ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Bocart, Fabian Y. R. P.; Hafner, Christian M.

Working Paper

Volatility of price indices for heterogeneous goods

SFB 649 Discussion Paper, No. 2012-039

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Bocart, Fabian Y. R. P.; Hafner, Christian M. (2012) : Volatility of price indices for heterogeneous goods, SFB 649 Discussion Paper, No. 2012-039, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/79628

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

SFB 649 Discussion Paper 2012-039

Volatility of price indices for heterogeneous goods

Fabian Y.R.P. Bocart*
Christian M. Hafner*

* Institut de statistique, Université catholique de Louvain, Belgium

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Volatility of price indices for heterogeneous goods

Fabian Y.R.P. Bocart* Christian M. Hafner†

May 23, 2012

Abstract

Price indices for heterogenous goods such as real estate or fine art constitute crucial information for institutional or private investors considering alternative investments in times of financial markets turmoil. Classical mean-variance analysis of alternative investments has been hampered by the lack of a systematic treatment of volatility in these markets. This may seem surprising as derivatives on subsets of the traded goods require a precise modelling and estimation of the underlying volatility. For example, in art markets, auction houses often give price guarantees to the seller that resemble put options. In this paper we propose a hedonic regression framework which explicitly defines an underlying stochastic process for the price index, allowing to treat the volatility parameter as the object of interest. The model can be estimated using maximum likelihood in combination with the Kalman filter. We derive theoretical properties of the volatility estimator and show that it outperforms the standard estimator. We show that extensions to allow for time-varying volatility are straightforward using a local-likelihood approach. In an application to a large data set of international blue chip artists, we show that volatility of the art market, although generally lower than that of financial markets, has risen over the last years and, in particular, during the recent European debt crisis.

Keywords: Volatility, heterogenous goods, hedonic regression, random effects
JEL classification: **C14, C43, Z11**

*Institut de statistique, Université catholique de Louvain, Belgium.
†Corresponding author. Institut de statistique and CORE, Université catholique de Louvain, Voie du Roman Pays, 20, 1348 Louvain-la-Neuve, Belgium. This research was supported by the *Deutsche Akademische Auslandsdienst* and *Deutsche Forschungsgemeinschaft* through the SFB 649 *Economic Risk*. We would like to thank Christian Gouriéroux, Wolfgang Härdle, and seminar participants at the *Deutsche Statistische Woche* in Leipzig, Weierstrass Institute Berlin, Humboldt-University Berlin, University of Cologne and CREST/Paris for helpful discussions and comments. A large part of the paper was written while the second author was visiting the Institute of statistics and econometric at Humboldt-University, whose hospitality is gratefully acknowledged.

1 Introduction

Over the last two decades, there has been a growing interest among scholars, business practitioners, and policy makers in price indices tracking the financial performance of a basket of heterogeneous goods. These price indices have typically been developed for physical assets that can be considered as investments, such as housing, art, wine, as well as many other collectibles (musical instruments, watches, jewelry, etc.). In addition to managing all risks specific to physical assets (forgery, theft, destruction, etc.), investors in physical assets must deal with the risks common to all financial investments: market risks, liquidity risks and counterparty risks. Obviously, prior to modelling and managing these financial risks, a pre-requisite is to have an estimate of the underlying time series of prices and volatility of returns.

Returns of baskets of physical assets need to be indirectly estimated because of the presence of heterogeneity in the series. Generally, two methodologies are used to cope with this problem: the repeat sale methodology (RSM) and the hedonic regression. Some advantages and disadvantages of hedonic regression as compared to RSM for estimating returns in the art market are discussed in Ginsburgh et al. (2006). Dorsey et al. (2010) discuss hedonic versus repeat-sales indices in the real estate market of Los Angeles and San Diego metropolitan areas. RSM can be viewed as a nested case of hedonic regression and consists of computing average returns of identical goods sold through time. A major critique is that RSM focuses on a small, biased sample of goods (see Collins et al., 2009). RSM has been used to develop real-estate price indices by Case and Shiller (1987) and Goetzmann (1992). Pesando (1993), Goetzmann (1993) and Mei and Moses (2002) use RSM to estimate returns in the art market.

The hedonic approach is to regress the price of each good on its characteristics, in order to control for variations due to observable differences between heterogeneous goods. The classical approach is to include time dummy variables in the regression, whose coefficients constitute the basis for building an index. Hedonic regression has been extensively used to build price indices. A few examples are de la Barre et al. (1994), Collins et al. (2009), Hodgson and Vorkink (2004), Renneboog and Spaenjers (2010) and Bocart and Hafner (2011) for art markets, Schulz and Werwatz (2004) and Gouriéroux and Laferrere (2009) for real estate, Combris et al. (1997) and Fogarty (2006) for wine and Graddy and Margolis (2011) for violins.

