MULTILEVEL MODELS WITH STOCHASTIC VOLATILITY FOR REPEATED CROSS-SECTIONS: AN APPLICATION TO TRIBAL ART PRICES

BY SILVIA CAGNONE[1], SIMONE GIANNERINI AND LUCIA MODUGNO[1]

University of Bologna

In this paper, we introduce a multilevel specification with stochastic volatility for repeated cross-sectional data. Modelling the time dynamics in repeated cross sections requires a suitable adaptation of the multilevel framework where the individuals/items are modelled at the first level whereas the time component appears at the second level. We perform maximum likelihood estimation by means of a nonlinear state space approach combined with Gauss–Legendre quadrature methods to approximate the likelihood function. We apply the model to the first database of tribal art items sold in the most important auction houses worldwide. The model allows to account properly for the heteroscedastic and autocorrelated volatility observed and has superior forecasting performance. Also, it provides valuable information on market trends and on predictability of prices that can be used by art markets stakeholders.

1. Introduction. The investigation of the relationship between the art market and financial markets has important implications for institutions as well as for auction houses, art merchants and individuals. In fact, also due to the recent financial crisis, there has been a sharp increase in the so-called alternative investments that comprise funds specialising in art. These appear to offer a highly beneficial diversification strategy with a complex correlation with traditional assets. Hence, the study of the features of this new asset class is important and cannot disregard the differences with respect to traditional stocks. For instance, art items are exchanged a few times and the transaction costs are considerable. Moreover, there is the so-called *aesthetic dividend* which plays a crucial role [see, e.g., Candela, Castellani and Pattitoni (2013), and references therein, Goetzmann (1993)].

The study of price determination and price indexes for art items is fundamental to auction houses and art merchants. Besides estimating the price of individual items, price indexes can be used to understand market trends, to assess the main social and economic factors that influence the art market, to understand whether investing in art would diversify risk in a long-term investment portfolio. For art

objects, the traditional view of a long run price related to the cost of production does not hold anymore. Some authors argue that the art market is inherently unpredictable since it is dictated by collectors' manias [Baumol (1986)]. However, there is an ever growing consensus, backed by empirical evidence, on the idea that "price fundamentals" can be objectively identified. The hedonic regression, also known as the *grey painting method*, is one of the most used approaches for modelling art prices. It was first proposed in Rosen (1974) and further investigated and applied in Agnello and Pierce (1996), Chanel (1995), Chanel, Gérard-Varet and Ginsburgh (1996), Collins, Scorcu and Zanola (2009), Ginsburgh and Jeanfils (1995), Locatelli Biey and Zanola (2005). According to this method, the price of an artwork item depends both on market trends and on a set of characteristics of the item itself. Such dependence is modelled through a fixed effect regression and the estimated regression coefficients can be interpreted as the price of each feature, the so-called *shadow price*. Hence, it is possible to predict the price of a given object by summing the prices of its features. Also, a time-dependent intercept can represent the value of the *grey painting* in that period, that is, the value of an artwork created by a standard artist, through standard techniques, with standard dimensions, etc. [Locatelli Biey and Zanola (2005)]. Eventually, the price index is built from the prices of the *grey painting* in different periods.

Despite its potential, the hedonic regression model has several shortcomings. First, as also remarked in Goetzmann, Mamonova and Spaenjers (2014) only a small fraction of the great variability of the price dynamics is explained. Second, most of the features are categorical so that the regression equation contains many dummy variables and the resulting models are not parsimonious. Most importantly, the time dynamics is not modelled directly but through dummy variables so that it is not possible to use the model to forecast the prices. Moreover, since it is practically impossible to follow the selling price of each artwork item over time, the available datasets have a structure of a repeated cross-section where at each time point a new sample is observed.

There is an increasing interest in modelling data that have the structure of repeated cross sections due to their ubiquitous appearance. Examples of social surveys with such structure are the *British Social Attitudes Survey*[2] and the UK *Family Expenditure Survey*,[3] the EU *Eurobarometer Surveys*,[4] and all opinion surveys. They are both easier and cheaper to gather than panels and do not suffer from some of the problems that affect the latter, such as unbalance due to attrition or mortality. Even if in repeated cross sections it is not possible to follow specific individuals over time, the (spatio-) temporal dynamics is important and cannot be

---

[2]http://natcen.ac.uk/our-research/research/british-social-attitudes/.
[3]https://discover.ukdataservice.ac.uk/series/?sn=200016.
[4]http://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm/General/index.

disregarded in that it allows to track trends and social changes. A first, straightforward solution to account for the time dynamics is to use some form of aggregation. For instance, many authors simply compute and work with averages over time [e.g., Box-Steffensmeier, De Boef and Lin (2004), MacKuen, Erikson and Stimson (1992), Scott and Smith (1974)]. Also, aggregations can be performed over cohorts of individuals as to obtain pseudo panels; see, for example, Deaton (1985). A sensitive issue in pseudo-panel models is the nonunique choice of the individual features used to construct the cohorts, in that it affects the consistency of estimators. This corresponds to the 'weak instruments' problem discussed, for example, in Bound, Jaeger and Baker (1995), Moffitt (1993). We argue that a more appropriate solution is the multilevel approach that accounts directly for the observed heterogeneity and avoids the loss of information due to aggregation also known as *ecological fallacy* [Goldstein (2010), Skrondal and Rabe-Hesketh (2004)]. However, the treatment of repeated cross-sectional data requires the extension of the classical multilevel specification by considering individual heterogeneity within time at the first level, and the variability over time at the second level. This specification has been adopted for the first time by DiPrete and Grusky (1990) and Browne and Goldstein (2010) in the frequentist and Bayesian frameworks, respectively. Along this line, Lebo and Weber (2015) adopt a simple two-step approach where, at the first step, a time-series model is applied to pooled data; at the second step, the individual heterogeneity is captured by means of a multilevel model applied to the "time-corrected" responses.

Motivated by the construction of a price index for auctioned items of tribal artworks, Modugno, Cagnone and Giannerini (2015) extended the multilevel approach for repeated cross sections by incorporating an autoregressive component at the second level of the mixed effects model. They implemented an EM iterative algorithm that allows to obtain full maximum likelihood parameter estimates and prediction of random effects simultaneously. This avoids data pooling and the loss of efficiency due to considering separate specifications for the individual heterogeneity and the time component. They found a considerable improvement over classical models in terms of prediction and forecasting but the assumption of normality of level-1 errors was violated, probably due to the presence of heteroscedasticity and kurtosis. They devised an ad hoc solution for deriving robust standard errors through the wild bootstrap scheme for multilevel models introduced in Modugno and Giannerini (2015). Similar findings concerning the nonnormality and the heteroscedasticity were reported for other specifications for the art market in Bocart and Hafner (2012) and Hodgson and Vorkink (2004). In Bocart and Hafner (2012), the problem was addressed by estimating a semiparametric time-varying volatility and Student's t error with skewness, whereas Hodgson and Vorkink (2004) did not assume any parametric form for the disturbances but retained the assumption of serial independence of random effects. Also, Bocart and Hafner (2015) modeled the volatility of price indexes by means of a smooth function of time as a component of an unbalanced panel model with AR(1) time effects.

They implemented a linear Gaussian state-space representation and estimated it through maximum likelihood combined with a Kalman filter, and as above, they found a violation of the normality assumption. The evidence reported in literature indicates that the volatility of prices plays a pre-eminent role; assuming it constant is not realistic and might cause estimation problems.

In this paper, we extend the model proposed in Modugno, Cagnone and Giannerini (2015) by including a stochastic volatility component at the second level by means of a nonlinear state-space approach. The specification is motivated by the analysis of the first world database of tribal art prices. This allows to account properly for the heteroscedastic and autocorrelated volatility of level-1 error terms and brings in several advantages. Stochastic volatility (SV) models are based on the assumption that the conditional variance of the observed variable depends on a latent variable that captures the flow of information arriving from the market. Similar to ARCH-type models for financial time series, SV models allow to account properly for fat tailed distributions, white-noise-type dependence, high and lag-decreasing autocorrelations of squared observations. We opt for a stochastic volatility component since ARCH-type models assume that the volatility is affected by past information through a deterministic function. Such a specification is not viable for repeated cross-sections.

