigma_{mm't}$.

The aim of this section is to study the Nadaraya-Watson estimator of the coefficients (betas) and its properties in a general setting. Neither the errors, nor the explanatory variables need to be stationary, only locally stationary. The possibility of a different number of explanatory variables in each equation and M different bandwidths will be considered. Under this framework, the estimators in Ang and Kristensen (2012) and Kristensen (2012) consider the estimation of single regressions or tv-SRE with different bandwidths instead of estimating the bound system or tv-SURE. Henderson et al. (2015) consider estimating a system of varying coefficients model for cross-sectional data using a common bandwidth. On the other hand, when several equations are considered with the same realized explanatory variables and a common bandwidth, the estimator becomes the same as in Ferreira et al. (2011). The novelty of this work is the study of the conditions in which a tv-SURE and a tv-SRE are different and therefore, the extra computational burden of a tv-SURE is justifiable in favour of a more efficient estimator.

The proposed Nadaraya-Watson tv-SURE estimator of model (2) at time t minimizes the following smoothed weighted sum of squared residuals,

$$\widehat{\beta}_t \equiv \arg\min \sum_{s=1}^{T} (Y_s - X_s\beta_t)^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} (Y_s - X_s\beta_t)$$

where $K_{H,ts} = diag(K_{h_1,ts}, ..., K_{h_M,ts})$ and $K_{h_m,ts} = (Th_m)^{-1}K((t-s)/(Th_m))$ is the matrix of weights introducing smoothness according to the vector of bandwidths, $H = (h_1, \ldots, h_M)^T$. Note that this minimization problem accounts for the time-varying structure of the variance-covariance matrix of the errors, Σ_t. The solution is the vector of coefficient estimates at time t below,

$$\widehat{\beta}_t = \left(\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \right)^{-1} \sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} Y_s. \quad (3)$$

The resulting tv-SURE estimator has a closed form that depends on the bandwidth vector H and the covariance matrices Σ_s.

Remark 1. If the components of H are large enough, the nonparametric tv-SURE estimator in equation (3) becomes the constant parametric estimator below,

$$\widehat{\beta}_t \equiv \widehat{\beta} = \left(\sum_{s=1}^{T} X_s^T \Sigma_s^{-1} X_s \right)^{-1} \sum_{s=1}^{T} X_s^T \Sigma_s^{-1} Y_s;$$

which allows for heteroscedasticity and time-varying contemporaneous correlation in the errors. Moreover, if the errors are stationary; that is, $\Sigma_t = \Sigma$ for all t, the estimator coincides with Zellner's estimator, see Zellner (1962),

$$\widehat{\beta} = \left(\sum_{s=1}^{T} X_s^T \Sigma^{-1} X_s \right)^{-1} \sum_{s=1}^{T} X_s^T \Sigma^{-1} Y_s.$$

Remark 2. If no smoothing over time is done, i.e. $K_{H,ts} = I$, estimator (3) looks like,

$$\widehat{\beta}_t = \left(X_t^T \Sigma_t^{-1} X_t \right)^{-1} X_t^T \Sigma_t^{-1} Y_t,$$

Note that the expression above needs the total number of variables in the system, P, to be smaller than the number of equations, M, to ensure identificability of the unknown coefficients.

Remark 3. If there is no contemporaneous correlation, i.e Σ_t is diagonal, $\Sigma_t = diag(\sigma_{mt}^2)_{m=1}^M$, then the estimator may be written as,

$$\widehat{\beta}_t = \left(\sum_{s=1}^{T} X_s^T \ diag(K_{h_m,ts}/\sigma_{ms}^2)_{m=1}^M \ X_s \right)^{-1} \sum_{s=1}^{T} X_s^T \ diag(K_{h_m,ts}/\sigma_{ms}^2)_{m=1}^M \ Y_s.$$

It is straightforward to check that in this case the beta estimator for each equation is,

$$\hat{\beta}_{mt} = \left(\sum_{s=1}^{T} (K_{h_m,ts}/\sigma_{ms}^2) x_{ms}^T x_{ms} \right)^{-1} \sum_{s=1}^{T} (K_{h_m,ts}/\sigma_{ms}^2) x_{ms}^T y_{ms}.$$

5

In other words, the estimation problem is equivalent to estimating each equation independently by time-varying weighted least squares, which is a particular case of the tv-SURE, the tv-SRE with heteroscedastic errors.

In the general framework, we state the following assumptions and results for consistency and asymptotic normality of the proposed estimator (3).

Assumption A.1 The coefficients are smooth functions of the rescaled time, that is, $\beta_{mit} = \beta_{mi}(t/T)$ for $m = 1, \ldots, M$ and $i = 1, \ldots, p_m$.

Assumption A.2 The weight function $K(u)$ is a symmetric second order kernel with compact support in $[-1, 1]$, Lipschitz continuous and its Fourier transform is absolutely integrable, such that $d_k = \int u^2 K^2(u) du$, $c_4 = \int K^4(u) du$ and $c_k = \int K^2(u) du$ are bounded.

Assumption A.3 The sequence $\{X_{mt}, u_{mt}\}$ is strong α-mixing with coefficients $\alpha(k)$ of order $6/5$, i.e. $\alpha(k) = O(k^{-\delta})$, with $\delta > 6/5$. All moments up to order $12 + \theta$ exist and they are uniformly bounded, for some positive θ.

Assumption A.4 The unconditional expectation $G_t = \mathbb{E}\left(X_t^T \Sigma_t^{-1} X_t\right)$ at each time t is symmetric and strictly positive definite. It can be decomposed as a smooth function of t/T (at least twice differentiable and uniformly bounded) and a term of order $O(T^{-1})$.

Assumption A.5 The error term u_t has zero mean and conditional covariance matrix $\Omega_t = \mathbb{E}(u_t u_t^T | X_t) = \mathbb{E}(u_t u_t^T)$, symmetric and positive definite.

Assumption A.6 At each time t, the matrix

$$\hat{G}_t = \sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \tag{4}$$

is positive definite and uniformly bounded from above and below.

Assumption A.7 The smoothing vector H is such that, for each m, $h_m = O(h_1)$ where $h_1 = h$ goes to zero and Th goes to infinity, as the sample size T goes to infinity.

Assumption *A.1* imposes the property of smoothness on the coefficients over time and when the Nadaraya-Watson estimator is used, it means that $\beta_{mi}(\cdot) \in C^2[0, 1]$. Assumption *A.2* is a technical assumption in nonparametric kernel estimation. Assumptions *A.3–A.4* impose the condition of locally stationarity on the generating distribution process, allowing for time-varying first and second moments on the explanatory variables. These types of processes are very useful and they can model nonstationary variables that, contrary to the unit root context, have a nonexplosive behaviour (see seminal works by Dahlhaus (1997, 2000)). Assumption *A.5* allows for heteroscedasticity and also for a time-varying contemporary correlation structure, always excluding exploiting patterns. Assumption *A.6* is a full rank condition for the identification of the estimator in order to ensure a unique solution of the optimization minimization problem. Finally, Assumption *A.7* is a standard condition in nonparametric estimation necessary to prove consistency.

To satisfy the reasonable property of consistency, we analyse the rate of convergence of the Mean Average Square Error, $MASE$, defined as,

$$MASE\left(\hat{\beta}_t\right) = MASE_t = \ tr\ \mathbb{E}\left[\left(\hat{\beta}_t - \beta_t\right)\left(\hat{\beta}_t - \beta_t\right)^T\right], \tag{5}$$

where tr refers to the trace. We may write the $MASE_t$ in terms of bias and variance as follows,

$$MASE_t \ = \ ||Bias(\hat{\beta}_t)||_2^2 + tr\ V(\hat{\beta}_t).$$

The next theorem proves the convergence of $MASE_t$ to zero as the sample size $T \to \infty$, which implies the consistency of the tv-SURE estimator.

Theorem 1 *Under assumptions A.1–A.7, the asymptotic bias and variance of the tv-SURE estimator are,*

$$Bias(\hat{\beta}_t) \ = \ \frac{d_k}{2}G_t^{-1}\mathbb{H}^2\left(G_t\beta_t'' + 2G_t'\beta_t'\right) + o\left(h^2\right)$$

$$V(\hat{\beta}_t) \ = \ \frac{c_k}{T}\mathbb{H}^{-1}G_t^{-1} + o\left(\frac{1}{Th}\right)$$

where β_t', β_t'', and G_t' denote the respective derivatives. $\mathbb{H} = diag(h_1 I_{p_1} \ldots h_M I_{p_M})$ is the matrix conformed by the smoothing parameters, where $I.$ is the identity matrix.

Theorem 1 provides the consistency of the estimator defined in (3). In addition, it provides the expressions of the bias and variance of the estimator, showing the crucial role of the bandwidth in both terms and its effect on the tv-SURE estimator. Note that from Theorem 1, $MASE = O(h^4 + (Th)^{-1})$ in this general tv-SURE framework, which is the usual order of convergence of classical time-varying nonparametric estimators, see Cai (2007) among others. However, there is an extra term containing G' contributing to the bias of our estimator in comparison to the time-varying case in Henderson et al. (2015).

There are two main differences from the parametric case that prevent the immediate generalization of Zellner's results to this nonparametric framework. The first difference comes from the bias. Since the nonparametric estimator is a biased estimator, the minimization of the mean square error is not only concerned with the minimization of the variance. The second main difference comes from the multivariate bandwidth whose presence prevents the G_t term from disappearing from the $Bias$ expression. Therefore, in the general nonparametric framework, the advantages of using tv-SURE over the tv-SRE are not straightforward.

In the particular setting where the bandwidth is the same for all equations and the errors and explanatory variables are stationary, the comparison of the two estimators is similar to the comparison of the parametric estimators. The next corollary states this result.

Corollary 2 *Under assumptions in Theorem 1, and in addition to $\mathbb{H} = hI_P$, $\Sigma_t \equiv \Sigma$ (i.e. $G_t \equiv G$), the bias and variance terms in the $MASE_t$ defined in (5) become,*

$$
\begin{aligned}
Bias(\hat{\beta}_t) &= \frac{d_k}{2} h^2 \beta_t'' + o\left(h^2\right) \\
V(\hat{\beta}_t) &= \frac{c_k}{Th} G^{-1} + o\left(\frac{1}{Th}\right)
\end{aligned}
$$

Remark 4. In this particular setting, it is easy to check that for equal regressors in all equations, $X_m = X_{m'}$ for $m \neq m'$, the tv-SURE and the tv-SRE estimators are the same, as in the parametric case. However, if the estimator considers different bandwidths for each equation, the equivalence does not hold, even under stationary errors.

In summary, the tv-SURE estimator has the usual asymptotic nonparametric convergence rates with an optimal rate of $O(T^{-4/5})$ for $h = O(T^{-1/5})$. However, in finite samples, the selection of one common bandwidth for all equations or the selection of several bandwidths makes a difference in the bias and variance values. The advantage of the tv-SURE over the tv-SRE estimator is clear under a common bandwidth. Surprisingly, this advantage is not obvious when selecting different bandwidths for each equation. Roughly speaking, when the bandwidths are different across equations, their interaction with the matrices G_t and G_t' affect the bias and variance in a cumbersome and not obvious way.

To overcome this drawback, a suboptimal bandwidth may be selected such that the bias is negligible (always asymptotically) with respect to the variance. In that case, the normal asymptotic distribution may be derived as stated in the theorem below.

Theorem 3 *Consider the same assumptions as in Theorem 1, and a suboptimal rate for the bandwidths such that $h = o(T^{-1/5})$. Then, the pointwise asymptotic distribution for the tv-SURE estimator of β_t is*

$$
\sqrt{T}\mathbb{H}^{1/2}\left(\widehat{\beta}_t - \beta_t\right) \rightarrow N\left(0, c_k G_t^{-1}\right)
$$

3   A simulation study

Theoretical results in the previous section show that the estimation of the tv-SRE is only equivalent to the estimation of the tv-SURE for the case of equal regressors across equations with stationary errors, while using a common bandwidth for all equations. Simulation results in this section confirm this finding. In addition, it is natural to wonder which of the two estimators fares best when a different bandwidth for each equation is used. There is no clear theoretical result and, therefore we resort to simulations to shine some light on this question.