The choice of an initial functional form to model returns is frequently debated in the literature. Empirically, Hansen (2009) finds that hedonic and repeat-sales methods provide similar estimates of price growth of Australian real estate when the sample is large. Dorsey et al. (2010) suggest that hedonic regression methods perform better at a local level to track prices of real estate in Los Angeles and San Diego. For the art market, Ginsburgh et al. (2006) show that

hedonic regression performs better than RSM when the sample size is small, while giving very similar results in large samples.

The goal of this article is to challenge the classical methodology of constructing the index using ordinary least squares (OLS), implicitly assuming deterministic prices, which is incoherent with a subsequent modelling of prices and returns as a stochastic process. Similar to the discussion of fixed versus random effects in the literature on panel data, we show that parameter estimation is more efficient exploiting the structure of a hypothesized random process. In particular, for an assumed random walk or stationary autoregressive process for the underlying market index, we derive explicitly the efficiency gains that can be achieved with maximum likelihood estimation compared to OLS. Since the parameters of interest are the variances of the two error components, these efficiency gains are equivalent to a more precise estimation of idiosyncratic and market volatility.

Interpreting the hedonic regression as an unbalanced panel model with time effects rather than individual effects, we further show that the fact of having an unbalanced panel deteriorates the properties of the estimators compared to the case of balanced panels with the same average number of observations, but that this negative effect disappears as the average number of observations per period increases. It should be understood that data of heterogeneous asset prices are typically highly unbalanced. In art markets, for example, sales are concentrated in spring and fall, with very few observations in summer.

Having in mind the large swings of volatility in financial markets, especially in crisis times, it is doubtful whether markets for heterogeneous goods have constant volatility. We therefore suggest a nonparametric extension of our model, allowing idiosyncratic and market volatility to be a smooth function of time that captures long-run trends in volatility. The functions can be conveniently estimated by local maximum likelihood.

We apply our methodology to the market of highly traded artworks in the period from 2000 to 2011. An ongoing debate about the diversification benefits of art in a portfolio has been taking place since Baumol (1986), and we contribute to this literature by explicitly delivering information about the associated risks of investing in this market. Our results suggest that long-run volatility of art followed a similar pattern as in financial markets, increasing during the financial crisis 2008/09 and the recent European debt crisis. On the other hand, the price trend seems to have been opposite during these two crises: While the trend was negative during the crisis following the Lehman Brothers bankruptcy, it was positive during the debt crisis. The latter observation might suggest that investors are increasingly considering art as an asset yielding diversification benefits in their portfolio during crisis times. Further studies using more data and explicitly modelling correlations between various markets for alternative investments

are needed to confirm this result.

The remainder of the paper is organised as follows. In Section 2, the basic model is presented. The third section introduces maximum likelihood estimation and compares efficiency of MLE with OLS. Section 4 discusses three extensions of the basic model, and Section 5 elaborates the results by applying the methodologies to empirical data on the art market. The last section closes this paper with final conclusions.

2 The model

As hedonic regression can be viewed as a generalization of the RSM approach, we consider an initial model that complies with the definition of a fully specified hedonic regression. However, the proposed estimation procedure can equally be applied to the RSM case.

Let there be N observed transactions and p_i denote the price of sale i. The logarithm of this price is usually modelled by the following hedonic regression model,

$$Y_i = \log p_i = \sum_{t=1}^{T} \beta_t d_{it} + \sum_{k=1}^{K} \alpha_k X_{ik} + u_i, \quad i = 1, ..., N. \tag{1}$$

The variable d_{it} is a dummy taking the value 1 if the object i was sold in period t, and 0 otherwise. The parameters β_t will be used to construct the pricing index. The parameters α_k are the coefficients of the explanatory variables, including a constant intercept term.

The time index $t = 1$ corresponds to the first period of the series and is used as benchmark. For identification, we set β_1 equal to zero. The K variables X_{ik} are all characteristics of the object i that have an impact on its price. For example, for a housing price index this would be variables such as the number of bathrooms and a dummy for a swimming-pool, for an art price index it would be the height, surface, and dummies for the artists, subject, etc. The price index, with base 100 in $t = 1$ is then defined as

$$\text{Index}_t = 100 \exp(\beta_t), \tag{2}$$

possibly corrected by a bias correction factor (see Jones and Zanola, 2010).