Model estimation is performed through maximum likelihood via a non-linear Gaussian filtering process in the spirit of Kitagawa (1987) and Tanizaki and Mariano (1998). The task poses several computational challenges related to the presence of time-varying latent variables that must be integrated out from the likelihood function so that there is no analytical solution. To this aim, Fridman and Harris (1998) proposed a nonlinear Kalman filter algorithm by expressing the likelihood function as a nested sequence of one-dimensional integrals approximated by the Gauss–Legendre numerical quadrature. Bartolucci and De Luca (2001) extended this approach by computing analytical first and second derivatives of the approximated likelihood. They applied a rectangular quadrature to approximate the integrals. More recently, Cagnone and Bartolucci (2017) approximated such integrals by using an adaptive Gauss–Hermite quadrature method. Here, we extend the procedures discussed in Fridman and Harris (1998) and Bartolucci and De Luca (2001) by implementing both the Gauss–Legendre quadrature and the rectangular quadrature methods to approximate the integrals involved in the likelihood. Eventually, we chose the Gauss–Legendre method for the application.

2. The first database of tribal art prices. The first database of tribal art prices was created in 2006 from the joint agreement of the department of Economics of the University of the Italian Switzerland, the Museum of the Extra-European cultures in Lugano, the Museo degli Sguardi in Rimini, and the Faculty of Economics of the University of Bologna, campus of Rimini. For each artwork item, there are 37 variables recorded from the catalogues released before the auctions. The variables include physical, historical, and market characteristics. Some

TABLE 1
Subset of variables classified by type: *Physical*, *Historical*, *and Market*

	Variable	Categories
Physical	Type of object	Furniture, Sticks, Masks, Religious objects, Ornaments, Sculptures, Musical instruments, Tools, Clothing, Textiles, Weapons, Jewels
	Material	Ivory, Wood, Metal, Gold, Stone, Terracotta, ceramic, Silver, Textile and hides, Seashell, Bone, horn, Not indicated
	Patina	Not indicated, Pejorative, Present, Appreciative
Historical	Continent	Africa, America Eurasia, Oceania
	Region	Central Africa, Southern Africa, Western Africa, Eastern Africa, Australia, Indonesia, Melanesia, Polynesia, Northern America, Northern Africa, Southern America, Mesoamerica, Far East, Micronesia, Indian Subcontinent, Southeastern Asia, Middle East
	Illustration on the catalogue	Absent, Black/white, Coloured,
	Illustration width	Absent, Miscellaneous, Quarter page, Half page, Full page, More than one page, Cover
	Description	Absent, Short visual, Visual, Broad visual, Critical, Broad critical
	Specialized bibliography	Yes, No
	Comparative bibliography	Yes, No
	Exhibition	Yes, No
	Historicization	Simple certification, Absent, Museum certification, Relevant museum certification
Market	Venue	New York, Paris,
	Auction house	Sotheby's, Christie's

of these are shown in Table 1 and most of them are categorical. After the auction, the information on the selling price is added to the record.

Figure 1 shows the boxplots of logged prices aggregated by semester; the number of items sold in each semester is reported inside the boxes. The structure of
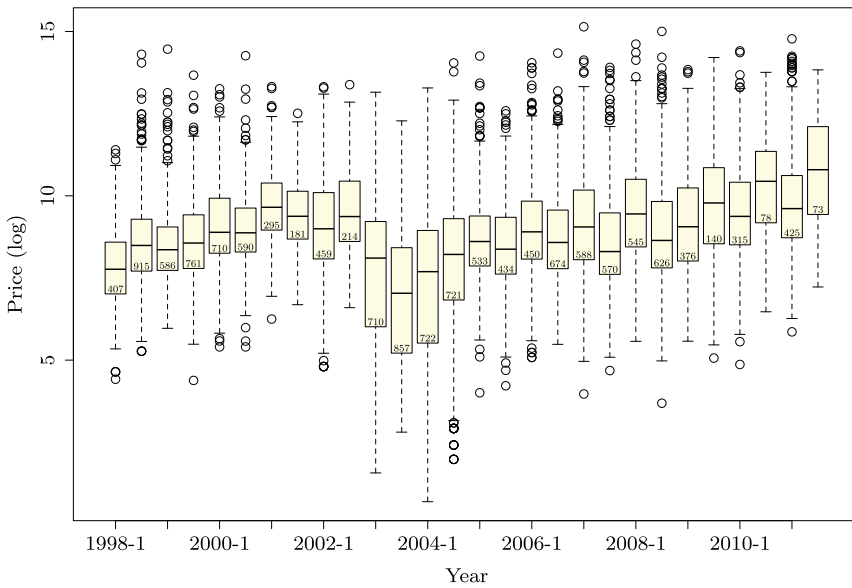
FIG. 1. *Boxplots of prices in (natural) logarithmic scale of the tribal art market by semester. The amount of items sold in a given semester is reported inside the boxes.*

the dataset emerges clearly from the graph: in every semester a different group of artworks is sold, for example, 407 items were auctioned in 1998-1, 915 objects different from the first set were sold in 1998-2, and so on. Hence, tribal art data has a structure like that of repeated cross-sectional surveys and the medians (black lines) give an idea of the trend of prices over time. In particular, note the consistent reduction in the number of auctioned items starting from 2009. Despite this, the overall turnover did not drop since the average price rose. This might indicate the adoption by market agents of hedging strategies against the economic crisis. Overall, we have $T = 28$ semesters, and n_t, the number of items sold in the semester t, varies between 73 (2011-2) and 915 (1998-2); the total sample size of sold items ($n = \sum_{t=1}^{T} n_t$) is 13,955. There are several reasons to aggregate the data in semesters rather than auction dates. First, auction dates are not equally spaced in time and, in our approach, this feature is essential to model time dependence. Treating auction dates as equally spaced would produce a severe bias on model identification as well as parameter estimation. Notice that unequally spaced observations could be modelled by adopting a completely different approach, for example, the continuous time framework described in Jones (1993), Chapter 3 for longitudinal data. Second, the art market of auction houses is naturally organized in semesters, and this is why the most important art indexes (e.g., artprice index, Mei and Moses fine art index) are semi-annual. The tribal art market makes no exception as the auction sessions are mostly concentrated in May/June and November/December and each session contains two to four auctions quite close in time.

The aggregation in semesters respects naturally this organization so that the results are meaningful from the economics point of view. It would be interesting to estimate the trend and other components at a monthly or quarterly frequency. Unfortunately, the tribal art database does not allow to consider a finer time scale. In fact, by organizing the data set in quarters we found out that, with few exceptions, there are no data both in Q1 and in Q3.

3. The model. In Section 3.1, we briefly review the model proposed in Modugno, Cagnone and Giannerini (2015) while in Section 3.2 we extend it by introducing the mixed effects model with stochastic volatility.

3.1. *A multilevel model with autoregressive random effects.* Let y_{it} be the natural logarithm of the observed price for item $i = 1, \ldots, n_t$ at time-point $t = 1, \ldots, T$ and let $\mathbf{x}_{it}$ be a corresponding column vector of k covariates. Since tribal art data can be thought to have a two-level structure where items represent level-1 units, and time points represent level-2 units, we consider the following random intercept model:

$$y_{it} = \beta_0 + u_t + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \qquad \varepsilon_{it}|\mathbf{x}_t \sim \mathrm{NID}(0, \sigma^2),$$

where u_t are time-specific random intercepts whose variance accounts for the unobserved heterogeneity between items within each time point; $\boldsymbol{\beta}$ is a vector of fixed slopes and β_0 is the overall mean. In repeated cross-sectional data, y_{it} and $y_{i(t+1)}$ are not the price of the same item i observed at successive time points since the two objects are physically different. Conditionally on the vector of covariates $\mathbf{x}_t$, level-1 errors ε_{it} (the error term for a given individual at a given time point) follows a normal distribution with constant variance. In other words, the art market is assumed to have a constant volatility over time. Note that ε_{it} is conditioned on the vector of the covariates $\mathbf{x}_t$ for all the individuals, that is, we assume strict exogeneity on the explanatory variables [Wooldridge (2010)]. Different from panel data, in repeated cross-sectional data the strict exogeneity assumption implies that, for each item, the covariates are uncorrelated with the error terms.