The simulated model consists of two equations with two variables each as follows,

$$y_{1t} = \beta_{11,t}x_{11,t} + \beta_{12,t}x_{12,t} + u_{1t},$$
$$y_{2t} = \beta_{21,t}x_{21,t} + \beta_{22,t}x_{22,t} + u_{2t} \quad t = 1, \ldots, T.$$

Two different specifications are taken for coefficients, regressors, the variance-covariance matrix of errors and the sample size. In total, there are 24 different settings generated in order to support the main results derived from the theory. Two scenarios are used for the time-varying coefficients depending on the degree of smoothness of the coefficients. As can be seen in Figure 1, in Scenario 1 (left plot), the coefficients of both equations have the same smoothness degree; while in Scenario 2 (right plot), the coefficients of the second equation are less smooth. The functional forms taken for both scenarios are detailed in Table 1. These two scenarios help to clarify the effect that coefficients' shape (first and second derivatives) have over the estimator bias.

Figure 1: Simulation time-varying coefficients.
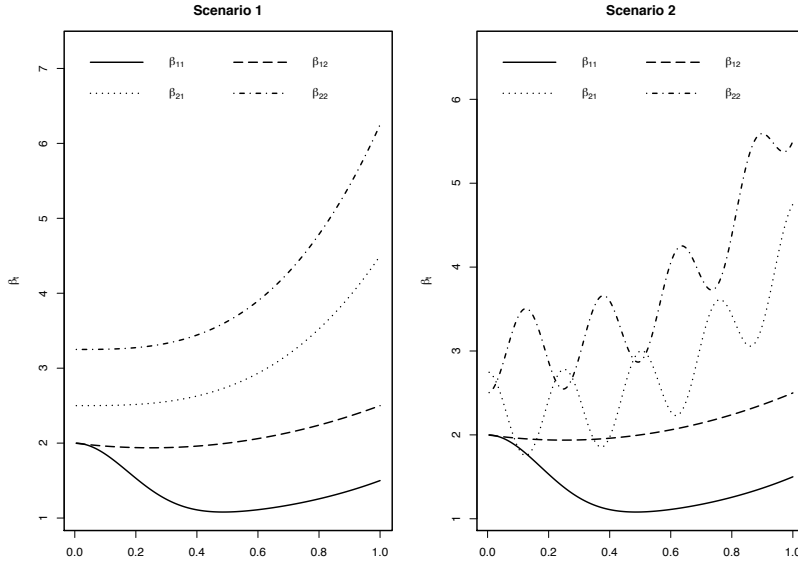
Coefficients for Scenario 1 (left) and Scenario 2 (right). The coefficients β_{11} and β_{12} correspond to the coefficients of the first equation and β_{21} and β_{22} to the second equation.

For the regressors, the two extreme cases are considered: orthogonal regressors across equations, $x_{1t} \perp x_{2t}$ and equal regressors across equations, i.e. $x_{1t} = (x_{11,t}, x_{12,t}) = (x_{21,t}, x_{22,t}) = x_{2t}$. Regressors are generated as normal independent random variables, $x_{i1,t} \sim IIDN(2,1)$ and $x_{i2,t} \sim IIDN(-3,1)$ for $i = 1, 2$. Regarding the disturbances, they were simulated from a multivariate

9

Table 1: Time-varying coefficients specification and errors deviations.

	Scenario 1	Scenario 2
β_{11t}	$0.5(t/T)^3 + \exp\left(-16(t/T)^2\right) + 1$	$0.5(t/T)^3 + \exp\left(-16(t/T)^2\right) + 1$
β_{12t}	$(t/T)^2 - 0.5(t/T) + 2$	$(t/T)^2 - 0.5(t/T) + 2$
β_{21t}	$2(t/T)^3 + 2.5$	$\cos^2(4\pi(t/T)) + 2(t/T)^3 + 1.75$
β_{22t}	$3(t/T)^3 + 3.25$	$\sin^2(4\pi(t/T)) + 3(t/T)^3 + 2.5$
σ_1	1.674	1.674
σ_2	7.813	8.3019

Coefficients (betas) in Scenario 1 have the same smoothness for both equations; while the betas in the second equation of Scenario 2 are less smooth than the betas in the first equation. The standard variation is constant for both equations and scenarios. It accounts for about 10% of the variance of the simulated dependent variable.

normal distribution,

$$u_t = (u_{1t}, u_{2t})' \sim N\left(\mathbf{0}, \Sigma_t = \begin{pmatrix} 0.3\ \sigma_1^2 & 0.3\ \rho_t \sigma_1 \sigma_2 \\ 0.3\ \rho_t \sigma_1 \sigma_2 & 0.3\ \sigma_2^2 \end{pmatrix}\right).$$

Errors are generated with a constant variance of about 10% of the variance of the dependent variables in order to make comparison easier. Thus, the variance of the error for the second equation is larger (see Table 1), especially in Scenario 2. The errors are contemporaneously correlated. The first case considers a constant contemporaneous correlation $\rho_t \equiv \rho = 0.9$, i.e. stationary errors. The second case considers a decreasing time-varying contemporaneous correlation, $\rho_t = \exp(-0.5(t/T)^2) - 0.2$, that ranges from 0.8 to 0.4. These two cases permit the study of the effect of the correlation not only on the variance but also on the bias of the estimator. Let us recall that the bias term contains the first derivative of the variance-covariance matrix.

Two sample sizes are used, $T = 200, 500$, and the number of replications is $R = 1000$. The performance comparison between the tv-SRE and the tv-SURE estimation is done using the mean of the relative bias squared and the mean of the relative variance, defined as follows,

$$MRBias^2(\hat{\beta}_{ij}) = MRBias_{ij}^2 = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{R}\sum_{r=1}^{R}\frac{\hat{\beta}_{ij,t}^{(r)}}{\beta_{ij,t}} - 1\right\}^2$$

$$MRVar(\hat{\beta}_{ij}) = MRVar_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left\{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{\beta}_{ij,t}^{(r)}}{\beta_{ij,t}}\right)^2 - \left(\frac{1}{R}\sum_{r=1}^{R}\frac{\hat{\beta}_{ij,t}^{(r)}}{\beta_{ij,t}}\right)^2\right\}$$

where $\hat{\beta}_{ij,t}^{(r)}$ is the estimate of $\beta_{ij,t}$ obtained in the rth replication.

For the tv-SRE, one bandwidth is selected for each equation independently by minimising the one-leave-out cross-validation function,

$$CV(\hat{h}_i) = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - x_{it}\hat{\beta}_{it,(-t)})^2 \tag{6}$$

10

for equations $i = 1, 2, \ldots, m$. Term $\hat{\beta}_{it,(-t)}$ is the estimator of β_{it} with bandwidth h_i without using tth datum (x_{it}, y_{it}).

For the tv-SURE estimator a multivariate cross-validation function is minimized to selected the bandwidths,

$$CV(\tilde{h}_1, \ldots, \tilde{h}_m) = \frac{1}{T} \sum_{t=1}^{T} (Y_t - X_t \hat{\beta}_{t,(-t)})^2. \qquad (7)$$

It is also possible to choose a common bandwidth for all equations. In our simulation study, the common bandwidth for the tv-SRE, denoted by $\hat{h}$, is chosen as the mean of $\hat{h}_1, \ldots, \hat{h}_m$ for simplicity. Similarly, the common bandwidth for the tv-SURE, $\tilde{h}$ is chosen as the mean value of $\tilde{h}_1, \ldots, \tilde{h}_m$.

Until now, it was assumed that the errors variance-covariance matrix was known. However, it is commonly estimated in practice, resulting in the Feasible tv-SURE estimator. This estimator consists of two steps, in the first Σ_t is estimated and plugged into the tv-SURE in the second step.

1. Estimate Σ_t based on the residuals of the tv-SRE. If Σ_t is known to be constant, the sample covariance matrix is a consistent estimator of it. If Σ_t changes over time, a nonparametric estimator such as in Aslanidis and Casas (2013) is a consistent alternative.

2. Estimate the coefficients of the tv-SURE by plugging $\hat{\Sigma}_t$ into (3).

3.1   Results

Table 2 presents the simulation results for orthogonal regressors, constant error variance and possibly time-varying contemporaneous correlation. Results are organized by blocks of two columns and four rows where element (i, j) with $i, j = 1, 2$ corresponds to the $MRBias_{ij}^2$ and elements in squared brackets correspond to the $MRVar_{ij}$. Panel A shows the results for the constant contemporaneous correlation cases with $\rho = 0.9$ for all cases. Panel B shows the results for a time-varying contemporaneous correlation where $\rho_t = \exp(-0.5(t/T)^2) - 0.2$. Both panels consider two scenarios for the time-varying coefficients, first with similar smoothness and second with different smoothness in the coefficients, as described in Table 1 and Figure 1. The tv-SRE and tv-SURE are compared for different bandwidth selections and sizes $T = 200, 500$.

As expected from Theorem 1, there are some general results that always hold. First, all estimation errors decrease as the sample size increases, as a consequence of consistency. Second, the $MRBias_{21}^2$ and $MRBias_{22}^2$ are always larger in Scenario 2 than in Scenario 1. This is reasonable because, as we have seen, the estimator bias term depends on the second derivative of β_t which is larger for the second equation coefficients in Scenario 2 than in Scenario 1. Third, the tv-SURE estimates have smaller mean relative variances than the tv-SRE for all comparable cases: 1) for a common bandwidth $\hat{h}$ and 2) for the same pair of bandwidths $(\hat{h}_1, \hat{h}_2)$. Therefore, in the orthogonal stationary regressors

11

Table 2: Mean relative bias squared and mean relative variance under cross-orthogonal regressors.

Panel A: Constant contemporaneous correlation ($\rho = 0.9$)

	tv-SRE ($\hat{h}$)	tv-SURE ($\hat{h}$)	tv-SRE ($\hat{h}_1, \hat{h}_2$)	tv-SURE ($\hat{h}_1, \hat{h}_2$)	tv-SURE ($\tilde{h}$)	tv-SURE ($\tilde{h}_1, \tilde{h}_2$)
Scenario 1: Coefficients with similar smoothness.						
T=200	1.101 0.102	1.073 0.081	0.417 0.042	0.344 0.083	0.791 0.062	0.182 0.054
	0.498 0.434	0.544 0.410	0.645 0.666	0.497 0.922	0.542 0.370	0.484 0.854
	[1.328] [0.239]	[0.280] [0.060]	[1.812] [0.317]	[0.388] [0.085]	[0.324] [0.069]	[0.493] [0.102]
	[6.307] [1.623]	[1.162] [0.412]	[5.371] [1.368]	[2.272] [0.726]	[1.291] [0.454]	[2.823] [0.863]
T=500	0.261 0.025	0.281 0.022	0.099 0.013	0.097 0.027	0.516 0.033	0.114 0.043
	0.109 0.131	0.125 0.124	0.207 0.240	0.135 0.323	0.189 0.187	0.324 0.708
	[0.653] [0.119]	[0.136] [0.029]	[0.843] [0.153]	[0.178] [0.039]	[0.136] [0.028]	[0.190] [0.040]
	[2.938] [0.849]	[0.593] [0.206]	[2.423] [0.701]	[1.076] [0.372]	[0.559] [0.195]	[1.085] [0.369]
Scenario 2: Coefficients with different smoothness.						
T=200	1.171 0.120	1.169 0.099	0.664 0.076	0.533 0.136	0.617 0.069	0.186 0.101
	3.131 1.715	4.438 1.872	3.815 2.132	4.227 2.824	3.947 1.535	1.706 3.005
	[1.005] [0.205]	[0.206] [0.055]	[1.248] [0.250]	[0.266] [0.075]	[0.284] [0.072]	[0.408] [0.106]
	[7.528] [1.714]	[1.502] [0.446]	[6.136] [1.439]	[2.650] [0.745]	[2.020] [0.582]	[3.872] [1.054]
T=500	0.469 0.029	0.491 0.027	0.177 0.014	0.136 0.073	1.676 0.116	0.746 0.154
	2.727 1.545	2.728 1.507	2.264 1.223	1.564 1.024	2.570 1.593	2.821 2.827
	[0.601] [0.121]	[0.123] [0.033]	[0.825] [0.163]	[0.171] [0.047]	[0.089] [0.021]	[0.123] [0.032]
	[4.720] [1.240]	[0.960] [0.281]	[3.901] [1.006]	[1.805] [0.521]	[0.619] [0.188]	[1.246] [0.360]