The regression (1) is generally estimated using Ordinary Least Squares (OLS). OLS estimators are efficient when errors u_i are normally distributed with constant variance, i.e., $u_i \sim N(0, \sigma_u^2)$. Empirical data, however, often violate this assumption. Hodgson and Vorkink (2004) and Seçkin and Atukeren (2006) focus on the normality part and propose a semiparametric estimator of the index based on a nonparametric error distribution, while maintaining

the assumption that u_i is i.i.d. and, hence, homoskedastic.

Furthermore, β_t is, by model assumption, a deterministic parameter rather than a stochastic process. To that extent, price indices built using OLS procedure cannot be interpreted as a random motion such as stock indices observed in financial markets. Nevertheless, it is standard practice to estimate β_t as if it was a deterministic parameter, and then continue working with the estimated β_t as if it was a realization of a stochastic process. As we will see, this methodological incoherence has important consequences for the properties of volatility estimators.

Note that model (1) can be written equivalently in the form

$$Y_{it} = \beta_t + X'_{it}\alpha + u_{it}, \quad t = 1, \ldots, T; \quad i = 1, \ldots n_t \tag{3}$$

where Y_{it} is the log price of the i-th sale at time t, and n_t is the number of sales at time t. The vector X_{it} contains the K explanatory variables of the i-th sale at time t, and α is a $(K \times 1)$ parameter vector. This model can be viewed as an unbalanced panel model with time effects. Individual effects are absent because the object of the i-th transaction at time t is not necessarily the same as the object of the i-th transaction at time t', $t' \neq t$. In fact, the ordering of the sales at a given time t is irrelevant as long as the error term u_{it} is i.i.d. across sales.

As is well known from the panel literature, the common OLS estimator of the hedonic regression (1) is equivalent to the fixed effects estimators $\hat{\alpha}_{FE}$ and $\hat{\beta}_{FE}$ of (3). Defining the $(n_t \times 1)$ vector $a_t = (1, \ldots, 1)'$, these are given by $\hat{\alpha}_{FE} = (\sum_t X_t Q_t X'_t)^{-1} \sum_t X_t Q_t Y_t$ and

$$\hat{\beta}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} (Y_{it} - X'_{it}\hat{\alpha}_{FE}), \quad t = 2, \ldots, T \tag{4}$$

where $Q_t = I_{n_t} - a_t a'_t / n_t$ is the projection matrix taking deviations with respect to time means. For example, a typical element of the matrix $Q_t X'_t$ is $X_{it} - \bar{X}_t$, where $\bar{X}_t = \sum_{i=1}^{n_t} X_{it}/n_t$. The fixed effects estimator has the advantage of being consistent even if X_t is endogenous with respect to time. However, it is inefficient under random effects, and as we will see this inefficiency is particularly strong for our object of interest, i.e., the volatility of β_t.

As an alternative, a random effects approach would assume that $\beta_t \sim N(0, \sigma^2_\beta)$, which yields the possibility to directly estimate volatility σ^2_β of the underlying random process. Identification is achieved by setting the expectation of β_t to zero, so that the restriction $\beta_1 = 0$ is not needed. Stacking for each t the observations Y_{it} into a $(n_t \times 1)$ column vector Y_t, and the explanatory variables into a $(K \times n_t)$, matrix X_t, the model can be written compactly as

$$Y_t = X'_t\alpha + a_t\beta_t + u_t, \quad t = 1, \ldots, T, \quad i = 1, \ldots n_t \tag{5}$$

where $u_t = (u_{1t}, \ldots, u_{n_t,t})'$. As in classical random effects models, we now need to impose exogeneity of the regressors with respect to the time component, i.e., $\mathrm{E}[\beta_t|X] = 0$. This allows to consider $\eta_t = a_t\beta_t + u_t$ as a composite error term with variance $\Omega_t = a_t a_t' \sigma_\beta^2 + \sigma_u^2 I_{n_t}$, and estimate α in the regression $Y_t = X_t'\alpha + \eta_t$ by feasible GLS, $\hat{\alpha}_{GLS} = (\sum_t X_t \hat{\Omega}_t^{-1} X_t')^{-1} \sum_t X_t \hat{\Omega}_t^{-1} Y_t$, where $\hat{\Omega}_t$ is a consistent estimator of Ω_t. In order to test the validity of the exogeneity assumption, a Hausman-type test statistic can be constructed as

$$H = (\hat{\alpha}_{FE} - \hat{\alpha}_{GLS})'(V_{FE} - V_{GLS})^{-1}(\hat{\alpha}_{FE} - \hat{\alpha}_{GLS}), \tag{6}$$

where $V_{GLS} = (\sum_t X_t \hat{\Omega}_t^{-1} X_t')^{-1}$, and $V_{FE} = \hat{\sigma}_u^2 (\sum_t X_t Q_t X_t')^{-1}$. Under the null hypothesis, H has an asymptotic χ^2 distribution with K degrees of freedom. If the null is not rejected, then the exogeneity assumption of X would appear reasonable and $\hat{\alpha}_{GLS}$ is consistent and efficient.

