

A multilevel model for repeated cross-sectional data with stochastic volatility

Lucia Modugno, Silvia Cagnone and Simone Giannerini

Abstract In this paper we propose a multilevel approach for the analysis of repeated cross-sectional data that exhibit volatility effects. We treat individuals as clustered within time-points so that the dynamics over time is modelled at the second level. Items sold in auction present a structure like that of repeated cross-sectional surveys since different goods are sold at different time-points. For prices of artworks, as well as for other assets (financial, insurance, etc.) the hypothesis of constant volatility appears unreasonable. In this work we combine a multilevel model with autoregressive random effects and a stochastic volatility model in order to account for the kurtosis and the volatility pattern of prices. We apply the model to Tribal art auction prices and show improvement over existing proposals both in terms of fit and forecasting.

Abstract *Il presente lavoro propone un approccio multilivello per l'analisi di dati cross-section ripetuti che presentano volatilità. Si considerano gli individui raggruppati per istante temporale cosicché la dinamica temporale è modellata al secondo livello. Gli oggetti venduti all'asta presentano una struttura di cross-section ripetute in quanto beni differenti vengono banditi in asta in momenti diversi. Per i prezzi di opere d'arte, come per altri tipi di asset, l'ipotesi di volatilità costante nel tempo è troppo restrittiva. Noi combiniamo un modello multilevel con effetti casuali autoregressivi e un modello con volatilità stocastica per tener conto dell'andamento della volatilità dei prezzi d'asta. Applicando il modello proposto a prezzi d'asta di opere d'arte tribale si ottiene un migliore adattamento ai dati e una migliore previsione rispetto ai modelli tradizionali.*

Lucia Modugno

Department of Statistical Sciences, University of Bologna, Via delle Belle Arti 41, 40126, Bologna, Italy, e-mail: lucia.modugno@unibo.it

Silvia Cagnone

Department of Statistical Sciences, University of Bologna, Via delle Belle Arti 41, 40126, Bologna, Italy e-mail: silvia.cagnone@unibo.it

Simone Giannerini

Department of Statistical Sciences, University of Bologna, Via delle Belle Arti 41, 40126, Bologna, Italy e-mail: simone.giannerini@unibo.it

Key words: Multilevel model, Repeated cross-sectional data, Stochastic volatility.

1 Introduction

The main aim of this work is to propose a multilevel approach ([8]) for repeated cross-sectional data. Differently from longitudinal data, repeated cross-sectional data consist of observations on individual survey respondents drawn from the same context (e.g. the same country) at many different time-points, and can therefore be treated as clustered within time-points ([3]). In this collection of data there are not specific individuals followed over time; nevertheless, there is often a time dynamics that manifests itself in different ways. To our knowledge, the analysis of repeated cross-sectional data in a multilevel framework is not well established both from the theoretical and the practical point of view. Most specifications are ad hoc solutions with no available software so that such models result poorly developed and rarely applied.

In [6] we proposed a specification that combines the multilevel framework and autoregressive effects for this kind of data by treating individuals as level-1 units and time-points as level-2 units. Clearly, the key trait of such specification is the possibility of modelling time effects by allowing serial correlation among level-2 units (time-points). Such specification has been successfully applied for modelling auction prices of artworks. Now we extend the framework and combine the multilevel model with Autoregressive Random Effects (ARE) with the stochastic volatility model. The proposal is able to account for kurtosis and volatility patterns that are often present in auction price dynamics. In fact, items sold in auctions present a structure like that of repeated cross-sectional surveys since different and strongly heterogeneous baskets of goods are sold in different time-points. For prices of heterogeneous investment goods, like artworks or financial assets, the hypothesis of constant volatility over time is not reasonable. Clearly, series of financial assets show high persistence in the autocorrelation of squared observations and leptokurtosis. Two classes of models that are often used to model and forecast unobserved volatility in asset returns are the ARCH-type models ([2]) and the stochastic volatility model ([9]).