Panel B: Time-varying contemporaneous correlation ($\rho_t = \exp(-0.5(t/T)^2) - 0.2$)

	tv-SRE ($\hat{h}$)	tv-SURE ($\hat{h}$)	tv-SRE ($\hat{h}_1, \hat{h}_2$)	tv-SURE ($\hat{h}_1, \hat{h}_2$)	tv-SURE ($\tilde{h}$)	tv-SURE ($\tilde{h}_1, \tilde{h}_2$)
Scenario 1: Coefficients with similar smoothness.						
	1.234 0.104	1.148 0.074	1.778 0.157	1.896 0.181	0.372 0.029	0.555 0.044
T=200	0.562 0.448	0.510 0.445	0.481 0.326	0.422 0.507	0.380 0.217	0.293 0.228
	[4.002] [0.662]	[0.917] [0.181]	[3.440] [0.576]	[1.329] [0.265]	[1.300] [0.252]	[1.755] [0.343]
	[1.719] [0.454]	[0.404] [0.134]	[2.015] [0.536]	[0.476] [0.175]	[0.533] [0.185]	[0.595] [0.227]
	0.261 0.028	0.290 0.021	0.268 0.028	0.300 0.022	0.094 0.013	0.063 0.023
T=500	0.095 0.120	0.092 0.112	0.095 0.119	0.090 0.114	0.064 0.062	0.058 0.091
	[1.765] [0.306]	[0.403] [0.084]	[1.759] [0.304]	[0.411] [0.087]	[0.595] [0.125]	[0.690] [0.155]
	[1.047] [0.305]	[0.229] [0.079]	[1.050] [0.307]	[0.231] [0.082]	[0.329] [0.118]	[0.449] [0.166]
Scenario 2: Coefficients with different smoothness.						
	1.213 0.141	1.174 0.169	1.073 0.128	0.917 0.309	0.519 0.140	0.794 0.253
T=200	3.090 1.702	3.195 1.767	3.277 1.778	3.417 2.194	3.848 1.588	3.804 1.388
	[3.462] [0.627]	[0.760] [0.193]	[3.655] [0.660]	[0.827] [0.221]	[1.088] [0.266]	[1.846] [0.428]
	[1.967] [0.461]	[0.493] [0.146]	[1.853] [0.438]	[0.532] [0.165]	[0.737] [0.218]	[0.974] [0.303]
	0.654 0.038	0.686 0.056	0.587 0.034	0.585 0.117	0.387 0.043	1.141 0.070
T=500	2.561 1.377	2.562 1.341	2.464 1.313	1.973 1.071	2.859 1.623	1.076 0.617
	[1.514] [0.289]	[0.323] [0.083]	[1.575] [0.299]	[0.345] [0.095]	[0.407] [0.104]	[0.926] [0.190]
	[1.223] [0.333]	[0.289] [0.091]	[1.204] [0.328]	[0.332] [0.108]	[0.367] [0.127]	[0.760] [0.227]

This table presents the results, in a matricial form (equation i, coefficient j), for the $MRBias^2_{ij}$ and the corresponding $MRVar_{ij}$ in square brackets. The regressors across equations are orthogonal and the errors are homoscedastic, but with different specifications for the contemporaneous correlation: constant correlation of 0.9 in Panel A and time-varying correlation in Panel B. Inside each panel results are presented accounting for both scenarios: equal and different degrees of smoothness in the coefficients. tv-SRE indicates that the regression equations are estimated separately whereas tv-SURE means that they are estimated as a system of equations bound by the variance-covariance matrix that takes into account their contemporaneous correlation. Bandwidths are selected in different ways: $\hat{h}_1$ and $\hat{h}_2$ are selected by cross validation of the tv-SRE, for each equation independent with $\hat{h}$ their mean. Bandwidths $\tilde{h}_1$ and $\tilde{h}_2$ are selected by cross validation of the tv-SURE and $\tilde{h}$ is their mean. Two sample sizes are simulated T=200, 500 with $R = 1000$ replications.

setting, with a constant error variance, possibly with time-varying contemporaneous correlation and equal bandwidths, the tv-SURE is more efficient than the tv-SRE.

The bandwidths in the tv-SRE($\hat{h}_1, \hat{h}_2$) are chosen by minimizing (6) for each equation i independently. Since the bandwidth used in the tv-SRE($\hat{h}$) is the average between $\hat{h}_1$ and $\hat{h}_2$, $\hat{h}$ will undersmooth one of the equations, reducing its estimation bias and increasing its variance; while it will oversmooth the other equation resulting in a bias increase and a variance reduction. However, this behaviour does not hold for the pairs [tv-SURE($\hat{h}$), tv-SURE($\hat{h}_1, \hat{h}_2$)], [tv-SURE($\tilde{h}$), tv-SURE($\tilde{h}_1, \tilde{h}_2$)] and [tv-SURE($\hat{h}_1, \hat{h}_2$), tv-SURE($\tilde{h}_1, \tilde{h}_2$)]. This is due to the influence of the contemporaneous correlation in the estimation of the tv-SURE.

For the variance, we see that it is always larger for the tv-SRE($\hat{h}_1, \hat{h}_2$) than for the tv-SURE($\hat{h}_1, \hat{h}_2$) and the tv-SURE ($\tilde{h}_1, \tilde{h}_2$). Also, when comparing methodologies with a common bandwidth for all equations, the tv-SURE($\hat{h}$) and the tv-SURE($\tilde{h}$) have smaller variances than the tv-SRE($\hat{h}$). In other words, for different types of bandwidth selection and orthogonal regressors, Zellner is also satisfied in the time-varying coefficients context. Regarding the bias, we see that there are no clear results, paritcularly when we use two different bandwidths.

In general, when it comes to comparing the results for the tv-SURE using either $\tilde{h}$, the average of the two optimal bandwidths of the tv-SURE, or $\hat{h}$, there are no large differences. On the other hand, the general performance seems to provide more stable results when using a common bandwidth, as expected from the theoretical results.

Table 3 presents the simulation results for the equal regressors across equations case in both equations and it is organized exactly as Table 2. The main result from this experiment is that the tv-SRE($\hat{h}$) and tv-SURE ($\hat{h}$) are exactly equal, as in Zellner: for equal regressors, a common bandwidth and stationary errors, the tv-SRE and tv-SURE are identical. As indicated in *Remark 4.*, there is no gain in estimating the equations jointly in this case, not even in the presence of high or time-varying contemporaneous correlation. When different bandwidths are used, we could find different results. However, the use of different bandwidths for each equation gives no advantage to the tv-SURE.

Table 4 presents the simulation results for the Feasible tv-SURE which estimates the errors variance-covariance matrix in the first step to plug it in the tv-SURE in the second step. All cases use the common bandwidth $\hat{h}$, so values of the $MRBias$ and $MRVar$ in this table must be compared with results of the tv-SURE($\hat{h}$) in Tables 2-3. The errors variance-covariance matrix is estimated nonparametrically with bandwidth selection made by cross-validation. The first two blocks consider constant contemporaneous correlation and others considers the time-varying contemporaneous correlation.

Table 4 also features results of the orthogonal and equal regressors cases and the two scenarios and sample sizes considered before. The comparison of the results in Table 4 and the corresponding results of the tv-SURE($\tilde{h}$) in Table 2 and Table 3, allows the effect of estimating the errors matrix to be analysed.

Table 3: Mean relative bias squared and mean relative variance under equal regressors.

Panel A: Constant contemporaneous correlation ($\rho = 0.9$)

	tv-SRE ($\hat{h}$)		tv-SURE ($\hat{h}$)		tv-SRE ($\hat{h}_1,\hat{h}_2$)		tv-SURE ($\hat{h}_1,\hat{h}_2$)		tv-SURE ($\tilde{h}$)		tv-SURE ($\tilde{h}_1,\tilde{h}_2$)	
Scenario 1: Coefficients with similar smoothness.												
T=200	0.811	0.085	0.811	0.085	0.247	0.043	0.615	0.303	0.438	0.060	0.316	0.194
	0.395	0.394	0.395	0.394	0.541	0.669	1.240	1.636	0.317	0.271	0.815	1.012
	[1.221]	[0.228]	[1.221]	[0.228]	[1.796]	[0.330]	[1.535]	[0.292]	[1.520]	[0.281]	[1.803]	[0.34]
	[5.417]	[1.451]	[5.417]	[1.451]	[4.317]	[1.174]	[6.244]	[1.709]	[6.688]	[1.801]	[6.696]	[1.821]
T=500	0.465	0.043	0.465	0.043	0.137	0.017	0.419	0.177	1.244	0.091	0.962	0.330
	0.210	0.199	0.210	0.199	0.406	0.391	0.879	0.911	0.469	0.463	1.602	1.883
	[0.534]	[0.114]	[0.534]	[0.114]	[0.771]	[0.163]	[0.677]	[0.147]	[0.428]	[0.091]	[0.570]	[0.122]
	[2.568]	[0.763]	[2.568]	[0.763]	[2.036]	[0.613]	[3.024]	[0.914]	[2.002]	[0.611]	[2.521]	[0.783]
Scenario 2: Coefficients with different smoothness.												
T=200	1.344	0.103	1.344	0.103	0.733	0.058	1.116	0.337	1.746	0.135	1.416	0.393
	3.005	1.769	3.005	1.769	4.025	2.462	5.590	4.082	3.645	2.195	6.504	4.793
	[1.260]	[0.235]	[1.260]	[0.235]	[1.655]	[0.305]	[1.623]	[0.302]	[1.147]	[0.212]	[1.544]	[0.284]
	[8.993]	[2.145]	[8.993]	[2.145]	[7.258]	[1.767]	[10.78]	[2.637]	[8.060]	[1.948]	[10.371]	[2.548]
T=500	0.574	0.065	0.574	0.065	0.111	0.022	0.395	0.189	1.601	0.123	0.900	0.386
	2.418	1.390	2.418	1.390	2.733	1.622	2.840	1.987	2.878	1.707	4.743	4.192
	[0.505]	[0.108]	[0.505]	[0.108]	[0.794]	[0.170]	[0.622]	[0.134]	[0.415]	[0.088]	[0.537]	[0.112]
	[3.568]	[0.945]	[3.568]	[0.945]	[2.699]	[0.734]	[3.711]	[1.042]	[2.811]	[0.770]	[3.095]	[0.883]

Panel B: Time-varying contemporaneous correlation ($\rho_t = \exp(-0.5(t/T)^2) - 0.2$)

	tv-SRE ($\hat{h}$)		tv-SURE ($\hat{h}$)		tv-SRE ($\hat{h}_1,\hat{h}_2$)		tv-SURE ($\hat{h}_1,\hat{h}_2$)		tv-SURE ($\tilde{h}$)		tv-SURE ($\tilde{h}_1,\tilde{h}_2$)	
Scenario 1: Coefficients with similar smoothness.												
T=200	0.851	0.095	0.851	0.095	0.959	0.101	1.917	0.162	1.246	0.118	2.824	0.224
	0.392	0.407	0.392	0.407	0.374	0.375	0.743	0.449	0.447	0.512	1.005	0.572
	[3.433]	[0.608]	[3.433]	[0.608]	[3.315]	[0.589]	[4.702]	[0.823]	[3.087]	[0.552]	[4.252]	[0.756]
	[1.843]	[0.481]	[1.843]	[0.481]	[1.917]	[0.500]	[2.463]	[0.637]	[1.664]	[0.441]	[2.340]	[0.604]
T=500	0.205	0.021	0.205	0.021	0.228	0.023	0.439	0.042	1.442	0.097	2.51	0.859
	0.118	0.101	0.118	0.101	0.108	0.092	0.255	0.120	0.507	0.518	1.172	1.341
	[2.026]	[0.396]	[2.026]	[0.396]	[1.971]	[0.385]	[2.686]	[0.528]	[1.219]	[0.239]	[1.805]	[0.339]
	[1.100]	[0.334]	[1.100]	[0.334]	[1.136]	[0.345]	[1.452]	[0.437]	[0.629]	[0.204]	[0.804]	[0.256]
Scenario 2: Coefficients with different smoothness.												
T=200	1.528	0.114	1.528	0.114	1.759	0.133	3.309	0.252	1.261	0.094	4.038	0.307
	3.257	1.947	3.257	1.947	2.796	1.678	3.222	1.960	2.788	1.672	4.063	1.873
	[2.683]	[0.450]	[2.683]	[0.450]	[2.545]	[0.426]	[3.955]	[0.703]	[2.954]	[0.495]	[4.150]	[0.727]
	[2.103]	[0.496]	[2.103]	[0.496]	[2.306]	[0.544]	[3.588]	[0.840]	[2.366]	[0.561]	[3.935]	[0.921]
T=500	0.485	0.058	0.485	0.058	0.313	0.045	2.191	0.749	0.714	0.072	3.065	0.202
	2.477	1.396	2.477	1.396	2.258	1.367	1.714	1.366	2.241	1.357	6.374	1.642
	[1.368]	[0.270]	[1.368]	[0.270]	[1.522]	[0.300]	[2.049]	[0.383]	[1.325]	[0.260]	[1.852]	[0.369]
	[0.914]	[0.242]	[0.914]	[0.242]	[0.826]	[0.221]	[1.191]	[0.321]	[0.887]	[0.245]	[1.469]	[0.390]

This table presents the results, in a matricial form (equation i, coefficient j), for the $MRBias^2_{ij}$ and the corresponding $MRVar_{ij}$ in square brackets. The regressors across equations are equal and the errors are homoscedastic, but with different specifications for the contemporaneous correlation: constant correlation of 0.9 in Panel A and time-varying correlation in Panel B. Inside each panel results are presented accounting for both scenarios: equal and different degrees of smoothness in the coefficients. tv-SRE indicates that the regression equations are estimated separately whereas tv-SURE means that they are estimated as a system of equations bound by the variance-covariance matrix that takes into account their contemporaneous correlation. Bandwidths are selected in different ways: $\hat{h}_1$ and $\hat{h}_2$ are selected by cross validation of the tv-SRE, for each equation independent with $\hat{h}$ their mean. Bandwidths $\tilde{h}_1$ and $\tilde{h}_2$ are selected by cross validation of the tv-SURE and $\tilde{h}$ is their mean. Two sample sizes are simulated T=200, 500 with $R = 1000$ replications.