The dynamics of u_t can be modelled at the second level by extending the above multilevel models as follows:

$$u_t = \rho u_{t-1} + \eta_t, \qquad \eta_t|\mathbf{x}_t \sim \mathrm{NID}(0, \sigma^2_\eta),$$

where $\eta_t \perp u_s$ and $\eta_t \perp \varepsilon_{it}$ for all $s < t$ and for all i. In this specification, the random effects follow an autoregressive process of order 1. We denote this model as Autoregressive Random Effects (ARE). Modugno, Cagnone and Giannerini (2015) introduced a full maximum likelihood estimation method via the EM algorithm to fit the ARE model to the tribal art data. The ARE model improves considerably over classical models in terms of prediction and forecasting. However, as it will be shown in the Application section, the assumption of normality of level-1 errors

is violated, probably due to the presence of heteroscedasticity and kurtosis. As mentioned above, the assumption of a constant volatility of prices of art assets is not realistic and might cause severe inference problems. In the following, we extend the ARE model by including a stochastic volatility component that accounts properly for the heteroscedastic and autocorrelated volatility of the level-1 error process.

3.2. *A multilevel model with autoregressive random effects and stochastic volatility.* We include a stochastic volatility component at level-1 of the ARE model as follows:

(1) $$y_{it} = \beta_0 + u_t + \mathbf{x}'_{it}\boldsymbol{\beta} + \exp(h_t/2)\varepsilon_{it},$$

(2) $$u_t = \rho u_{t-1} + \eta_t,$$

(3) $$h_t = \alpha + \delta h_{t-1} + \sigma_\nu \nu_t,$$

for $i = 1, \ldots, n_t$ and $t = 1, \ldots, T$. As before, u_t is the time dependent random effect whereas h_t is the latent variable that represents the volatility component at time t. Both u_t and h_t follow a stationary autoregressive process, so that $|\rho| < 1$ and $|\delta| < 1$. Moreover, we assume that ε_{it}, η_t and ν_t are mutually independent, with $\varepsilon_{it} \sim \text{NID}(0, 1)$, $\eta_t \sim \text{NID}(0, \sigma_\eta^2)$, and $\nu_t \sim \text{NID}(0, 1)$, respectively. Note that we ruled out the random walk assumption for the random effects as this corresponds to the weak efficiency assumption for the art market, that is, it is not possible to predict future prices on the basis of the filtration that contains past information. In the literature, there is some evidence against this assumption [Ballesteros (2011), Goetzmann (1995)]. It seems that the art market tends towards greater efficiency over time and this might be ascribed to the increasing availability of auction information to prospective art buyers and sellers; still, it shows windows of (nontrivial) predictability.

The assumption of stationarity for u_t and h_t allows to avoid the problems of unknown initial values so that we do not have to use, for instance, a diffuse initialization. As also described in Durbin and Koopman (2012) (Chapters 5.6.2 and 9.5), the stationary unconditional distributions are taken as initial conditions for the two processes: $u_1 \sim N(0, \sigma_u^2)$ and $h_1 \sim N(\mu_h, \sigma_h^2)$ with $\sigma_u^2 = \sigma_\eta^2/(1 - \rho^2)$, $\mu_h = \alpha/(1 - \delta)$ and $\sigma_h^2 = \sigma_\nu^2/(1 - \delta^2)$. Under these assumptions, the conditional densities result:

$$y_{it}|\mathbf{x}_{it} \sim \text{NID}(\beta_0 + \mathbf{x}'_{it}\boldsymbol{\beta}, \sigma_u^2 + \sigma_{h*}^2),$$

(4) $$u_t|u_{t-1} \sim \text{NID}(\rho u_{t-1}, \sigma_\eta^2),$$

$$h_t|h_{t-1} \sim \text{NID}(\alpha + \delta h_{t-1}, \sigma_\nu^2),$$

where $\sigma_{h*}^2 = \text{Var}(\exp(h_t/2)\varepsilon_{it}) = \exp(\mu_h + 0.5\sigma_h^2)$ is the variance of the level-1 error term, assuming the stochastic volatility model for it. Thus, the sum of the

variance components $\sigma_u^2 + \sigma_{h^*}^2$ in (4) is the variability of the responses due to the random part of the model. The overall unconditional variability of Y is the sum of all the variance components of the model:

$$\text{Var}(Y) = \sigma_{\mathbf{x}^*}^2 + \sigma_u^2 + \sigma_{h^*}^2,$$

where $\sigma_{\mathbf{x}^*}^2 = T^{-1} \sum_t n_t^{-1} \boldsymbol{\beta}'(\tilde{\mathbf{x}}_t'\tilde{\mathbf{x}}_t)\boldsymbol{\beta}$, being $\tilde{\mathbf{x}}_t$ the centered variables. The specification also implies that the associations among different items are explained through the autoregressive process u_t. Indeed, the covariances of the response variable within and between time result:

$$\text{Cov}(y_{it}, y_{jt}|\mathbf{x}_t) = \sigma_u^2, \qquad i \neq j,$$

$$\text{Cov}(y_{it}, y_{js}|\mathbf{x}_t, \mathbf{x}_s) = \rho^{t-s}\sigma_u^2, \qquad i \neq j, s < t.$$

We call the new specification Stochastic Volatility and Autoregressive Random Effects (SVARE) model. The system (1)–(3) is a nonlinear state-space representation. Hence, model estimation can be performed by using maximum likelihood via a non-Gaussian filtering process and poses several non-trivial challenges that we describe and address in the following section.

4. Model estimation.

4.1. *The likelihood function.* We perform maximum likelihood estimation based on the following likelihood function:

$$L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}) = \int_{\mathbf{h}} \int_{\mathbf{u}} f(\mathbf{y}|\mathbf{u}, \mathbf{h}, \mathbf{x}) f(\mathbf{u}, \mathbf{h}) \, d\mathbf{u} \, d\mathbf{h}$$

(5)
$$= \int \cdots \int \left[\prod_{t=1}^{T} f(\mathbf{y}_t|u_t, h_t, \mathbf{x}_t) f(u_t|u_{t-1}) \right.$$

$$\left. \times f(h_t|h_{t-1}) \right] du_T \cdots du_1 \, dh_T \cdots dh_1,$$

where $\boldsymbol{\theta} = \{\beta_0, \boldsymbol{\beta}', \rho, \sigma_\eta, \alpha, \delta, \sigma_v\}$ is the vector of parameters, $f(\mathbf{y}_t|u_t, h_t, \mathbf{x}_t) = \prod_{i=1}^{n_t} f(y_{it}|u_t, h_t, \mathbf{x}_{it})$ for $t = 1, \ldots, T$, $f(u_1|u_0) = f(u_1)$ and $f(h_1|h_0) = f(h_1)$.