The present work is motivated by the analysis of the first world database of Tribal art prices. Such database, built by a team of researchers of the University of Bologna (Rimini campus), in conjunction with other institutions, contains information on over 20000 artwork items auctioned by the most important auction houses from 1998 to 2011. We apply the proposed model on Tribal art prices and compare it with the traditional hedonic regression model ([7]).

2 A multilevel model with a stochastic volatility component

The multilevel model with time series components in [6] fits repeated cross-sectional data by considering units grouped in time-points. It relaxes the usual assumption of independence among random effects of the multilevel model and treats them as a time series at the second level. Consider a random intercept model with k level-1 covariates:

$$y_{it} = \beta_{0t} + \mathbf{x}_{it}^T \boldsymbol{\beta} + \varepsilon_{it}, \quad \varepsilon_{it} | \mathbf{x}_{it}^T \sim \text{NID}(0, \sigma^2)$$

for $i = 1, \dots, n_t$ and $t = 1, \dots, T$. The slopes $\boldsymbol{\beta}$'s are fixed slopes and the β_{0t} 's are time-specific random intercepts accounting for the heterogeneity between individuals within each time point. The variability of the time-specific random intercepts over time can be modeled as follows

$$\beta_{0t} = \beta_0 + u_t,$$

where u_t represents the deviation of the group-specific intercept β_{0t} from the overall mean, β_0 . The assumption of independence among random effects of the classic multilevel model is relaxed by assuming an autoregressive process of order 1 for level-2 errors (ARE - Autoregressive Random Effects):

$$u_t = \rho u_{t-1} + \eta_t, \quad \eta_t | \mathbf{X}_t \sim \text{NID}(0, \sigma_\eta^2),$$

with $|\rho| < 1$, that guarantees stationarity, $\eta_t \perp u_s$ and $\eta_t \perp \varepsilon_{it}$ for all $s < t$ and for all i .

We further extend the ARE model by including the log-normal stochastic volatility model in the multilevel structure (SVARE - Stochastic Volatility and Autoregressive Random Effects) as follows:

$$\begin{aligned} y_{it} &= \beta_0 + u_t + \mathbf{x}_{it}^T \boldsymbol{\beta} + \exp(h_t/2) \varepsilon_{it} \\ u_t &= \rho u_{t-1} + \eta_t \\ h_t &= \alpha + \delta h_{t-1} + \sigma_v v_t \\ \text{for } i &= 1, \dots, n_t \text{ and } t = 1, \dots, T, \end{aligned}$$

for $i = 1, \dots, n_t$ and $t = 1, \dots, T$. Both the volatility component h_t and the random effect u_t are latent variables with autoregressive behaviour assumed strictly stationary (therefore $|\rho| < 1$ and $|\delta| < 1$). As initial conditions for the two autoregressive processes, we use the unconditional distributions of h_t and u_t , that is $h_1 \sim N(\frac{\alpha}{1-\delta}, \frac{\sigma_v^2}{1-\delta^2})$ and $u_1 \sim N(0, \frac{\sigma_\eta^2}{1-\rho^2})$. Moreover, it is assumed that

$$\begin{aligned} \varepsilon_{it} &\sim \text{NID}(0, 1), & \eta_t &\sim \text{NID}(0, \sigma_\eta^2), & v_t &\sim \text{NID}(0, 1) \\ \eta_t &\perp \varepsilon_{it} \perp v_t, & \eta_t &\perp u_s, & \forall s < t \end{aligned}$$

Under these assumptions, the conditional densities result

$$\begin{aligned}
y_{it}|u_t, h_t &\sim \text{NID}(\beta_0 + u_t + \mathbf{x}_{it}^T \boldsymbol{\beta}, \exp(h_t)) \\
u_t|u_{t-1} &\sim \text{NID}(\rho u_{t-1}, \sigma_\eta^2) \\
h_t|h_{t-1} &\sim \text{NID}(\alpha + \delta h_{t-1}, \sigma_v^2).
\end{aligned}$$