Table 4: Mean relative bias squared and mean relative variance for the Feasible tv-SURE estimator.

	Constant correlation.				Time-varying correlation.			
	Orthogonal reg.		Equal reg.		Orthogonal reg.		Equal reg.	
Scenario 1: Coefficients with similar smoothness.								
T=200	1.074	0.083	0.811	0.085	1.140	0.073	0.851	0.095
	0.528	0.412	0.395	0.394	0.515	0.447	0.392	0.407
	[0.284]	[0.061]	[1.221]	[0.228]	[0.914]	[0.181]	[3.433]	[0.608]
	[1.208]	[0.424]	[5.417]	[1.451]	[0.398]	[0.133]	[1.843]	[0.481]
T=500	0.280	0.022	0.465	0.043	0.289	0.021	0.205	0.021
	0.124	0.123	0.210	0.199	0.091	0.113	0.118	0.101
	[0.136]	[0.029]	[0.534]	[0.114]	[0.404]	[0.084]	[2.026]	[0.396]
	[0.597]	[0.208]	[2.568]	[0.763]	[0.228]	[0.078]	[1.100]	[0.334]
Scenario 2: Coefficients with different smoothness.								
T=200	1.166	0.105	1.344	0.103	1.176	0.159	1.528	0.114
	4.267	1.886	3.005	1.769	3.211	1.754	3.257	1.947
	[0.209]	[0.057]	[1.260]	[0.235]	[0.833]	[0.205]	[2.683]	[0.450]
	[1.541]	[0.461]	[8.993]	[2.145]	[0.479]	[0.149]	[2.103]	[0.496]
T=500	0.708	0.054	0.574	0.065	0.672	0.046	0.485	0.058
	2.625	1.288	2.418	1.390	2.526	1.343	2.477	1.396
	[0.099]	[0.027]	[0.505]	[0.108]	[0.361]	[0.092]	[1.368]	[0.270]
	[0.650]	[0.212]	[3.568]	[0.945]	[0.292]	[0.090]	[0.914]	[0.242]

This table presents the results, in a matricial form (equation i, coefficient j), for the $MRBias_{ij}^2$ and the corresponding $MRVar_{ij}$ in square brackets. Errors are homoscedastic but present different specifications for the contemporaneous correlation which in the left block column is assumed to be constant and time-varying for the right one. Inside each block column results are presented according to the cases of orthogonal and equal regressors differentiated by subcolumns and also to both scenarios, equal and different smoothness degree for the coefficients. The feasible tv-SURE estimator is used with a common bandwidth h. Elements of the variance and covariance matrix of the errors are estimated by smoothing residuals and using cross validation for the bandwidths involved. For each context two samples are used T=200, 500 and $R = 1000$ replications.

Apart from a higher bias and variance error criteria due to the higher level of uncertainty, it can be noted that the performance of the Feasible tv-SURE estimator is very similar the performance of the tv-SURE.

We see that the estimation of a (Feasible) tv-SURE model is systematically more efficient than the estimation of a tv-SRE, except for the unique case of equal regressors, for which there is no gain. In summary, we can say that the tv-SURE outperforms the tv-SRE when a common bandwidth is used and the regressors are clearly different across equations. For other cases, we recommend the use of the tv-SRE.

4   Empirical application: Time-varying five-factor model

This section is concerned with the application of the tv-SURE to asset pricing in an international framework. Recently, several studies have argued that the three-factor model by Fama and French (1993) does not explain the whole variation in average returns. Thus, Novy-Marx (2013) show that there is a difference in expected returns for high profitability companies in comparison to weak profitability companies, while Aharoni et al. (2013) show that companies with higher capital investment also have higher expected returns. To improve the

expected returns forecast, Fama and French (2015) added two new factors that measure the differences in profitability (robust and weak) and investment (conservative and aggressive), creating their five-factor model (FF5F). Later Fama and French (2017) use the FF5F to analyse the international market mainly in its local version, that is, returns are intend to be explained by the factors of the same geographical zone. The model is as follows,

$$R_{it} - RF_{it} = a_i + b_i\ (RM_{it} - RF_{it}) + s_i\ SMB_{it} + h_i\ HML_{it} + r_i\ RMW_{it} + c_i\ CMA_{it} + \epsilon_{it} \quad i = 1, ..., N \quad t = 1, ...T, \tag{8}$$

where R_{it} is the return of the asset of certain portfolio for region/market i at time t, RF is the risk free return rate and RM represents the return of the market portfolio. The rest of factors are calculated as differences between the returns on diversified portfolios of small and big stocks for SMB, high and low book-to-market stocks for HML, robust and weak operating profitability for RMW and stocks that have low and high investment capabilities due to their conservative or aggressive characteristic respectively for CMA. Finally ϵ is the error term, N the number of international markets and T the sample size.

Theoretically, if model (8) captures the whole variation in excess returns, the intercept a_i should be zero for all i. Nonetheless Fama and French (2017) show empirically that the FF5F model still fails to account for the full variation of the excess returns since the estimation provides intercepts statistically significant. In particular, the average excess returns of small stocks and low profitability firms that invest aggressively are not totally explained.

Here, we propose the time-varying FF5F, an extension of Fama and French (2015) and Fama and French (2017) with coefficients, variances and contemporaneous correlations possibly varying over time. The objective is to evaluate how the sensitivities of expected returns to each risk factor evolve over time. In addition, the use of cross markets information in the tv-SURE might result in a better adjustment of the confidence intervals. The time-varying five-factor model is specified as follows,

$$R_{it} - RF_{it} = a_{it} + b_{it}\ (RM_{it} - RF_{it}) + s_{it}\ SMB_{it} + h_{it}\ HML_{it} + r_{it}\ RMW_{it} + c_{it}\ CMA_{it} + \epsilon_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T. \tag{9}$$

where

$$E(\epsilon_{it}\epsilon_{js}) = \begin{cases} \sigma_{iit} = \sigma_{it}^2 & i = j, \quad t = s \\ \sigma_{ijt} & i \neq j, \quad t = s \\ 0 & t \neq s \end{cases}$$

The data set contains the five factors from four different international markets: North America (NA) with the US and Canada; Japan (JP); Europe (EU) consisting of Austria, Belgium, Denmark, Finland, France, Germany Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom; and Asia Pacific (AP) with Australia, New Zealand, Hong Kong, and Singapore (excluding Japan). For the dependent variable, we

have selected the excess returns of portfolios formed on size and book-to-market, obtained from the Kenneth R. French data library. The period runs from July 1990 to August 2016 and it has a monthly frequency. We have focussed only on the following four portfolios: Small/Low, Small/High, Big/Low and Big/High.

A time-varying estimator of the sensitivities will provide the evolution of the sensitivities to the risk factors along time. Moreover, since the risk factors are different across markets the tv-SURE will take advantage of the possible remaining correlation between errors. Moreover, the procedure will allows us to obtain restricted estimators to test certain restrictions, as will be discussed in the next section.

The time-varying coefficients estimates for the four portfolios are shown in Figures 2 to 7 for the European market. The European market is used as an example to show the possibilities for dynamic investment strategies of a time varying model like (9). Figures for the rest of the markets are not included to avoid a very long article, although there are available from the authors upon request. The cross-restrictions among markets is studied in Section 5.

In the whole estimation process, the variance-covariance matrix has been estimated nonparametrically, as in the simulation study of the feasible tv-SURE. The respective bandwidths have always been selected using cross-validation. The 95% pointwise confidence intervals have been calculated using bootstrap, equivalent to the algorithm in Chen et al. (2017).

Figure 2 presents the estimated time-varying intercepts and their 95% pointwise bootstrap confidence interval of the European market portfolios. Results are for the portfolios Small/Low (top left), Small/High (top right), Big/Low (bottom left) and Big/High (bottom right). It may be observed that the intercept for the Big/Low portfolio is different from zero during the whole sample period, meaning that the five factors do not explain the total variation in expected returns. For the portfolios Small/Low (top left) and Big/High (bottom right) the dynamics of the intercept evolve in a very similar way, approaching zero in the last two decades. Therefore, it seems that the time-varying FF5F is explaining the variation in expected returns in the European market better during the present century.

Figure 3 shows the estimates of coefficients b_t in the European market. They account for the sensitivity to the market risk of each portfolio at time t. It may be observed that all marginal effects have positive confidence intervals, close to one. The relation between the expected returns and the market expected returns seems to be constant for the Big/Low portfolio. On the other hand, there seems to be a negative correlation between the sensitivities in portfolio Small/High and portfolios Small/Low and Big/High. The former's partial effect decreases at the end of the nineties and slightly increases after the Global Financial Crisis (GFC), while the coefficient does the opposite in the latter two portfolios.

Figure 4 shows the estimates of the sensitivities to the SMB risk factor in the European market. This accounts for the spread in returns of small companies over big companies. Clearly, small capitalization firms outperform big capitalization firms. As expected, the effect of size in portfolios with small capitalization stocks have a positive significant effect on the excess returns. On

17

Figure 2: Time-varying estimates of a_t from the time-varying five-factor model in the European market.
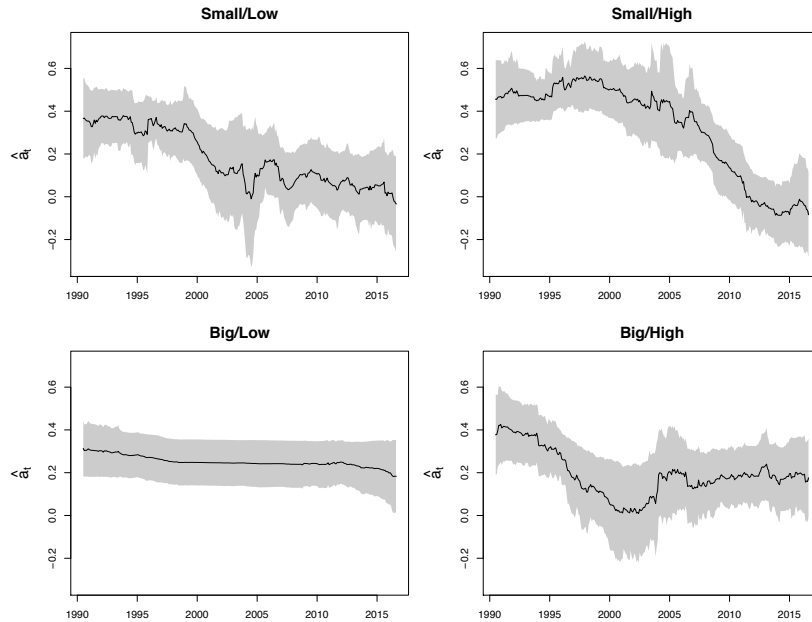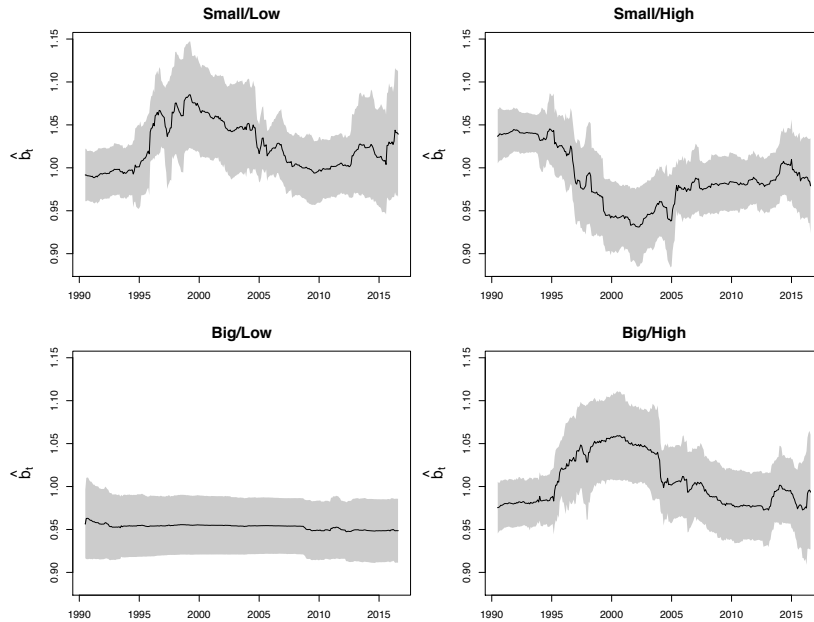

Time-varying pointwise estimates for the intercepts are in continuous lines and the bands represent the bootstrap 95% pointwise confidence interval corresponding to portfolios Small/Low (top left), Small/High (top right), Big/Low (bottom left) and Big/High (bottom right). Results are only shown for the European market, although the rest are available from the authors upon request.

the contrary, the effect is significantly negative in portfolios of big capitalization firms. Here, all coefficients are fairly constant, with the Big/Low being the most stable over time.