The computation of $L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x})$ requires solving a $2T$-dimensional integral which is computationally unfeasible. We address the issue by applying an iterated numerical integration procedure introduced by Kitagawa (1987) for non-Gaussian filtering problems. The procedure is based upon rephrasing the likelihood (5) as

$$L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x})$$

(6)
$$= \int\int f(\mathbf{y}_1|u_1, h_1, \mathbf{x}_1) f(u_1) f(h_1) \int\int f(\mathbf{y}_2|u_2, h_2, \mathbf{x}_2) f(u_2) f(h_2)$$

$$\cdots \int\int f(\mathbf{y}_T|u_T, h_T, \mathbf{x}_T) f(u_T) f(h_T) \, du_T \, dh_T \cdots du_2 \, dh_2 \, du_1 \, dh_1.$$

The resulting bivariate integrals can be approximated by using numerical quadrature techniques. The most used techniques for stochastic volatility models are the rectangular quadrature (RQ) [Bartolucci and De Luca (2001)] and the Gauss–Legendre quadrature rule (GL) [Fridman and Harris (1998)]. The choice between the two methods should be closely related to the data analyzed. In fact, the results of existing literature [see, e.g., Cagnone and Bartolucci (2017)] indicate that the two methods perform similarly when the squared coefficient of variation of the volatility is equal or greater than 1, whereas the rectangular quadrature outperforms the Gauss–Legendre for squared coefficient of variation equal to 0.1. The application of the two quadrature based methods to equation (6) produces the following approximated likelihood function:

$$
\tilde{L}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}) = (r_u r_h)^T \sum_{i_1}^{n_u} w_{ui_1} \sum_{j_1}^{n_h} w_{hj_1} f(\mathbf{y}_1|u_{i_1}^*, h_{j_1}^*, \mathbf{x}_1) f(u_{i_1}^*) f(h_{j_1}^*)
$$

$$
(7) \qquad \times \sum_{i_2}^{n_u} w_{ui_2} \sum_{j_2}^{n_h} w_{hi_2} f(\mathbf{y}_2|u_{i_2}^*, h_{j_2}^*, \mathbf{x}_2) f(u_{i_2}^*|u_{i_1}^*) f(h_{j_2}^*|h_{j_1}^*)
$$

$$
\cdots \sum_{i_T}^{n_u} w_{ui_T} \sum_{j_T}^{n_h} w_{hi_T} f(\mathbf{y}_T|u_{i_T}^*, h_{j_T}^*, \mathbf{x}_T) f(u_{i_T}^*|u_{i_{T-1}}^*) f(h_{j_T}^*|h_{j_{T-1}}^*),
$$

where, using the Gauss–Legendre quadrature method, $\{u_i^*\}$, with $i = 1, \ldots, n_u$, and $\{h_j^*\}$, with $j = 1, \ldots, n_h$, are sets of Gauss–Legendre quadrature points, w_{ui} and w_{hj} are the corresponding weights. The constants r_u and r_h are defined as

$$
r_u = \left(\frac{b - a}{2}\right); \qquad r_h = \left(\frac{e - d}{2}\right),
$$

where $[a, b]$ and $[d, e]$ are finite integration limits which replace the infinite ones for the random effects and the volatility process, respectively. With the rectangular quadrature method, the quadrature points are chosen as equidistant in the ranges $[a, b]$ and $[d, e]$ and the weights w_{ui} and w_{hj} are set equal to 1 [Bartolucci and De Luca (2001)].

Under both methods, the choice of the grids and the number of evaluation points is crucial for numerical precision. First, as proposed in Fridman and Harris (1998), the grids for the two latent processes are centered 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 nonnegligible mass. Second, the number of quadrature points, n_u and n_h, are chosen according to the degree of smoothness of the integrands, that is, 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. As we will discuss in Section 5, in our case we chose the Gauss–Legendre quadrature method.

Estimation via the non-Gaussian filtering process consists in maximizing the approximated likelihood of equation (7) and it is based on a recursive algorithm described in the supplementary material [Cagnone, Giannerini and Modugno (2017)]. We also derive optimal estimators for the unobserved-state vectors **u** and **h** by performing filtering and smoothing, which differ in the conditioning information set. These procedures, together with one-step-ahead prediction, are described in detail in the supplementary material.

5. Application to tribal art prices.　In this section, we illustrate the application of our model to the first database of ethnic artworks. The responses are the logged prices for 28 semesters for the overall sample size of 13,955 items. As is customary with prices, we take the natural logarithm as to linearize the effects and obtain distributions closer to the Gaussian. We take the fixed effects hedonic specification (FE) as the benchmark model. For the selection of covariates, we applied stepwise (forward and backward) techniques combined with parsimony and art economics arguments. First of all, note that tribal art is considered an *anonymous* art in that the geographic/ethnic provenance plays the role that the artist's name has in Western art and strongly characterizes the object. However, the variables Continent, Region and Ethnic group are nested so that they are collinear. Given that there are 17 regions and 361 ethnic groups, we decided to include the Region as a reasonable compromise between fitting capability and parsimony. Then we proceeded by applying stepwise forward and backward methods with both the AIC and BIC criteria. Whilst the AIC criterion did not rule out any variable, the BIC criterion kept 10 variables out of 14 both in forward and backward procedures. Among the excluded variables there was *Historicization*, which art economics experts consider an important feature. For this reason, we added it to the model.

Table 2 reports the parameter estimates for the three models: fixed effects (FE), autoregressive random effects (ARE) and stochastic volatility with autoregressive random effects (SVARE), fitted on the same data set with the same set of covariates. The asymptotic standard errors for the FE and SVARE models are derived from the Hessian matrix of the likelihood functions. The robust standard error for the ARE model are derived by means of the wild bootstrap for multilevel models introduced in Modugno and Giannerini (2015).

As mentioned above, we chose the Gauss–Legendre quadrature method to approximate the likelihood of the SVARE model. In our application, the squared coefficient of variation of the volatility is greater than 1 and, in agreement with the results on the classical stochastic volatility models, we found that GL and RQ perform similarly, also in terms of both computational time and number of iterations to convergence. This is shown in Table 2 of the supplementary material [Cagnone, Giannerini and Modugno (2017)] where we have estimated the SVARE model by varying the number of quadrature points for the two methods. According to the rule given in Section 4.1, we first set $n_u = n_h = 21$ and then increased them up to 61. The estimates indicate a robust fit for all the parameters, with the exception of α.

TABLE 2

Significant parameter estimates for models FE, ARE, and SVARE with standard errors in parentheses. The complete set of estimates is available in the supplementary material [Cagnone, Giannerini and Modugno (2017)]. For each categorical variable, the baseline category is indicated

	FE	ARE	SVARE
σ	1.073 (0.006)	1.074 (0.008)	–
σ_η	–	0.326 (0.057)	0.324 (0.044)
ρ	–	0.833 (0.148)	0.862 (0.084)
α	–	–	−0.018 (0.006)
δ	–	–	0.910 (0.055)
σ_ν	–	–	0.414 (0.059)
β_0	–	6.793 (0.652)	7.018 (0.112)
Type of object: baseline Furniture			
– Sticks	−0.143 (0.069)	−0.144 (0.075)	−0.238 (0.064)
– Masks	0.270 (0.057)	0.269 (0.063)	0.253 (0.052)
– Ornaments	−0.306 (0.063)	−0.306 (0.074)	−0.232 (0.058)
– Sculptures	0.149 (0.054)	0.147 (0.058)	0.142 (0.049)
– Tools	−0.173 (0.055)	−0.175 (0.062)	−0.167 (0.050)
– Weapons	−0.240 (0.069)	−0.240 (0.074)	−0.212 (0.065)
Material: baseline Ivory			
– Wood	0.279 (0.054)	0.280 (0.065)	0.194 (0.051)
– Gold	0.231 (0.082)	0.231 (0.101)	0.332 (0.077)
– Bone, horn	−0.413 (0.091)	−0.412 (0.121)	−0.344 (0.089)
Patina: baseline Not indicated			
– Pejorative	0.446 (0.104)	0.445 (0.106)	0.365 (0.104)
– Appreciative	0.183 (0.029)	0.182 (0.035)	0.149 (0.028)
Region: baseline Central Africa			
– Southern Africa	−0.473 (0.075)	−0.474 (0.105)	−0.462 (0.074)
– Western Africa	−0.273 (0.029)	−0.274 (0.032)	−0.278 (0.026)
– Eastern Africa	−0.355 (0.071)	−0.355 (0.083)	−0.382 (0.065)
– Indonesia	−0.359 (0.071)	−0.361 (0.073)	−0.323 (0.069)
– Polynesia	0.477 (0.045)	0.477 (0.053)	0.441 (0.041)
– Northern America	0.525 (0.047)	0.525 (0.055)	0.418 (0.042)
– Mesoamerica	0.245 (0.053)	0.243 (0.061)	0.155 (0.047)
– Indian Subcontinent	0.824 (0.285)	0.815 (0.283)	0.760 (0.242)
Illustration width: baseline Absent			
– Col. miscellaneous	0.979 (0.048)	0.980 (0.066)	0.964 (0.046)
– Col. quarter page	1.825 (0.048)	1.825 (0.063)	1.567 (0.046)
– Col. half page	2.224 (0.056)	2.223 (0.071)	1.943 (0.053)
– Col. full page	2.564 (0.060)	2.565 (0.077)	2.319 (0.056)
– Col. more than one	3.060 (0.065)	3.063 (0.086)	2.857 (0.061)
– Col. cover	3.375 (0.171)	3.376 (0.242)	3.122 (0.168)
– b/w miscellaneous	1.116 (0.107)	1.114 (0.092)	0.920 (0.085)
– b/w quarter page	0.834 (0.067)	0.835 (0.070)	0.673 (0.057)
– b/w half page	1.403 (0.138)	1.403 (0.129)	1.223 (0.106)
– b/w full page	2.204 (0.542)	2.204 (0.642)	1.876 (0.431)