We perform a full maximum likelihood estimation method by treating this model as a nonlinear state space model (see [4]) and by maximizing the likelihood obtained by integrating the latent variables out of the joint density of the responses. The likelihood takes the following form

$$\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{y}) &= \int_{\mathbf{h}} \int_{\mathbf{u}} f(\mathbf{y}|\mathbf{u}, \mathbf{h}) f(\mathbf{u}, \mathbf{h}) d\mathbf{u} d\mathbf{h} = \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(\mathbf{y}_1|u_1, h_1) f(u_1) f(h_1) \\
&\quad \prod_{t=2}^T f(\mathbf{y}_t|u_t, h_t) f(u_t|u_{t-1}) f(h_t|h_{t-1}) du_T \dots du_1 dh_T \dots dh_1
\end{aligned}$$

where $f(\mathbf{y}_t|u_t, h_t) = \prod_{i=1}^{n_t} f(y_{it}|u_t, h_t)$ for $t = 1, \dots, T$.

Computing the $2T$ -dimensional integral is a challenging task. However, the likelihood may be approximated by applying an iterated numerical integration procedure introduced in [5] for non-Gaussian filtering problems. In particular, we use the Gauss-Legendre quadrature rules (see [4]). The resulting approximated likelihood function is

$$\begin{aligned}
\tilde{L}(\boldsymbol{\theta}|\mathbf{y}) &= \left(\frac{b-a}{2}\right)^T \left(\frac{e-d}{2}\right)^T \sum_{i_1}^{n_u} w_{ui_1} \sum_{j_1}^{n_h} w_{hj_1} \left[\prod_{i=1}^{n_1} f(y_{i1}|u_{i_1}^*, h_{j_1}^*) \right] f(u_{i_1}^*) f(h_{j_1}^*) \\
&\quad \sum_{i_2}^{n_u} w_{ui_2} \sum_{j_2}^{n_h} w_{hj_2} \left[\prod_{i=1}^{n_2} f(y_{i2}|u_{i_2}^*, h_{j_2}^*) \right] f(u_{i_2}^*|u_{i_1}^*) f(h_{j_2}^*|h_{j_1}^*) \dots \\
&\quad \dots \sum_{i_T}^{n_u} w_{ui_T} \sum_{j_T}^{n_h} w_{hj_T} \left[\prod_{i=1}^{n_T} f(y_{iT}|u_{iT}^*, h_{j_T}^*) \right] f(u_{iT}^*|u_{iT-1}^*) f(h_{j_T}^*|h_{j_{T-1}}^*)
\end{aligned} \tag{1}$$

where $\{u_i^*\}$, with $i = 1, \dots, n_u$, and $\{h_j^*\}$, with $j = 1, \dots, n_h$, are sets of Gauss-Legendre quadrature points, w_{ui} and w_{hj} are the corresponding weights and $[a, b]$ and $[d, e]$ are finite integration limits which replaces the infinite ones for the random effects and the volatility process respectively. The choice of the grids and the number of evaluation points is crucial for the numerical precision. First, as proposed in [4], we center the grids for the two latent processes on $\mu_u = 0$ and $\mu_h = \alpha/(1 - \delta)$ with a width of $3\sigma_u = 3\sigma_\eta/(\sqrt{1 - \rho^2})$ and $3\sigma_h = 3\sigma_v/(\sqrt{1 - \delta^2})$; this allows the grids to cover the support of the unconditional distributions with non negligible mass. Second, the numbers of quadrature points, n_u and n_h , are chosen by considering the degree of smoothness of the integrands, thus such that the average distance between

two points is less or equal to $\sigma_\eta/2$ for the random effect process and $\sigma_v/2$ for the volatility process.

In order to compute and maximize the approximated likelihood Eq. 1, we transform it in a more compact form which is an extension of those in [4] and [1]. Also, we perform filtering and smoothing to achieve optimal estimators of the unobserved-state vectors $\mathbf{u}$ and $\mathbf{h}$, and we show how to obtain one-step-ahead prediction values.