Figure 5 shows the estimates of the sensitivities to the book-to-market ratio (HML) that aims to identify undervalued or overvalued companies. As expected, portfolios with high book-to-market companies have a significant positive effect on excess returns while companies with a low book-to-market ratio have a significant negative effect. Again, the coefficients in the Big/Low portfolio seem to be constant for the European market. The sensitivities in the other portfolios vary over time. In particular, their values have increased slightly after the GFC.

Since Novy-Marx (2013) showed evidence of profitability premium, investors have taken this element into account when trading stocks. Profitable firms tend to be growth firms which increase in value very quickly. Figure 6 shows the time varying effects of the RMW factor for the four European portfolios. Estimated coefficients in the Big/Low portfolio are constant and positive, meaning that companies with robust profitability perform better than the average. For

18

Figure 3: Time-varying estimates of b_t from the time-varying five-factor model in the European market.
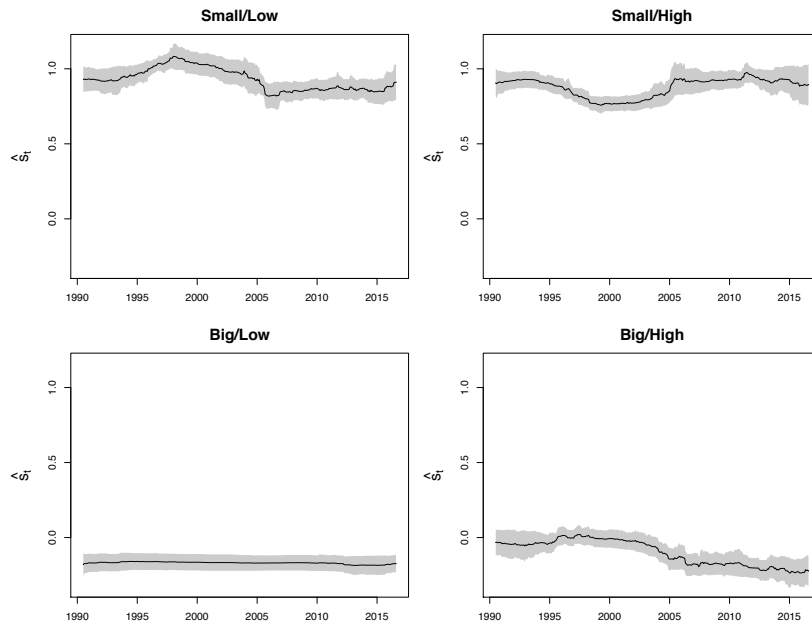

Time-varying pointwise estimates for the coefficients of the excess market return are in continuous lines and the bands represent the bootstrap 95% pointwise confidence interval corresponding to portfolios Small/Low (top left), Small/High (top right), Big/Low (bottom left) and Big/High (bottom right). Results are only shown for the European market, although the rest are available from the authors upon request.

the Small/High portfolio the effect is negative at the beginning of the sample, rising to around zero to later take a drop just before the GFC and steadily increasing during the last decade. This means that robust profitability companies of Small/High portfolios have recently been good investments in Europe. Conversely, profitable companies in Small/Low and Big/High portfolios have followed the opposite path over time. Meaning that weak profitability companies of these two portfolios are performing better than the average.

Finally, Figure 7 shows the effect of the investment factor (CMA). The question is whether it makes sense to hold an asset of a company that has announced great investments in, for example, a new project. This of course depends on other factors such as its size and book-to-market ratio. The effect of CMA in Big/Low stocks expected returns are borderline over zero, while stocks in Small/Low and Big/High portfolios with high investments do result in expected returns under the average with very large variances during the last decade. Only expected returns in the Small/High portfolio have benefited from great investments during the period of 1995-2005.
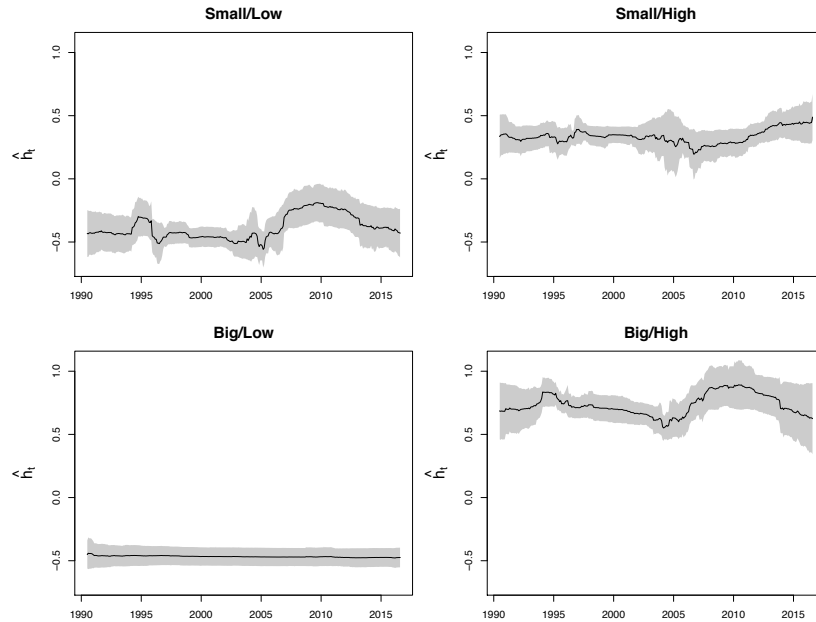
Figure 4: Time-varying estimates of s_t from the time-varying five-factor model in the European market.

Time-varying pointwise estimates for the coefficients of SMB are in continuous lines and the bands represent the bootstrap 95% pointwise confidence interval corresponding to portfolios Small/Low (top left), Small/High (top right), Big/Low (bottom left) and Big/High (bottom right). Results are only shown for the European market, although the rest are available from the authors upon request.

In summary, the use of a tv-SURE in portfolio management provides two new insights with respect to the constant estimators. On the one hand, the tv-SURE allows the use of the information contained in the correlation structure between the different international markets. On the other hand, it provides a description of the dynamics of each risk factor sensitivity during the sample period. As a general result, we observe that our results are coherent with the results obtained in Fama and French (2017). Their estimates are around the mean of our time varying estimates. This mean might be interpreted as the long-run effect, as defined in Ang and Kristensen (2012). In addition, the tv-SURE gives more detailed information, we see that the expected returns sensitivities to the risk factors are very stable for the Big/Low portfolio in the European market, while they change with time for the other three portfolios. This makes the returns of the Big/Low portfolios the most predictable. Moreover, we see a very similar pattern (in terms of variability over time) for Small/Low and Big/High coefficients. Therefore, some effects are still constant, but other effects change over time and the tv-SURE is able to perceive these dynamics. Although

Figure 5: Time-varying estimates of h_t from the time-varying five-factor model in the European market.
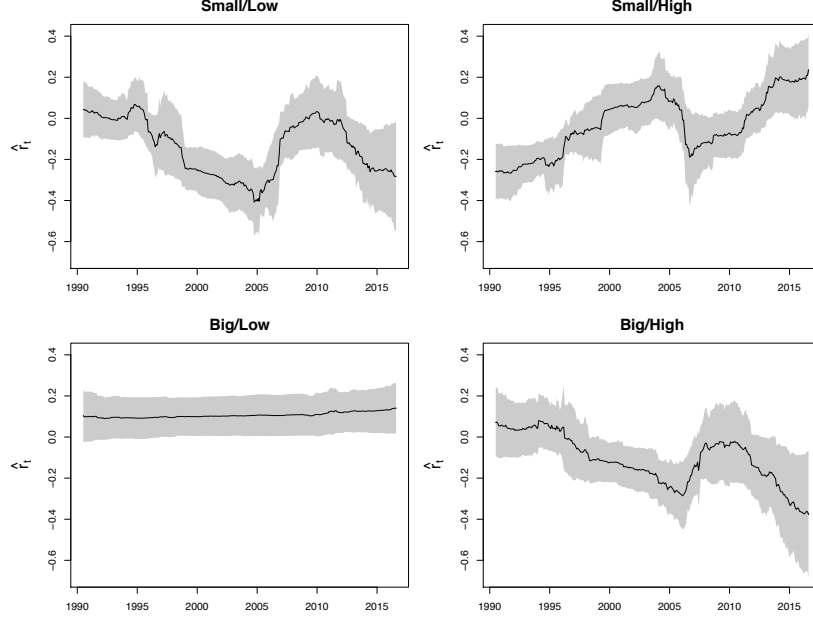

Time-varying pointwise estimates for the coefficients of HML are in continuous lines and the bands represent the bootstrap 95% pointwise confidence interval corresponding to portfolios Small/Low (top left), Small/High (top right), Big/Low (bottom left) and Big/High (bottom right). Results are only shown for the European market, although the rest are available from the authors upon request.

a long term investment might not be affected by the choice of a time-varying model, short-term strategies will definitely benefit from a more precise estimate at each period of time.

5 Estimation and testing under cross-linear restrictions

In the empirical analysis, we have estimated the sensitivities of each portfolio expected returns to the risk factors for the European market. A natural question is to check whether these sensitivities are the same across regions. For instance, we might ask if the effect of excess market returns are the same for European and North American Small/Low portfolios. One of the main aspects of the tv-SURE is the possibility of performing estimations with cross-restrictions due to its nature. The estimation procedure considers a set of general linear restrictions given by,

Figure 6: Time-varying estimates of r_{it} from the time-varying five-factor model in the European market.

Time-varying pointwise estimates for the coefficients of RMW are in continuous lines and the bands represent the bootstrap 95% pointwise confidence interval (CI) corresponding to portfolios Small/Low (top left), Small/High (top right), Big/Low (bottom left) and Big/High (bottom right). Results are only shown for the European market, although the rest are available from the authors upon request.

$$R_t\beta_t = r_t \qquad t = 1, 2, \ldots, T \qquad (10)$$

where R_t is the $(q \times P)$ matrix accounting for the linear relations between the betas, r_t is a q order vector and $q < P$ is the number of restrictions considered. Note that we are allowing for cross-restrictions to vary over time.