TABLE 2
(*Continued.*)

	FE	ARE	SVARE
Description on the catalogue: baseline Absent			
– Short visual descr.	−0.182 (0.085)	−0.184 (0.102)	−0.257 (0.076)
– Broad visual descr.	0.784 (0.094)	0.784 (0.110)	0.675 (0.084)
– Critical descr.	0.774 (0.094)	0.772 (0.113)	0.652 (0.084)
– Broad critical descr.	1.668 (0.103)	1.666 (0.120)	1.487 (0.092)
CABS (Yes vs. No)	0.249 (0.031)	0.248 (0.033)	0.227 (0.028)
CABC (Yes vs. No)	0.332 (0.024)	0.332 (0.027)	0.306 (0.022)
CAES (Yes vs. No)	0.204 (0.034)	0.205 (0.038)	0.205 (0.030)
Historicization: baseline Absent			
– Simple certification	0.083 (0.025)	0.083 (0.029)	0.102 (0.023)
Paris (vs. New York)	−0.307 (0.030)	−0.302 (0.035)	−0.197 (0.026)
Christie's (vs. Sotheby's)	−0.253 (0.026)	−0.255 (0.030)	−0.335 (0.025)

This is in agreement with the results in literature that show a high mean square error for the estimator for α [see, e.g., Bartolucci and De Luca (2001)]. Also, the RQ with 61 quadrature points did not converge so that we chose the GL method with 51 points. This is the solution with the highest log-likelihood and the best forecasting performance among the GL results (see Table 3 of the supplementary material). Moreover, 51 points should be enough to ensure an accurate approximation of the standard errors. Alternative estimation methods, like the EM algorithm or Markov chain Monte Carlo techniques in the Bayesian context, could also be used and will be subject of further investigations.

In order to assess whether the specifications proposed manage to model satisfactorily the time dynamics and the heterogeneity observed, we have implemented a series of diagnostic tests. Table 3 reports information on the goodness of fit of the models. In particular, we have computed a version of the R^2 coefficient for mixed effects models, with and without stochastic volatility. The R^2 for the ARE

TABLE 3
Log-likelihood, number of parameters, information criteria, and coefficient of determination R^2 for the hedonic regression (FE), ARE, and SVARE models

	FE	ARE	SVARE
loglik	−20,788.86	−20,855.52	−20,167.98
n. par	93	68	70
AIC	41,763.73	41,847.05	40,475.96
BIC	42,465.28	42,360.01	41,004.01
R^2	0.65	0.61	0.70

TABLE 4
*p-values of the Shapiro–Wilk test and indexes of skewness and kurtosis
for the FE, ARE, and SVARE (level-1) residuals*

	FE	ARE	SVARE
Shapiro–Wilk p-value	<2.2e−16	<2.2e−16	8.9e−08
Skewness	−0.157	−0.161	0.021
Kurtosis	1.102	1.111	0.497

and SVARE models has been generalized with respect to the classical R^2 (for fixed effects regression) to account properly for the proportion of variability explained by the different components of the models. Following the lines of Xu (2003), the R^2 for the ARE model can be obtained as follows:

$$(8) \qquad R^2 = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_u^2 + \hat{\sigma}_{\mathbf{x}^*}^2 + \hat{\sigma}^2}.$$

It expresses the proportion of the variability of Y explained by both the random intercept and the covariates. In the same way, the coefficient of determination for the SVARE model can be defined as

$$(9) \qquad R^2 = 1 - \frac{\hat{\sigma}_{h*}^2}{\hat{\sigma}_u^2 + \hat{\sigma}_{\mathbf{x}^*}^2 + \hat{\sigma}_{h*}^2}.$$

In this case, the R^2 can be interpreted as the proportion of variability of Y explained by the random intercept and the covariates, assuming the stochastic volatility model for the level-1 error term. As for the FE model, we consider the classical adjusted R^2.

In Table 4, we present the results of some diagnostic tests and indicators on the residuals. In particular, the first row of Table 4 shows the p-values for the Shapiro–Wilk test for normality of the residuals of the three models; the `shapiro.test` function in R limits the sample size to 5,000. The results presented are the median p-values over 20,000 random subsamples of size 5,000 drawn from the original sample. The last two rows show the indexes of skewness and kurtosis b_1 and b_2 as in Joanes and Gill (1998) computed on level-1 residuals. We derive the standardized residuals for the ARE model from the best linear unbiased predictors (BLUP) of the random effects whereas for the SVARE model we use the smoothed values for both the random effects and the volatility. Their expressions are reported in the supplementary material [Cagnone, Giannerini and Modugno (2017)].

As concerns the time dynamics we assess the adequateness of the models by computing the sample global and partial autocorrelation functions over time-varying quantities such as level-2 residuals. Moreover, we use the metric entropy measure S_k defined as

$$(10) \qquad S_k = \frac{1}{2} \int \int \left[\{ f_{(X_t, X_{t+k})}(x_1, x_2) \}^{1/2} - \{ f_{X_t}(x_1) f_{X_{t+k}}(x_2) \}^{1/2} \right]^2 dx_1 \, dx_2,$$

TABLE 5
Prediction/forecasting performance of the three models over 100
out-of-sample units within the time span 1998–2011 (*rows* 2–4)
and over 73 *units of the out-of-sample semester*, 2011-2 (*rows* 5–7)

		FE	**ARE**	**SVARE**
Prediction	MAE	0.66	0.66	0.63
	RMSE	0.82	0.82	0.79
	MAPE	7.4	7.4	7.1
Forecast	MAE	1.20	0.87	0.83
	RMSE	1.55	1.21	1.15
	MAPE	10.7	7.9	7.6

where f_{X_t} and $f_{(X_t, X_{t+k})}$ denote the probability density function of X_t and of the vector (X_t, X_{t+k}), respectively. The measure is a particular member of the family of relative entropies, which includes as a special case nonmetric entropies often referred to as Shannon or Kullback–Leibler divergence. It can be interpreted as a nonlinear autocorrelation function and possesses many desirable properties. We use S_k as in Giannerini, Maasoumi and Bee Dagum (2015) to test for nonlinear serial dependence and as in Granger, Maasoumi and Racine (2004) to test for serial independence (see the supplementary material for more details). The tests are implemented in the R package tseriesEntropy [Giannerini (2015)]. In the spirit of time-series analysis, if the specification is appropriate then the residuals behave as a white noise process and diagnostic tests can suggest directions to improve the existing model.

Finally, Table 5 compares the prediction/forecasting capability of the three models under scrutiny. It reports the prediction error over 100 (out of sample) items within the time span 1998–2011, and the forecasting performance over all the 73 observations of semester 2011-2. Such observations have not been included in the model fit so that the measures reflect a genuine forecasting performance The aggregate measures of prediction error are the Mean Absolute (Prediction) Error (MAE), the Root Mean Square (Prediction) Error (RMSE), and the Mean Absolute Percentage Error (MAPE), given by

$$ \text{MAE} = \frac{1}{n^*} \sum_{i=1}^{n^*} |e_i|; \qquad \text{RMSE} = \sqrt{\frac{1}{n^*} \sum_{i=1}^{n^*} e_i^2}; \qquad \text{MAPE} = \frac{1}{n^*} \sum_{i=1}^{n^*} \left| \frac{100 e_i}{y_i} \right|; $$

where $e_i = y_i - \hat{y}_i$ is the forecasting/prediction error for item i, and n^* is the number of predicted responses. The MAPE is scale independent and allows to compare the performance of different models and also different data sets. It is meaningful if the scale has a meaningful origin and it is best suited to data sets without zeroes and without values close to zero. In our case, these conditions are fulfilled. Moreover, the MAPE takes values in the interval $[0, \infty]$, where the minimum value zero

indicates perfect fit/forecast. Hence, it provides also a measure of absolute performance and a value below 10% is usually taken as an indication of a very good fit.