In a second step, the realizations of β_t can be estimated by $\hat{\beta}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} (Y_{it} - X_{it}'\hat{\alpha}_{GLS})$. These $\hat{\beta}_t$ will have a mean close to zero, but $\hat{\beta}_1$ is not necessarily close to zero. One can apply the adjustment $\hat{\beta}_t - \hat{\beta}_1, t = 1, \ldots, T$, if the usual standardization $\beta_1 = 0$ is required in order to obtain an index value of 100 at the beginning of the sample.

We now extend the classical random effects model by introducing assumptions about the dynamics of β_t. In particular, we will assume an autoregressive process or order one, AR(1), including the random walk as a special case:

$$\beta_t = \phi\beta_{t-1} + \xi_t, \tag{7}$$

with $|\phi| \leq 1$ and $\beta_0 = 0$. The system (5)-(7) is a state space representation. If one imposes a normality assumption on both error terms, maximum likelihood and the Kalman filter can be applied to efficiently estimate the state variables β_t. This will be discussed in Section 3.

Before that, however, let us discuss in this dynamic framework the properties of the fixed effects estimator for β_t and the implied estimators of σ_u^2 and σ_ξ^2. Let us assume for simplicity that ϕ is known. For example, a typical choice would be to set $\phi = 1$, meaning that log-prices follow a random walk, and the sequence ξ_t represents the returns. One could estimate ϕ, assuming stationarity, in a two step procedure where in a first step, consistent fixed effects estimates of β_t are obtained, and in a second step, the AR(1) model (7) is estimated. It is however more common to directly assume a random walk for log-prices, which also simplifies the analysis of volatility estimators. Possible model extensions, allowing e.g. for autocorrelations of returns ξ_t, are delegated to Section 4.

Our assumptions are summarized in the following.

(A1) The error terms u_{it} and ξ_t are mutually independent, i.i.d. with mean zero, variances σ_u^2

and σ_ξ^2, respectively, and finite fourth moments.

(A2) The number of observations, n_t, is a positive integer i.i.d. random variable, satisfying $P(n_t \geq 2) > 0$.

Consider the following estimator of σ_u^2:

$$\hat{\sigma}_u^2 = \left(1 - \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}(Y_{it} - \hat{\beta}_t - X_{it}'\hat{\alpha})^2$$

If $n_1 = n_2 = \ldots = n_T = N$, then the estimator is given by

$$\hat{\sigma}_u^2 = \frac{1}{(N-1)T}\sum_{t=1}^{T}\sum_{i=1}^{N}(Y_{it} - \hat{\beta}_t - X_{it}'\hat{\alpha})^2$$

Estimated returns, $\hat{\xi}_t$ say, are obtained by $\hat{\xi}_t = \hat{\beta}_t - \phi\hat{\beta}_{t-1}$, and the variance of returns is estimated by

$$\hat{\sigma}_\xi^2 = \frac{1}{T}\sum_{t=1}^{T}(\hat{\xi}_t - T^{-1}\sum_{j=1}^{T}\hat{\xi}_j)^2 - (1+\phi^2)\hat{\sigma}_u^2\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}$$

For the particular case $n_t = N, t = 1, \ldots, T$, and $\phi = 0$, this estimator becomes

$$\hat{\sigma}_\xi^2 = \frac{1}{T}\sum_{t=1}^{T}(\hat{\xi}_t - T^{-1}\sum_{j=1}^{T}\hat{\xi}_j)^2 - \frac{1}{N}\hat{\sigma}_u^2$$

which is the well known variance estimator in panel data analysis with time and cross section units reversed, see e.g. equation (3.10) of Arellano (2003).

Proposition 1 *Under (A1) and (A2), $\hat{\sigma}_u^2$ and $\hat{\sigma}_\xi^2$ are $\sqrt{T}$-consistent estimators of σ_u^2 and σ_ξ^2, respectively.*

We can further derive the asymptotic distribution of the OLS estimator of $\theta = (\sigma_u^2, \sigma_\xi^2)$, but need an additional distributional assumption.

(A1') The error terms u_{it} and ξ_t are mutually independent with $u_{it} \sim N(0, \sigma_u^2)$ and $\xi_t \sim N(0, \sigma_\xi^2)$.

Clearly, (A1') encompasses and substitutes (A1).