The restricted tv-SURE estimator minimizes

$$\min \sum_{s=1}^{T} (Y_s - X_s\beta_t)^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} (Y_s - X_s\beta_t)$$
$$\text{s.t. } R_t\beta_t = r_t. \qquad (11)$$

The Lagrangian function associated with this constrained optimization prob-

Figure 7: Time-varying estimates of c_{it} from the time-varying five-factor model in the European market.
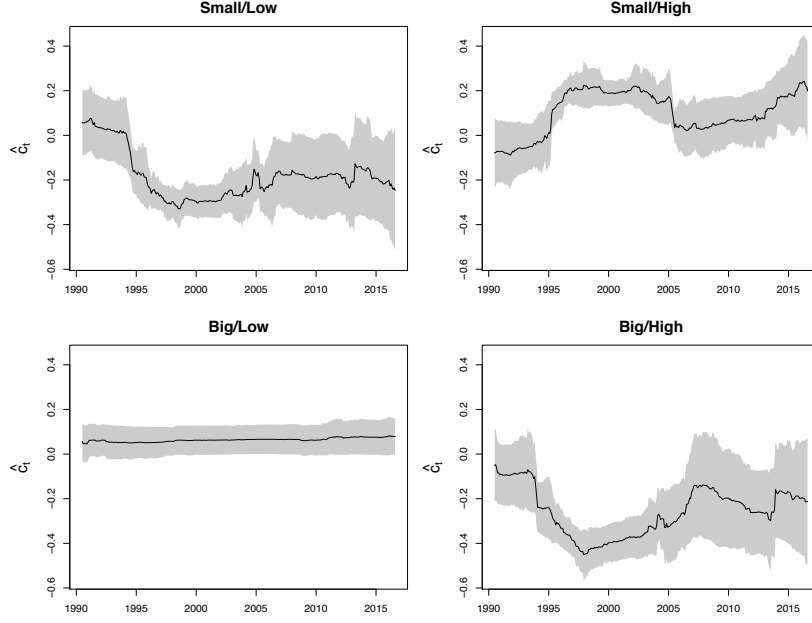
Time-varying pointwise estimates for the coefficient of CMA are in continuous lines and the bands represent the bootstrap 95% pointwise confidence interval (CI) corresponding to portfolios Small/Low (top left), Small/High (top right), Big/Low (bottom left) and Big/High (bottom right). Results are only shown for the European market, although the rest are available from the authors upon request.

lem is defined, in matrix notation, as,

$$\mathcal{S}^R(\beta_t, \lambda_t) = \sum_{s=1}^{T} (Y_s - X_s\beta_t)^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} (Y_s - X_s\beta_t) + 2\lambda_t^T (R_t\beta_t - r_t) \quad (12)$$

where λ_t is a $(q \times 1)$ vector that contains the Lagrange multipliers. Let us denote by $\hat{\beta}_t^R$ and $\widehat{\lambda}_t^R$ the solution to the constrained optimization problem

$$(\hat{\beta}_t^R, \widehat{\lambda}_t^R) \equiv \operatorname{argmin}_{\beta_t, \lambda_t} \mathcal{S}^R(\beta_t, \lambda_t). \quad (13)$$

A closed expression for the vector containing the coefficient estimators will be obtained from the first order conditions in (13) as

$$\hat{\beta}_t^R = \hat{\beta}_t - \hat{G}_t^{-1} R_t^T \left[R_t \hat{G}_t^{-1} R_t^T \right]^{-1} (R_t\hat{\beta}_t - r_t) \quad (14)$$

where $\hat{\beta}_t$ is the unconstrained tv-SURE estimator derived in (3), that is, the one

obtained when no restrictions are imposed (i.e. $\lambda_t = 0$), and $\hat{G}_t$ is the matrix in Assumption A.6.

The estimator for the time-varying Lagrange multiplier is also derived from the first order conditions,

$$\hat{\lambda}_t = \left[R_t M_t R_t^T\right]^{-1} (R_t \hat{\beta}_t - r_t) \tag{15}$$

which measures the cost of imposing non-true restrictions. Thus, its value increases in the same direction as the difference $R_t \hat{\beta}_t - r_t$. The sufficient conditions for a unique solution of the optimization problem in (13) are guaranteed by assumptions *A.1*, *A.5* and *A.6*. Nevertheless, in the restricted tv-SURE, assumption *A.6* may be substituted by *Assumption* A.6' below.

Assumption A.6' At each time t, matrix

$$\hat{G}_t + R_t^T R_t \tag{16}$$

is positive definite and uniformly bounded from above and below.

The main asymptotic results of the restricted tv-SURE are the following stated.

Theorem 4 *Under assumptions A.1–A.7, the asymptotic bias and variance for the restricted estimator are*

$$Bias(\hat{\beta}_t^R) = \frac{d_k}{2}(I - Q_t R_t) G_t^{-1} \mathbb{H} G_t \left(\beta_t'' + 2G_t' \beta_t'\right) + Q_t(r_t - R_t Q \beta_t) + o\left(h^2\right) \tag{17}$$

and

$$V(\hat{\beta}_t^R) = \frac{c_k}{T}(I - Q_t R_t)(\mathbb{H} G_t)^{-1}(I - Q_t R_t)^T + O((Th)^{-1}) \tag{18}$$

where $Q_t = G_t^{-1} R_t^T \left(R_t G_t^{-1} R_t^T\right)^{-1}$

The estimator is not consistent when the restriction does not hold, since the bias remains as a nonzero constant, even for large T. The next corollary provides the differences in bias and variance between the tv-SURE and the restricted tv-SURE.

Corollary 5 *Given the assumptions in Theorem 4,*

$$Bias(\hat{\beta}_t) - Bias(\hat{\beta}_t^R) = \frac{d_k}{2} Q_t R_t G_t^{-1} \mathbb{H} \left(\frac{\partial^2 \beta_t}{\partial t \partial t} G_t + 2(G_t') \frac{\partial \beta_t}{\partial t}\right)$$

if and only if the restriction is fulfilled, $R_t \beta_t = r_t$. For the variances,

$$V(\hat{\beta}_t) - V(\hat{\beta}_t^R) = \frac{c_k}{T} \left[(\mathbb{H} G_t)^{-1} - (I - Q_t R_t)(\mathbb{H} G_t)^{-1}(I - Q_t R_t)^T\right],$$

Remark 6 *There is no clear advantage in employing the restricted estimator when using different bandwidths for each equation. The gain of the restricted tv-SURE over the unrestricted tv-SURE is only comparable for a common bandwidth. Similarly to the comparison of the tv-SRE and the tv-SURE which was only possible for a common bandwidth. The following corollary summarizes this result.*

Corollary 7 *Given the assumptions in Theorem 4, and if a common h is used,*

$$V(\hat{\beta}_t) - V(\hat{\beta}_t^R) = \frac{c_k}{Th} \left[G_t^{-1} - (I - Q_t R_t)(\mathbb{H} G_t)^{-1}(I - Q_t R_t)^T \right],$$

is a positive semidefinite matrix.

5.1 Testing

In general, restrictions may be justified by a theoretical model and/or the need to test it. For instance, it may be of interest to analyse the statistical equality of the sensitivities related to a given risk factor across different regions. The confidence intervals in Figure 8 show that it is not always clear when betas for two different regions are equal for the whole sample period. To account for the whole curve of betas, we propose the Kolmogorv-Smirnov-type test developed in Ferreira and Gil-Bazo (2004).

Assume, without loss of generality, a system of two equations and two regressors per equation,

$$y_{1t} = \beta_{11t} x_{11t} + \beta_{12t} x_{12t} + u_{1t}, \quad Var(u_{1t}) = \sigma_{1t}^2$$
$$y_{2t} = \beta_{21t} x_{21t} + \beta_{22t} x_{22t} + u_{2t}.$$

Consider testing for the restriction with null hypothesis $H_0 : \beta_{11t} = \beta_{21t}$, for all t versus the alternative hypothesis $H_a : \beta_{11t} > \beta_{21t}$ for some t. This is equivalent to $H_a : \beta_{11t} = \beta_{21t} + \alpha_t$, with $\alpha_t > 0$ for some t.

Under the null ($\beta_{11t} = \beta_{21t}$) and assuming that the coefficients are known, the distribution of the partial sums of length n defined as $S_n(s) = \sqrt{T}^{-1} \sum_{t=1}^{Ts} \{y_{1t} - \beta_{21t} x_{11t} - \beta_{12t} x_{12t}\} = \sqrt{T}^{-1} \sum_{t=1}^{Ts} \{u_{1t}\}$ is known. In the previous formula, T is the length of the our series and $s \in [0,1]$ is a fraction of it. Using Donsker's theorem in Billingsley (1968), $S_n(s) \sim_D B(V(s))$ where B is a Brownian motion with variance $V(s) = \int_0^s \sigma_1^2(u) du$. Kolmogorov-Smirnov type statistics for the –greater than– and –two sided– alternatives are defined as,

$$KS_1 = \max_{0 \le s \le 1} S_n(s) \sim_D \sqrt{V(1)} \max_{0 \le v \le 1} B(v), \tag{19}$$

$$KS_2 = \max_{0 \le s \le 1} |S_n(s)| \sim_D \sqrt{V(1)} \max_{0 \le v \le 1} |B(v)| \tag{20}$$

Figure 8: Time-varying five-factor model estimates for different markets and risk factors.
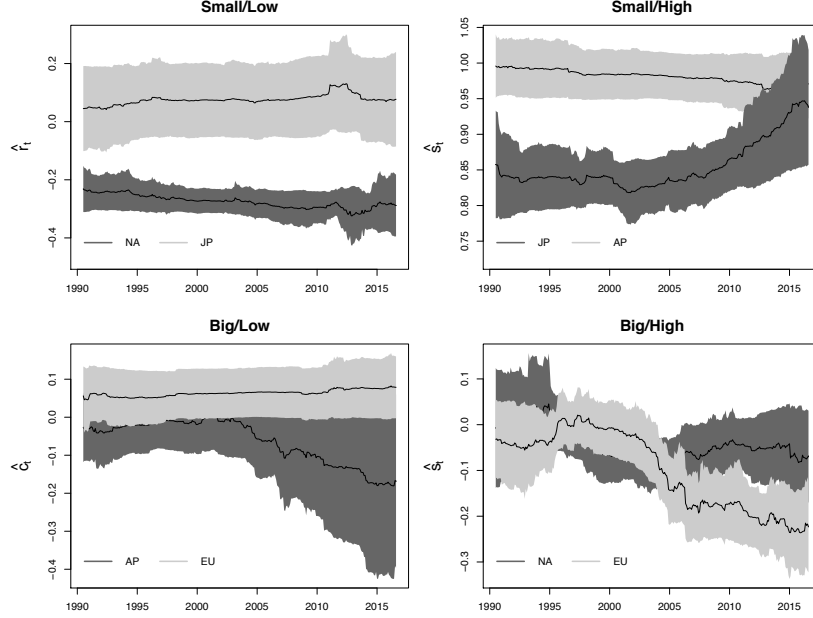
This figure shows a representation of some evolutions analysed for different coefficients in different regions. Time-varying pointwise estimates for the coefficients of RMW corresponding to portfolios Small/Low compare the North American and Japanese markets (top left), HML for Small/High portfolios in the Japanese and European markets (top right), CMA for Big/Low Asia in the Pacific and European markets (bottom left), and SMB corresponding to Big/High portfolios of the North American and European markets (bottom right). The bands represent the bootstrap 95% pointwise confidence interval.

respectively. The asymptotic distributions of the first statistic is known. Provided by the reflection principle of Wiener processes,

$$P\left(\frac{1}{\sqrt{V(1)}}\max_{0\leq v\leq 1}B(v)>1\right)=2P\left(\frac{1}{\sqrt{V(1)}}B(1)>1\right)=2P(Z>1)$$

for Z a standard normal random variable. The distribution of KS_2 can be easily tabulated.

Since the time-varying coefficients are unknown, the distribution of estimated partial sums $\hat{S}_n(s)=\sqrt{T}^{-1}\sum_{t=1}^{Ts}\{\hat{u}_{1t}\}$ is distorted from the true distribution of $S_n(s)$. To overcome this caveat, as in Ferreira and Gil-Bazo (2004), a uniformly subsample from the $\hat{u}_{1t}$ is drawn to calculate $\hat{S}_n(s)$.

Table 5 shows the p-values of the described test when applied to pairs of betas from the application in Section 4. The null hypothesis is $H_0 : \beta_{ji,t}=\beta_{jm,t}\ \forall t$

and the alternative hypothesis is $H_a : \beta_{ij,t} \neq \beta_{im,t}$, where j refers to the risk factor, $i \neq m$ refers to the region and t is time. Only p-values below 0.05 are shown. The first thing to notice is that differences between regions occur mainly for portfolios Small/Low and Big/High. Often, the sensitivities from the Japanese market are statistically different to those from the Asia Pacific and European markets. In addition, for the Big/High portfolio the North America market is different from the other three. Figure 8 presents the evolution of several coefficients with the 95% confidence interval and provides a way to compare the coefficients graphically.

The formal KS test may answer the question of whether there is any period in which two markets behave differently. A further study of the power of the test to be a better support for investment decisions will be relegated to further research.

Table 5: Results of a two sided KS-type test. The p-values of the test are displayed in brackets.