5.1. *FE and ARE models.*   The parameter estimates of the FE and ARE models are very similar (first two columns of Table 2), still there are some differences: first, the ARE fit is more parsimonious and results in a smaller BIC (see Table 3); also, it provides a decomposition of the total variability of the response in between-time and within-time variability. Furthermore, in the FE model the time dynamics is modelled through 28 dummy variables while in the ARE model the time dynamics is fully captured through the AR(1) specification with the two parameters, ρ and σ_η; see Figures 1 and 2 of the supplementary material. The same tests performed on level-2 residuals of the ARE model show no structure (see Figures 3 and 4 of the supplementary material). Finally, the ARE model provides a superior one-step-ahead forecasting of the price whereas the prediction performance is the same as that of the FE model (see Table 5). Note that the R^2 for the ARE model (0.61) is lower than that of the FE model (0.65) and this could be due to the impact of the 28 time dummies on the coefficient which is notoriously biased towards overfitting. Interestingly though, the better performance of the FE model in terms of explained variability does not imply a better forecast as the ARE specification manages to model the time dynamics with the autoregressive component.

The Shapiro–Wilk test (see Table 4) points to a deviation from normality in level-1 residuals of the ARE model (whereas it does not reject the assumption of normality for level-2 residuals). As discussed above, this is consistent with the findings in literature and might be due to heteroscedasticity. In fact, similarly to other assets, level-1 residuals show a leptokurtic behaviour as shown in Figure 4 and by looking at the kurtosis index in Table 4. Furthermore, we reject the assumption of homogeneity of the variance across time points, tested through a nonparametric version of the Levene (1960) rank-based test.

The plot of s_t^{ARE}, the standard deviations of level-1 residuals $\hat{\varepsilon}_{it}$ in Figure 2(left) provides a visual evidence of volatility patterns. The entropy measure S_k shown in Figure 2(right) confirms the presence of a linear serial dependence (the test for nonlinearity does not reject, see Figure 6 of the supplementary material) and the correlograms of Figure 3 indicate a AR(1)-type dependence structure for the volatility. In the following subsection, we account for the observed heterogeneity by fitting the multilevel model with autoregressive random effects and stochastic volatility (SVARE).

5.2. *SVARE model.*   The point estimates of the SVARE model are in agreement with those of the FE/ARE models. In most cases, the significance of the parameters does not change and this indicates the overall consistency of the multilevel approach. Nevertheless, there are important differences. In fact, SVARE estimates account properly for the volatility and reflect more closely the impact
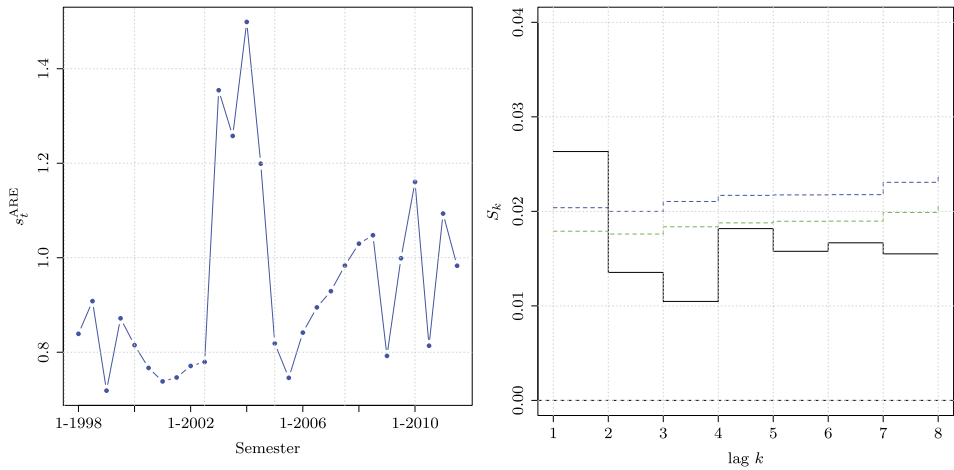
FIG. 2. *Standard deviations of ARE level-1 residuals s_t^{ARE}: time plot (left) and entropy measure of dependence (right). The confidence bands correspond to the null hypothesis of serial independence at levels 90% and 95% up to 10 lags/semesters.*

of the covariates on artwork prices. This is reinforced by the standard error of the estimates which are invariably the lowest among the three models. The most noticeable differences can be found in the coefficient of the category Material-Gold: from 0.231 for the FE/ARE model to 0.332 for the SVARE model. Gold and wood are still the materials with the highest estimated coefficients, which in the hedonic regression framework, are interpreted as the prices of each feature, the so-called *shadow prices*. Also, the absolute impact of the venue reduces from 0.302 (ARE)
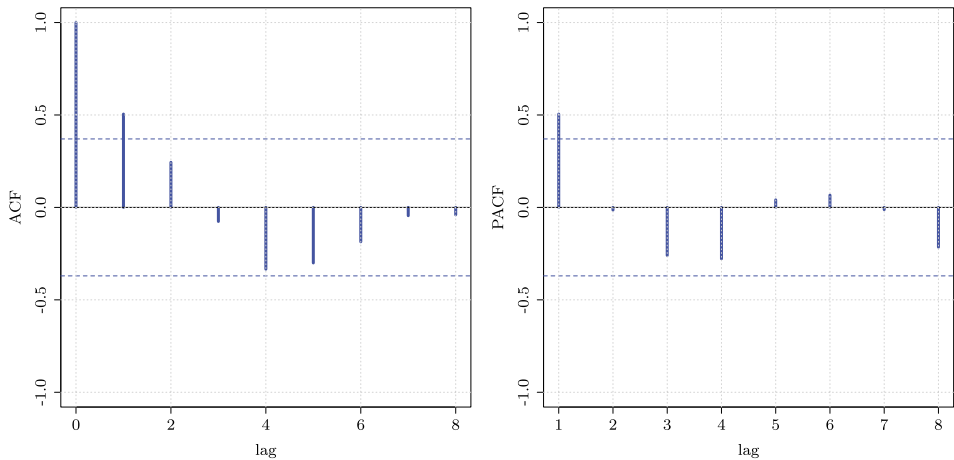

FIG. 3. *Global (left) and partial (right) empirical autocorrelation functions of standard deviations of ARE level-1 residuals.*

to 0.197 (SVARE) while the auction house effect increases from 0.255 (ARE) to 0.335 (SVARE). This might indicate that the auction house is more likely to have a significant effect on the selling price rather than the venue. In any case, the shadow price of the auction house confirms the international leadership of Sotheby's in the tribal art market as in other art markets; this is known among operators as *Sotheby's effects*. Analogously, the reduced shadow price of Paris with respect to New York is not unexpected and reflects its more recent entrance in this market. Moreover, the most appraised objects result masks and sculptures and these are also the most traditional. The patina, especially when it worsens the object appearance, is a valued feature of some tribal objects, since in many cases, it would derive from the settling of organic liquids during sacrificial rites, and thus its presence would witness its real usage [Biordi and Candela (2007)]. Also, the use of catalogues as a marketing tool by auction houses appears important in fetching good prices. In fact, the shadow prices tend to increase as the importance given to the object on the catalogue through illustrations and descriptions increases. Finally, investors tend to pay more for objects with a relevant *pedigree*, for example, for those boasting either object-specific (CABS) or just comparative (CABC) citations and for those that have been previously exhibited (CAES).