3 Application of the SVARE model

The potential of the proposed model is shown by means of an application to the first database on Ethnic artworks. In literature, art prices are typically modelled through the hedonic regression model, a time-varying fixed-effect model that takes into account the heterogeneity of artworks by explaining prices through object features and constructs a price index by neutralizing the effect of quality. Here we first apply the ARE model that allows to overcome the well-known drawbacks of the fixed-effects regression model. Although the model provides similar estimates and interpretation of results of the hedonic model, the resulting fit is more parsimonious and results in a smaller BIC (see Table 1); also, it provides a decomposition of the total variability of the response in between-time and within-time variability. Finally, the ARE model allows better one-step-ahead forecasting of the responses i.e. the prices of objects that will be sold one semester later. However, the assumption of normality for the (first level) errors of both models is questionable and, similarly to other assets, the residuals show a leptokurtic behaviour. Furthermore, the Levene's test rejects the hypothesis of homogeneity of variances across groups. Hence, we fit the proposed SVARE model to take into account the time-dependent heteroscedasticity of the errors.

Preliminary results of the SVARE model with no covariates, show that, unlike competing models, the SVARE specification fits the trend both of the mean and of the volatility of artwork prices (Figure 1). The resulting predicted volatility values provide useful information on the predictability of the prices that can be exploited by art market stakeholders. Moreover, as shown in Table 1, the SVARE model presents a better fit in terms of information criteria and superior forecasting performance.

Table 1 Loglikelihood and Information criteria for the hedonic regression model (Hed), ARE and SVARE. The results refer to models without covariates.

	Hed	ARE	SVARE
loglik	-14380.460	-14446.162	-14097.464
AIC	28816.920	28900.324	28212.929
BIC	29028.640	28930.570	28280.981

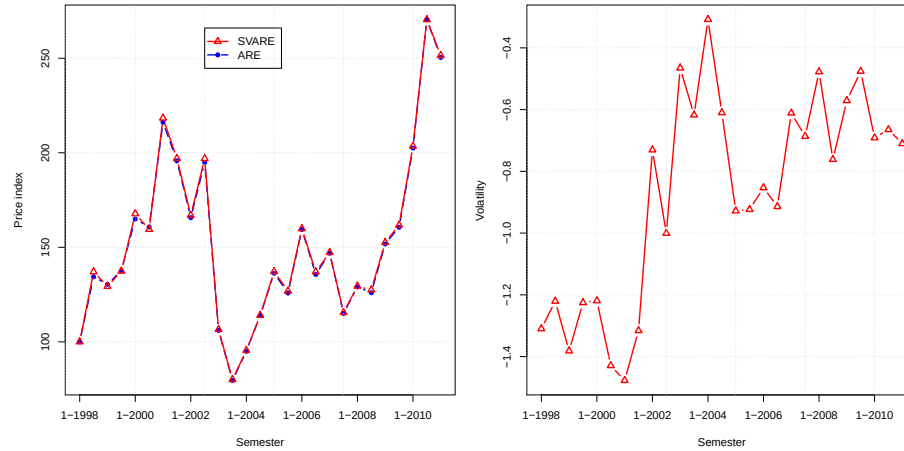


Fig. 1 (Left) Fitted price index for ARE and SVARE models. (Right) Predicted volatility for the SVARE model.

References

1. Bartolucci, F., De Luca, G.: Maximum likelihood estimation of a latent variable time-series model. *Applied Stochastic Models in Business and Industry* **17**, 5–17 (2001)
2. Engle, R.F.: Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50**(4), 987–1007 (1982)
3. Firebaugh, G.: *Analyzing Repeated Surveys*. CA: Sage, Thousand Oaks (1997)
4. Fridman, M., Harris, L.: A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models. *Journal of Business & Economic Statistics* **16**(3), 284–291 (1998)
5. Kitagawa, G.: Non-gaussian state-space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association* **82**, 1032–1063 (1987)
6. Modugno, L., Cagnone, S., Giannerini, S.: The analysis of tribal art prices: a multilevel model with autoregressive components. Available via *ArXiv e-prints*, 1304.4876 (2013)
7. Rosen, S.: Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy* **82**, 34–55 (1974)
8. Skrondal, A., Rabe-Hesketh, S.: *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Chapman & Hall/CRC, New York (2004)
9. Taylor, S.: *Modelling Financial Time Series*. Wiley, New York (1986)