	Small/Low	Smal/High	Big/Low	Big/High
RM-RF	JP $\neq$ AP (0.03) JP $\neq$ EU (0.02)			NA $\neq$ JP (0.03) NA $\neq$ AP (0.03) JP $\neq$ AP (0.05)
SMB	JP $\neq$ AP (0.006) JP $\neq$ EU (0.03)	JP $\neq$ AP (0.02)	JP $\neq$ AP (0.02)	NA $\neq$ AP (0.01) NA $\neq$ EU (0.05) JP $\neq$ AP (0.008)
HML	JP $\neq$ AP (0.03) JP $\neq$ EU (0.02)			NA $\neq$ AP (0.02) JP $\neq$ AP (0.008) JP $\neq$ EU (0.03)
RMW	NA $\neq$ AP (0.01)			NA $\neq$ JP (0.002) NA $\neq$ AP (0.006) NA $\neq$ EU (0.004)
CMA	NA $\neq$ JP (0.05) NA $\neq$ AP (0.05) JP $\neq$ AP (0.03) JP $\neq$ EU (0.03)	JP $\neq$ AP (0.03) AP $\neq$ EU (0.03)	JP $\neq$ AP (0.03) AP $\neq$ EU (0.02)	NA $\neq$ AP (0.004) JP $\neq$ AP (0.03) JP $\neq$ EU (0.03)

Results from testing pairwise betas from different markets and equal factors.

6 Conclusion

We have proposed a nonparametric estimator to account for time-varying coefficients in a Seemingly Unrelated Regression Equations model (tv-SURE estimator). We have proved that, in general, Zellner's results can be extended to this framework when using a common bandwidth for all equations. At first glance, this might be a shocking result, since the selection of one bandwidth for each equation provides more flexibility in the estimation procedure. However, the theoretical and the simulation studies show how a multiple selection leads to worse results. Roughly speaking, the use of one different bandwidth for each equation seems to distort the way in which the tv-SURE estimator binds the whole system of equations. The asymptotic results show the role of the bandwidth and where the gain in variance comes from. The simulation study

supports the advantage of using a tv-SURE over a tv-SRE when the covariates are not closely related and there is correlation among the errors.

We have applied this methodology to the five factors model considered in Fama and French (2015) and Fama and French (2017) with portfolio data from four international markets: North America, Europe, Asia Pacific and Japan. The results describe the dynamics of the excess returns sensitivity to each risk factor across different portfolios and markets. Generally speaking, portfolios classified as Big/High and Small/Low show the greatest changes over time and portfolio, while Big/Low portfolios seem to have the more stable sensitivities to the risk factors.

Finally, the tv-SURE provides the possibility to estimate models with linear cross-restrictions, as well as their testing. As general results, we find that 1) the returns in the North American market respond to changes in the five risk factors very differently to returns in the other markets, when investing in Big/High portfolios; 2) the Japanese returns respond differently than the Asia Pacific and European returns to all risk factors, except for the RMW, when investing in Small/Low portfolios; and 3) not many differences have been found among markets on the other two portfolios, Small/High and Big/Low.

Acknowledgement

References

Aharoni, G., Grundy, B., and Zeng, Q. Stock returns and the miller modigliani valuation formula: Revisiting the fama french analysis. *Journal of Financial Economics*, 110:347–357, 2013.

Ando, T. and Bai, J. Asset pricing with a general multifactor structure. *Journal of Financial Econometrics*, 13:556–604, 2015.

Ang, A. and Kristensen, D. Testing conditional factor models. *Journal of Financial Economics*, 106:132–156, 2012.

Aretz, K., Bratram, S., and Pope, P. Macroeconomic risks and characteristic-based factor models. *Journal of Banking & Finance*, 34:1383–1399, 2010.

Aslanidis, N. and Casas, I. Nonparametric correlation models for portfolio allocation. *Journal of Banking & Finance*, 37:2268–2283, 2013.

Billingsley, P. *Convergence of probability measures*. Wiley Series in probability and Mathematical Statistics: Tracts on probability and statistics. Wiley, 1968.

Cai, Z., Li, Q., and Park, J. Y. Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148:101–113, 2009.

Cai, Z., Ren, Y., and Yang, B. A semiparametric conditional capital asset pricing model. *Journal of Banking and Finance*, 61:117–126, 2015.

Cai, Z. Trending time-varying coefficient time series with serially correlated errors. *Journal of Econometrics*, 136:163–188, 2007.

Chen, X. B., Gao, J., Li, D., and Silvapulle, P. Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. *Journal of Business & Economic Statistics*, online:1–13, 2017.

Dahlhaus, R. Fitting time series models to nonstationary processes. *Annals of Statistics*, 25:1–37, 1997.

Dahlhaus, R. A likelihood approximation for locally stationary processes. *Annals of Statistics*, 28:1762–1794, 2000.

Das, M. Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics*, 124:335–361, 2005.

Esteban, M. V., Ferreira, E., and Orbe-Mandaluniz, S. Nonparametric methods for estimating and testing for constant betas in asset pricing models. *Applied Economics*, 47:2577–2607, 2015.

Fama, E. and French, K. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56, 1993.

Fama, E. F. and French, K. R. A five-factor asset pricing model. *Journal of Financial Economics*, 116:1–22, 2015.

Fama, E. F. and French, K. R. International tests of a five-factor asset pricing model. *Journal of Financial Economics*, 123:441–463, 2017.

Ferreira, E. and Gil-Bazo, J. Beyond single-factor affine term structure models. *Journal of Financial Econometrics*, 2:565–591, 2004.

Ferreira, E., Gil-Bazo, J., and Orbe, S. Conditional beta pricing models: A nonparametric approach. *Journal of Banking & Finance*, 35:3362–3382, 2011.

French, K. R. Data library. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Accessed: 2016-09-01.

Granger, C. W. Non-Linear Models: Where Do We Go Next - Time Varying Parameter Models? *Studies in Nonlinear Dynamics & Econometrics*, 12:1–11, 2008.

Henderson, D., Kumbhakar, S. C., Li, Q., and Parmeter, C. Smooth coefficient estimation of a seemingly unrelated regression. *Journal of Econometrics*, 189:148–162, 2015.

Kristensen, D. Non-parametric detection and estimation of structural change. *Econometrics Journal*, 15:420–461, 2012.

Novy-Marx, R. The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108:1 – 28, 2013.

Orbe, S., Ferreira, E., and Rodriguez-Poo, J. An algorithm to estimate time-varying parameter sure models under different types of restriction. *Computational Statistics & Data Analysis*, 42:363–383, 2003.

Orbe, S., Ferreira, E., and Rodriguez-Poo, J. Nonparametric estimation of time varying parameters under shape restrictions. *Journal of Econometrics*, 126:53–77, 2005.

Phillips, P. C. B., Li, D., and Gao, J. Estimating smooth structural change in cointegration models. *Journal of Econometrics*, 196:180–195, 2017.

White, H. and Domowitz, I. Nonlinear regression with dependent observations. *Econometrica*, 52:143–161, 1984.

Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348–368, 1962.

APPENDIX

Proof of Theorem 1

This theorem is a generalization of Theorem 1 in Ferreira et al. (2011). The main differences lie in: 1) the data matrix X_s which in this proof is a diagonal matrix of the matrices of each equation; and 2) the smoothing parameter H, which is a diagonal matrix (one value for each equation). The rest of the variables and parameters have the same structure so we just reproduce the main different steps.

We need the following lemma.

Lemma 8

$$\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \longrightarrow (a.s.) \quad G_t$$

$$\sum_{s=1}^{T} X_s^T K_{H,ts} \Sigma_s^{-1} K_{H,ts} X_s \longrightarrow (a.s.) \quad c_K G_t$$

Recall $K_{H,ts} = diag(K_{h_1,ts} \ldots K_{h_M,ts})$, where $K_{h_m,ts} = h_m^{-1} K(\frac{t-s}{Th_m})$.

First, the error committed in the estimation is given by the following expression,

$$
\begin{aligned}
\widehat{\beta}_t - \beta_t &= \left(\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \right)^{-1} \sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} Y_s - \beta_t \\
&= \left(\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \right)^{-1} \sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s (\beta_s - \beta_t) \\
&\quad + \left(\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \right)^{-1} \sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s
\end{aligned}
$$

To overcome the issue of having a random denominator, we redefine the bias and variance expressions substituting $W_t = G_t^{-1} \sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s$. Hence, the redefined bias is $Bias^*(\widehat{\beta}_t) = Bias(W_t \widehat{\beta}_t)$. For technical reasons, we use different bandwidths for W_t^* and for $\widehat{\beta}_t$, say h^* and h respectively, such that the following condition holds,

$$\frac{E \parallel W_t^* - I \parallel^2}{E \parallel \widehat{\beta}_t - \beta_t \parallel^2} = o(1), \tag{21}$$

as T goes to infinity. This condition establishes that W_t^* goes to the identity at a faster rate than the mean squared error goes to zero. The proof continues using G_t^{-1} instead of the random denominator.

i) The Bias

$$Bias^*(\widehat{\beta}_t) \quad = \quad G_t^{-1} \sum_{s=1}^{T} E(X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s)(\beta_s - \beta_t) \qquad (22)$$

$$+ \quad G_t^{-1} \sum_{s=1}^{T} E(X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s) \qquad (23)$$

$$= \quad G_t^{-1} \sum_{s=1}^{T} E(X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s)(\beta_s - \beta_t), \qquad (24)$$

since $E(X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s | X_s) = 0$.

Consider the $P \times 1$ order vector

$$Z_t = \sum_{s=1}^{T} E(X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s)(\beta_s - \beta_t) \qquad (25)$$

Define γ_{ijt} as the generic term of $\Gamma_t = E(X_s^T X_s)$, σ_{ijt} a generic term of Σ_t and σ_t^{ij} for Σ_t^{-1}.

We can write $X_{Mt}^* = K_{H,ts}^{1/2} X_{Mt}$

$$\begin{bmatrix} \sigma_t^{11} X_{1t}^{*T} X_{1t}^* & \cdots & \sigma_t^{1M} X_{1t}^{*T} X_{Mt}^* \\ \sigma_t^{21} X_{2t}^{*T} X_{1t}^* & \cdots & \sigma_t^{2M} X_{2t}^{*T} X_{Mt}^* \\ \cdots & \ddots & \vdots \\ \sigma_t^{M1} X_{Mt}^{*T} X_{1t}^* & \cdots & \sigma_t^{MM} X_{Mt}^{*T} X_{Mt}^* \end{bmatrix} \begin{bmatrix} (\beta_{1s} - \beta_{1t}) \\ \vdots \\ (\beta_{Ms} - \beta_{Mt}) \end{bmatrix}.$$

Note that the kernel matrix weight $K_H(t - s)$ is block diagonal, the matrix containing the moments of the explanatory variables is a kind of weighted cross matrix of variables corresponding to different equations and the final vector stacks subvectors containing the coefficients of different equations. Given the mentioned structure it is easy to obtain that the ith element of equation m is given by,

$$\sum_{s=1}^{T} \sum_{m'=1}^{M} \sum_{j=1}^{P_m} K_{h_m,ts} \sigma_s^{mm'} X_{mis} X_{m'js} (\beta_{m'js} - \beta_{m'jt}) \qquad (26)$$

From (22) and taking into account (26)

$$\sum_{m'=1}^{M} \sum_{j=1}^{P_m} \sum_{s=1}^{T} K_{h_m,ts} \sigma_s^{mm'} E(X_{mis} X_{m'js}) (\beta_{m'js} - \beta_{m'jt})$$

Using the Taylor expansion with $t - s = T h_m \eta_m$, and writing $\gamma_{mi,m'j,t} = E(X_{mit} X_{m'jt})$,

$$\sum_{s=1}^{T} K_{h_m,ts}\sigma_s^{mm'} E(X_{mis}X_{m'js})(\beta_{m'js} - \beta_{m'jt}) = \int K(u_m)\left(\sigma_t^{mm'} - u_m h_m \frac{\partial\sigma_t^{mm'}}{\partial t} + o(h_m^2)\right)$$

$$\left(\gamma_{mi,m'j,t} - h_m u_m \frac{\partial\gamma_{mi,m'j,t}}{\partial t} + o(h_m^2)\right)\left(-\frac{\partial\beta_{m'jt}}{\partial t} h_m u_m + \frac{1}{2}\frac{\partial^2\beta_{m'jt}}{\partial t\partial t}(h_m u_m)^2 + o(h_m^2)\right) du_m =$$

$$\frac{d_k h_m^2}{2}\left[\sigma_t^{mm'}\gamma_{mi,m'j,t}\frac{\partial\partial\beta_{m'jt}}{\partial t\partial t} + 2\left(\gamma_{mi,m'j,t}\frac{\partial\sigma_t^{mm'}}{\partial t} + \frac{\partial\gamma_{mi,m'j,t}}{\partial t}\sigma_t^{mm'}\right)\frac{\partial\beta_{m'jt}}{\partial t}\right] + o(h_m^2).$$

$$E(Z_{mit}) = \frac{d_k}{2}\sum_{m'=1}^{M} h_m^2 \sum_{j=1}^{P_m}\left[\sigma_t^{mm'}\gamma_{mi,m'j,t}\frac{\partial\partial\beta_{m'jt}}{\partial t\partial t} + 2\left(\gamma_{mi,m'j,t}\frac{\partial\sigma_t^{mm'}}{\partial t}\right.\right.$$

$$+ \left.\left.\frac{\partial\gamma_{mi,m'j,t}}{\partial t}\sigma_t^{mm'}\right)\frac{\partial\beta_{m'jt}}{\partial t}\right] + o\left(\sum_{m=1}^{M} h_m^2\right) = O\left(\sum_{m=1}^{M} h_m^2\right) = O(h^2)$$

where we recall that $\mathbb{H} = diag(h_1 I_{p_1} \ldots h_M I_{p_M}) \neq H$.