The SVARE model provides also the key information deriving from the volatility parameter $\hat{\delta} = 0.910$, which agrees with those of models for financial time series reported in literature and indicates a nonnegligible volatility persistence. Indeed, the goodness of fit of the SVARE model increases noticeably as witnessed by both the information criteria and from the R^2 in Table 3.

Also for the SVARE model, the Shapiro–Wilk test of normality of level-1 residuals rejects the null hypothesis (Table 4). Nevertheless, the leptokurtic behaviour of residuals is considerably reduced with respect to both the ARE and FE models. This is shown in Table 4 (the skewness disappears and the kurtosis is more than halved) and in Figure 4 where we show the densities of level-1 residuals for the ARE and SVARE models. Note the agreement of SVARE residuals with the standard Normal density (dotted in the figure). See also Figures 11 and 12 of the supplementary material for a normal qq-plot of level-1 residuals and a log-density plot for the two models. As in the FE/ARE case, we compute the diagnostic tests of dependence on level-2 residuals $\hat{\eta}_t$ and $\hat{v}_t$. Both the correlograms and the entropy measure S_k indicate the absence of any dependence structure (see Figures 7–10 of the supplementary material) so that we may argue that the SVARE specification manages to capture the volatility dynamics. Finally, from Table 5 it emerges clearly that the SVARE specification performs best among competitors in terms of both prediction and forecasting. In particular, besides allowing comparisons, the MAPE for the SVARE model (7.1%) is also an indication of a good fit in absolute terms.

6. Conclusions. The SVARE specification provides a natural and convenient framework for modelling the trends in the mean and in the volatility of artwork
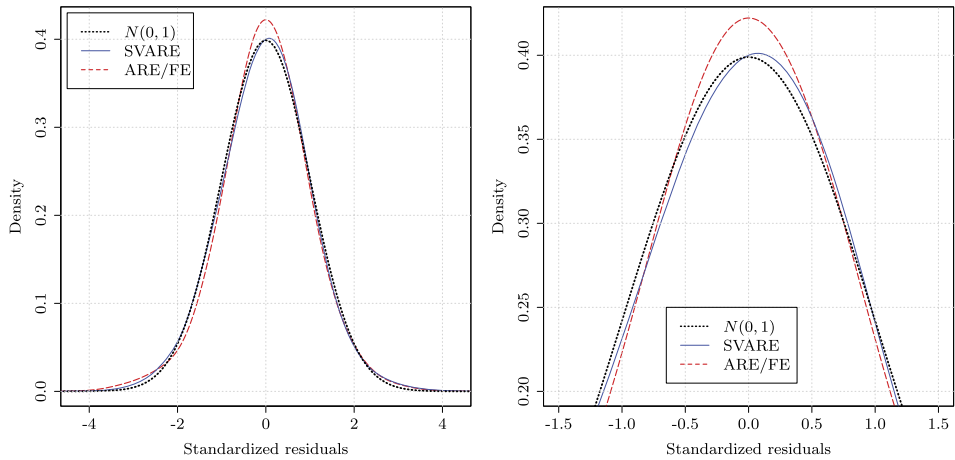
FIG. 4. *Density of level-*1 *residuals. The right panel is a zoomed detail of the peak that highlights the kurtosis.*

prices. The model does not assume that the observations form a panel, which is clearly not the case for auction data, nor it needs repeated sales data. The stochastic volatility component accounts (to some extent) for the deviation from normality observed in the residuals of models without it, especially regarding the skewness and the kurtosis. As witnessed by the modified R^2 measure, there is a gain in the explained heterogeneity with respect to the ARE model. Also, we observe a superior forecasting ability. Of course, the dynamic of art prices still retains a proportion of unexplained, maybe unexplainable, variability. For instance, there might be complex interactions with the buy-in phenomenon [Collins, Scorcu and Zanola (2009)] or with the selling probability [Candela, Castellani and Pattitoni (2012)]. Moreover, modelling the so-called *superstars* would require a different class of models, possibly rooted in extreme value theory. Last but not least, the values of tribal artworks are deemed to be *private* rather than *public* since are more dictated by the personal judgements of passionate collectors rather than the common consensus of some community and this complicates the modelling task. Still, the multilevel model with stochastic volatility provides important additional information on the predictability of the prices, and hence, on investment risks, that can be exploited by art market stakeholders for informed decision making. This can be best appreciated by looking at Figure 5 where we show the biannual price indexes obtained through the FE, ARE, and SVARE fits (left panel), together with the predicted volatility values of the SVARE model, $\exp(\hat{h}_t/2)$ (right panel). The indexes are computed with fixed base in semester $b =$ "1-1998" as

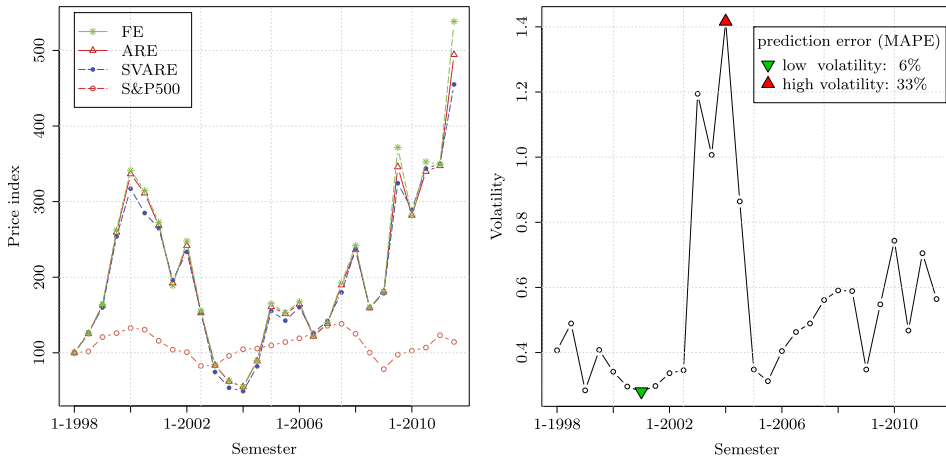$$I_t = \frac{e^{\hat{\beta}_{0t}}}{e^{\hat{\beta}_{0b}}} \times 100,$$

FIG. 5. (*left panel*) *Price index of the tribal art market for the FE, ARE, and SVARE models, together with the S&P*500 *index* (*base*: *"semester* 1-1998*"*). (*right panel*) *Plot of the estimated volatility of the SVARE model. The semesters with lowest and highest volatility are indicated with coloured triangles and the associated prediction errors are reported in the box.*

where, for multilevel models, the $\hat{\beta}_{0t}$s are the best linear unbiased predictors (BLUP values) $\hat{\beta}_0 + \hat{u}_t$. With the exceptions of few time points, the indexes for the three models are similar. However, by performing a prediction exercise with the SVARE fit on the prices of items sold in semesters 1-2001 (lowest volatility) and 1-2004 (highest volatility), we obtain a MAPE of 6% and 33%, respectively (see the right panel of Figure 5). This shows that the information deriving from the volatility is essential to the prediction of price dynamics and to the characterization of its complexity. The price indexes and the volatility shown in Figure 5 also highlight the peculiarity of the tribal art market. Indeed, both the traditional art market and financial markets showed a level drop and clusters of high volatility after the financial crisis of 2008; see the behaviour of the S&P500 index reported in Figure 5(left). This is not the case with the tribal art market. In fact, the tribal art index reached both its minimum level and highest volatility in 2004. Moreover, after 2008, it showed a sharp increase together with mild volatility. Note also the drop in volatility in semester 1-2009. There is no apparent dependence/correlation between the tribal art market and other market segments (e.g., traditional art, financial, gold). Hence, tribal artworks can be counted among the *alternative* investments that contribute to the diversification of a portfolio and enhance its performance.

SUPPLEMENTARY MATERIAL

Supplement to "Multilevel models with stochastic volatility for repeated cross-sections: An application to tribal art prices" (DOI: 10.1214/17-AOAS1035SUPP; .pdf). The online supplement contains six technical Appendices with detailed material on the following topics:

1. Recursive algorithm for computing the likelihood;
2. Filtering, Smoothing, and Prediction;
3. Application to Tribal Art prices: full table of the estimates;
4. Application to Tribal Art prices: choice of the quadrature based method;
5. Application to Tribal Art prices: entropy based diagnostic tests for serial independence and nonlinearity;
6. Software implementation.

REFERENCES

AGNELLO, R. and PIERCE, R. (1996). Financial returns, price determinants, and genre effects in American art investment. *J. Cult. Econ.* **20** 359–383.

BALLESTEROS, T. (2011). Efficiency tests in the art market using cointegration and the error correction model. *Social Science Research Network*. DOI:http://dx.doi.org/10.2139/ssrn.1696785.

BARTOLUCCI, F. and DE LUCA, G. (2001). Maximum likelihood estimation of a latent variable time-series model. *Appl. Stoch. Models Bus. Ind.* **17** 5–17. MR1819006

BAUMOL, W. (1986). Unnatural value: Or art investment as floating crap game. *Am. Econ. Rev.* **76** 10–14.

BIORDI, M. and CANDELA, G. (2007). L'arte etnica: Tra cultura e mercato. *Skira*.

BOCART, F. Y. R. P. and HAFNER, C. M. (2012). Econometric analysis of volatile art markets. *Comput. Statist. Data Anal.* **56** 3091–3104. MR2943883

BOCART, F. Y. R. P. and HAFNER, C. M. (2015). Volatility of price indices for heterogeneous goods with applications to the fine art market. *J. Appl. Econometrics* **30** 291–312. MR3322720

BOUND, J., JAEGER, D. and BAKER, R. (1995). Problems with instrumental variables estimation when the correlation between instruments and the endogenous explanatory variable is weak. *J. Amer. Statist. Assoc.* **90** 443–450.

BOX-STEFFENSMEIER, J. M., DE BOEF, S. and LIN, T.-M. (2004). The dynamics of the partisan gender gap. *Am. Polit. Sci. Rev.* **98** 515–528.

BROWNE, W. and GOLDSTEIN, H. (2010). MCMC sampling for a multilevel model with noninde-pendent residuals within and between cluster units. *J. Educ. Behav. Stat.* **35** 453–473.

CAGNONE, S. and BARTOLUCCI, F. (2017). Adaptive quadrature for maximum likelihood estimation of a class of dynamic latent variable models. *Comput. Econ.* **49** 599–622.

CAGNONE, S., GIANNERINI, S. and MODUGNO, L. (2017). Supplement to "Multilevel models with stochastic volatility for repeated cross-sections: An application to tribal art prices." DOI:10.1214/17-AOAS1035SUPP.

CANDELA, G., CASTELLANI, M. and PATTITONI, P. (2012). Tribal art market: Signs and signals. *J. Cult. Econ.* **36** 289–308.

CANDELA, G., CASTELLANI, M. and PATTITONI, P. (2013). Reconsidering psychic return in art investments. *Econom. Lett.* **118** 351–354.

CHANEL, O. (1995). Is art market behaviour predictable? *Eur. Econ. Rev.* **39** 519–527.

CHANEL, O., GÉRARD-VARET, L. and GINSBURGH, V. (1996). The relevance of hedonic price indices. The case of paintings. *J. Cult. Econ.* **20** 1–24.

COLLINS, A., SCORCU, A. and ZANOLA, R. (2009). Reconsidering hedonic art price indexes. *Econom. Lett.* **104** 57–60. MR2542857

DEATON, A. (1985). Panel data from time series of cross sections. *J. Econometrics* **30** 109–126.

DIPRETE, T. and GRUSKY, D. (1990). The multilevel analysis of trends with repeated cross-sectional data. *Sociol. Method.* **20** 337–368.

DURBIN, J. and KOOPMAN, S. J. (2012). *Time Series Analysis by State Space Methods*, 2nd ed. *Oxford Statistical Science Series* **38**. Oxford Univ. Press, Oxford. MR3014996

FRIDMAN, M. and HARRIS, L. (1998). A maximum likelihood approach for non-Gaussian stochastic volatility models. *J. Bus. Econom. Statist.* **16** 284–291.

GIANNERINI, S. (2015). tseriesEntropy: Entropy Based Analysis and Tests for Time Series. R package version 0.5-13.

GIANNERINI, S., MAASOUMI, E. and BEE DAGUM, E. (2015). Entropy testing for nonlinear serial dependence in time series. *Biometrika* **102** 661–675. MR3394282

GINSBURGH, V. and JEANFILS, P. (1995). Long-term comovements in international markets for paintings. *Eur. Econ. Rev.* **39** 538–548.

GOETZMANN, W. (1993). Accounting for taste: Art and financial markets over three centuries. *Am. Econ. Rev.* **83** 1370–1376.

GOETZMANN, W. (1995). The informational efficiency of the art market. *Manage. Finance* **21** 25–34.

GOETZMANN, W., MAMONOVA, E. and SPAENJERS, C. (2014). The economics of aesthetics and three centuries of art price records. Working Paper 20440, National Bureau of Economic Research.

GOLDSTEIN, H. (2010). *Multilevel Statistical Models*, 4th ed. Wiley, Chichester.

GRANGER, C. W., MAASOUMI, E. and RACINE, J. (2004). A dependence metric for possibly nonlinear processes. *J. Time Series Anal.* **25** 649–669. MR2086354

HODGSON, D. and VORKINK, K. (2004). Asset pricing theory and the valuation of Canadian paintings. *Canadian Journal of Economics/Revue canadienne déconomique* **37** 629–655.

JOANES, D. and GILL, C. (1998). Comparing measures of sample skewness and kurtosis. *J. R. Stat. Soc., Ser. D Stat.* **47** 183–189.

JONES, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach. Monographs on Statistics and Applied Probability* **47**. Chapman & Hall, London. MR1293123

KITAGAWA, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *J. Amer. Statist. Assoc.* **82** 1032–1063. MR0922169

LEBO, M. and WEBER, C. (2015). An effective approach to the repeated cross-sectional design. *Amer. J. Polit. Sci.* **59** 242–258.

LEVENE, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics* (I. Olkin, ed.) 278–292. Stanford Univ. Press, Stanford, CA. MR0120709

LOCATELLI BIEY, M. and ZANOLA, R. (2005). The market for picasso prints: A hybrid model approach. *J. Cult. Econ.* **29** 127–136.

MACKUEN, M., ERIKSON, R. and STIMSON, J. (1992). Peasants or bankers? The American electorate and the U.S. economy. *Am. Polit. Sci. Rev.* **86** 597–611.

MODUGNO, L., CAGNONE, S. and GIANNERINI, S. (2015). A multilevel model with autoregressive components for the analysis of tribal art prices. *J. Appl. Stat.* **42** 2141–2158. MR3373724

MODUGNO, L. and GIANNERINI, S. (2015). The wild bootstrap for multilevel models. *Comm. Statist. Theory Methods* **44** 4812–4825. MR3424810

MOFFITT, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *J. Econometrics* **59** 99–123.

ROSEN, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *J. Polit. Econ.* **82** 34–55.

SCOTT, A. J. and SMITH, T. M. F. (1974). Analysis of repeated surveys using time series methods. *J. Amer. Statist. Assoc.* **69** 674–678.

SKRONDAL, A. and RABE-HESKETH, S. (2004). *Generalized Latent Variable Modeling*: *Multi-level*, *Longitudinal*, *and Structural Equation Models*. Chapman & Hall/CRC, Boca Raton, FL. MR2059021

TANIZAKI, H. and MARIANO, R. (1998). Nonlinear and nonnormal state-space modeling with Monte-Carlo stochastic simulations. *J. Econometrics* **83** 263–290.

WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. MR2768559

XU, R. (2003). Measuring explained variation in linear mixed effects models. *Stat. Med.* **22** 3527–3541.

S. CAGNONE
S. GIANNERINI
L. MODUGNO
DIPARTIMENTO DI SCIENZE STATISTICHE
VIA DELLE BELLE ARTI 41
40126, BOLOGNA
ITALY
E-MAIL: silvia.cagnone@unibo.it
                simone.giannerini@unibo.it
                lucia.modugno@unibo.it