$$Bias(\widehat{\beta}_t) = \frac{d_k}{2}G_t^{-1}\mathbb{H}^2\left(\frac{\partial^2\beta_t}{\partial t\partial t}G_t + 2(A_{1t} + A_{2t})\frac{\partial\beta_t}{\partial t}\right) + o(h^2).$$

where

$$A_{1t} = \begin{bmatrix} (\partial\sigma_t^{11}/\partial t)E(X_{1t}^T X_{1t}) & \ldots & (\partial\sigma_t^{1M}/\partial t)E(X_{1t}^T X_{Mt}) \\ (\partial\sigma_t^{21}/\partial t)E(X_{2t}^T X_{1t}) & \ldots & (\partial\sigma_t^{2M}/\partial t)E(X_{2t}^T X_{Mt}) \\ \ldots & \ddots & \vdots \\ (\partial\sigma_t^{M1}/\partial t)E(X_{Mt}^T X_{1t}) & \ldots & (\partial\sigma_t^{MM}/\partial t)E(X_{Mt}^T X_{Mt}) \end{bmatrix}$$

$$= \begin{bmatrix} (\partial\sigma_t^{11}/\partial t)\Gamma_{11t} & \ldots & (\partial\sigma_t^{1M}/\partial t)\Gamma_{1Mt} \\ (\partial\sigma_t^{21}/\partial t)\Gamma_{21t} & \ldots & (\partial\sigma_t^{2M}/\partial t)\Gamma_{2Mt} \\ \ldots & \ddots & \vdots \\ (\partial\sigma_t^{M1}/\partial t)\Gamma_{M1t} & \ldots & (\partial\sigma_t^{MM}/\partial t)\Gamma_{MMt} \end{bmatrix}$$

$$A_{2t} = \begin{bmatrix} \sigma_t^{11}\,\partial E(X_{1t}^T X_{1t})/\partial t & \ldots & \sigma_t^{1M}\,\partial E(X_{1t}^T X_{Mt})/\partial t \\ \sigma_t^{21}\,\partial E(X_{2t}^T X_{1t})/\partial t & \ldots & \sigma_t^{2M}\,\partial E(X_{2t}^T X_{Mt})/\partial t \\ \ldots & \ddots & \vdots \\ \partial\sigma_t^{M1}\,\partial E(X_{Mt}^T X_{1t})/\partial t & \ldots & \sigma_t^{MM}\,\partial E(X_{Mt}^T X_{Mt})/\partial t \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_t^{11}\,\partial\Gamma_{11t}/\partial t & \ldots & \sigma_t^{1M}\,\partial\Gamma_{1Mt}/\partial t \\ \sigma_t^{21}\,\partial\Gamma_{21t}/\partial t & \ldots & \sigma_t^{2M}\,\partial\Gamma_{2Mt}/\partial t \\ \ldots & \ddots & \vdots \\ \partial\sigma_t^{M1}\,\partial\Gamma_{M1t}/\partial t & \ldots & \sigma_t^{MM}\,\partial\Gamma_{MMt}/\partial t \end{bmatrix}$$

Thus,

$$||Bias(\widehat{\beta}_t)||^2 = \frac{d_k^2}{4}\left\|G_t^{-1}\mathbb{H}^2\left(\frac{\partial^2\beta_t}{\partial t\partial t}G_t + 2(A_{1t}+A_{2t})\frac{\partial\beta_t}{\partial t}\right)\right\|_2^2 + o\left(\sum_{m=1}^M h_m^4\right).$$

Note that

$$\frac{\partial G_t}{\partial t} = \frac{\partial E(X_t^T\Sigma_t^{-1}X_t)}{\partial t} = \left(\Psi_t \circ \frac{\partial E(X_t^T X_t)}{\partial t}\right) + \left(\frac{\partial\Psi_t}{\partial t}\circ E(X_t^T X_t)\right)$$

where

$$\Psi_t = \begin{bmatrix} \sigma_t^{11}\,J_{p_1,p_1} & \cdots & \sigma_t^{1M}\,J_{p_1,p_M} \\ \sigma_t^{21}\,J_{p_2,p_1} & \cdots & \sigma_t^{2M}\,J_{p_2,p_M} \\ \cdots & \ddots & \vdots \\ \partial\sigma_t^{M1}\,J_{p_M,p_1} & \cdots & \sigma_t^{MM}\,J_{p_M,p_M}, \end{bmatrix}$$

J_{p_m,p'_m} is the matrix of order $p_m \times p'_m$ whose elements are all equal to 1. and $\circ$ denotes the Hadamard product.

Then,

$$||Bias(\widehat{\beta}_t)||^2 = \frac{d_k^2}{4}\left\|G_t^{-1}\mathbb{H}^2\left(G_t\frac{\partial^2\beta_t}{\partial t\partial t} + 2\frac{\partial G_t}{\partial t}\frac{\partial\beta_t}{\partial t}\right)\right\|_2^2 + o\left(\sum_{m=1}^M h_m^4\right).$$

and the result follows.
ii) Variance
The variance term is given by

$$
\begin{aligned}
Var(\widehat{\beta}_t) &= Var(\widehat{\beta}_t - \beta_t) = \\
&= Var\left[\left(\sum_{s=1}^{} X_s^T K_{H,ts}^{1/2}\Sigma_s^{-1}K_{H,ts}^{1/2}X_s\right)^{-1}\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2}\Sigma_s^{-1}K_{H,ts}^{1/2}X_s(\beta_s - \beta_t)\right. \\
&\quad \left. + \left(\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2}\Sigma_s^{-1}K_{H,ts}^{1/2}X_s\right)^{-1}\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2}\Sigma_s^{-1}K_{H,ts}^{1/2}u_s\right],
\end{aligned}
$$

and using the redefined variance term, $Var^*(\widehat{\beta}_t) = Var(W_t^*\widehat{\beta}_t)$, it follows

$$Var^*(\widehat{\beta}_t) = Var\left[G_t^{-1}\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2}\Sigma_s^{-1}K_{H,ts}^{1/2}X_s(\beta_s - \beta_t)\right] + Var\left[G_t^{-1}\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2}\Sigma_s^{-1}K_{H,ts}^{1/2}u_s\right]$$

(27)

As proved in Ferreira et al. (2011), the leading term is

$$G_t^{-1} Var \left[\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s \right] G_t^{-1} \tag{28}$$

and following similar arguments as there,

$$Var \left[\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s \right] = E \left[\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s \left(\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s \right)^T \right]$$

$$= E \left[\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s \sum_{s=1}^{T} u_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \right]$$

Using iterated expectations and the fact that $E(u_{ms} u_{m's'} | X) = 0$, for all $s \neq s'$, the expectation above is equal to

$$= E \left[\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s u_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \right]$$

$$= E \left[\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} \Sigma_s K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \right]$$

$$= E \left[\sum_{s=1}^{T} X_s^T \Sigma_s^{-1/2} K_{H,ts} \Sigma_s^{-1/2} \Sigma_s \Sigma_s^{-1/2} K_{H,ts} \Sigma_s^{-1/2} X_s \right]$$

$$= E \left[\sum_{s=1}^{T} X_s^T \Sigma_s^{-1/2} K_{H,ts} K_{H,ts} \Sigma_s^{-1/2} X_s \right]$$

$$= \sum_{s=1}^{T} E \left[X_s^T K_{H,ts} \Sigma_s^{-1/2} K_{H,ts} X_s \right]$$

$$= c_k (T \mathbb{H})^{-1} G_t + o((T \mathbb{H})^{-1})$$

given results from Lemma 6.1.

Thus taking into account expression (27), the asymptotic variance is given by,

$$Var(\widehat{\beta}_t) = \frac{c_k}{T} G_t^{-1} \mathbb{H}^{-1} G_t G_t^{-1} + o((Th)^{-1}) =$$

$$\frac{c_k}{T} (\mathbb{H} G_t)^{-1} + o((Th)^{-1})$$

$$V(\widehat{\beta}_t) = \frac{c_k}{T} tr(\mathbb{H} G_t)^{-1} + o((Th)^{-1})$$

35

Proof of Theorem 3

Consider the sequence of variables Z_t defined as

$$Z_t = \sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} u_s$$

Theorem 2.4 in White and Domowitz (1984) applies and, since the bias is negligible with respect to the variance, the result is straightforward.

Proof of Theorem 4

Note that $\widehat{\beta}_t^R$ can be written as $(I - \hat{Q}_t R_t)\hat{\beta}_t - \hat{Q}_t r_t$, where $\hat{Q}_t = \hat{G}_t^{-1} R_t^T [R_t \hat{G}_t^{-1} R_t^T]^{-1}$ and $\hat{G}_t = \left(\sum_{s=1}^{T} X_s^T K_{H,ts}^{1/2} \Sigma_s^{-1} K_{H,ts}^{1/2} X_s \right)$.

Then, we can write $\widehat{\beta}_t^R = (I - \hat{Q}_t R_t)\hat{\beta}_t - \hat{Q}_t r_t$, and the result follows straightforward from Theorem 1.

Research Papers
2017

CREATES
Center for Research in Econometric
Analysis of Time Series

2017-16: Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris: Does the ARFIMA really shift?

2017-17: Massimo Franchi and Søren Johansen: Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles

2017-18: Matias D. Cattaneo, Michael Jansson and Kenichi Nagasawa: Bootstrap-Based Inference for Cube Root Consistent Estimators

2017-19: Daniel Borup and Martin Thyrsgaard: Statistical tests for equal predictive ability across multiple forecasting methods

2017-20: Tommaso Proietti and Alessandro Giovannelli: A Durbin-Levinson Regularized Estimator of High Dimensional Autocovariance Matrices

2017-21: Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Variance swap payoffs, risk premia and extreme market conditions

2017-22: Jakob Guldbæk Mikkelsen: Testing for time-varying loadings in dynamic factor models

2017-23: Roman Frydman, Søren Johansen, Anders Rahbek and Morten Nyboe Tabor: The Qualitative Expectations Hypothesis: Model Ambiguity, Concistent Representations of Market Forecasts, and Sentiment

2017-24: Giorgio Mirone: Inference from the futures: ranking the noise cancelling accuracy of realized measures

2017-25: Massimiliano Caporin, Gisle J. Natvik, Francesco Ravazzolo and Paolo Santucci de Magistris: The Bank-Sovereign Nexus: Evidence from a non-Bailout Episode

2017-26: Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Decoupling the short- and long-term behavior of stochastic volatility

2017-27: Martin M. Andreasen, Jens H.E. Christensen and Simon Riddell: The TIPS Liquidity Premium

2017-28: Annastiina Silvennoinen and Timo Teräsvirta: Consistency and asymptotic normality of maximum likelihood estimators of a multiplicative time-varying smooth transition correlation GARCH model

2017-29: Cristina Amado, Annastiina Silvennoinen and Timo Teräsvirta: Modelling and forecasting WIG20 daily returns

2017-30: Kim Christensen, Ulrich Hounyo and Mark Podolskij: Is the diurnal pattern sufficient to explain the intraday variation in volatility? A nonparametric assessment

2017-31: Martin M. Andreasen, Jens H.E. Christensen and Glenn D. Rudebusch: Term Structure Analysis with Big Data

2017-32: Timo Teräsvirta: Nonlinear models in macroeconometrics

2017-33: Isabel Casas, Eva Ferreira and Susan Orbe: Time-varying coefficient estimation in SURE models. Application to portfolio management