(A3) $\mathrm{E}[|u_{it}|^{4+\delta}] < \infty$ and $\mathrm{E}[|\xi_t|^{4+\delta}] < \infty$, for $\delta > 0$.

Proposition 2 *Under (A1'),(A2) and (A3),*

$$\sqrt{T}(\hat{\theta}-\theta) \to_d N\left(0, 2\sigma_u^4 \lim_{T\to\infty} E[\Sigma_T]\right), \quad \Sigma_T = \begin{pmatrix} \Sigma_{uu,T} & \Sigma_{uv,T} \\ \Sigma_{uv,T} & \Sigma_{vv,T} \end{pmatrix}$$

where

$$\Sigma_{uu,T} = \frac{T^{-1}\sum_{t=1}^{T}(n_t-1)/n_t^2}{\left\{T^{-1}\sum_{t=1}^{T}(n_t-1)/n_t\right\}^2}, \tag{8}$$

$$\Sigma_{uv,T} = -(1+\phi^2)\Sigma_{uu,T}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}, \tag{9}$$

$$\Sigma_{vv,T} = \left(\frac{\sigma_\xi^2}{\sigma_u^2}+\frac{1}{n_t}+\frac{\phi^2}{n_{t-1}}\right)^2 + \frac{2\phi^2}{n_{t-1}^2} + (1+\phi^2)^2\Sigma_{uu,T}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\right)^2 + Var(\frac{1}{n_t}) \tag{10}$$

For the balanced case, i.e., $n_t = N, a.s.$, $t = 1, \ldots, T$, this result reduces to

$$\Sigma_T = \Sigma = \begin{pmatrix} \frac{1}{N-1} & -\frac{(1+\phi^2)}{N(N-1)} \\ -\frac{(1+\phi^2)}{N(N-1)} & \left(\frac{\sigma_\xi^2}{\sigma_u^2}+\frac{1+\phi^2}{N}\right)^2 + \frac{2\phi^2}{N^2} + \frac{(1+\phi^2)^2}{N^2(N-1)} \end{pmatrix}$$

Note that for the large N, large T case, we would obtain $\sqrt{NT}(\hat{\sigma}_u^2 - \sigma_u^2) \to_d N(0, 2\sigma_u^2)$ and $\lim_{N,T\to\infty} \mathrm{Cov}(\hat{\sigma}_u^2, \hat{\sigma}_\xi^2) = 0$. Hence, both variance estimators are independent if sufficient cross-sectional data is available. However, $\sqrt{NT}(\hat{\sigma}_\xi^2 - \sigma_\xi^2)$ diverges since additional cross-sectional data does not increase the information about σ_ξ^2.

In order to assess the effects of an unbalanced panel on the efficiency compared with the balanced panel case, let us assume that $n_t - 1$ follows a Poisson distribution with parameter λ, $Po(\lambda)$. Figure 2 plots the relative efficiencies of the estimators of σ_u^2 and σ_ξ^2, calculated as the ratio of the asymptotic variances under the assumption of a fixed design with $N = 1 + \lambda$ (numerator), and an unbalanced $Po(\lambda)$ design (denominator). While this relative efficiency only depends on the distribution of n_t for σ_u^2, it depends on the population parameters σ_u^2 and σ_ξ^2 for the estimation of σ_ξ^2. For the calculation, we used $\sigma_u^2 = 1$ and $\sigma_\xi^2 = 0.01$, which corresponds to typical empirical estimates (see Section 5). Clearly, the unbalanced design decreases the efficiency of both estimators, but the relative inefficiency disappears as the average number of observations, given by $1 + \lambda$, increases.

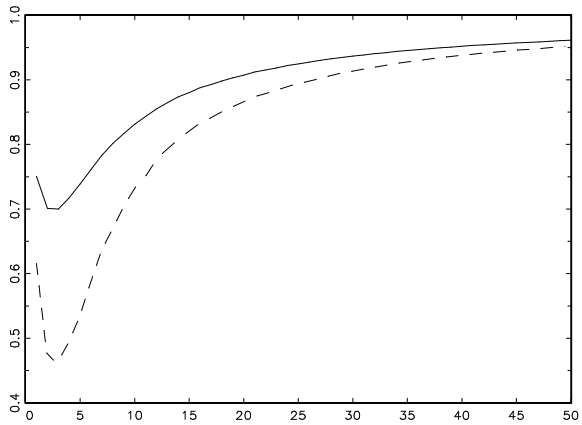

Figure 1: *Relative efficiency of the estimators of σ_u^2 (solid line) and σ_ξ^2 (dashed line), calculated as the ratio of the asymptotic variances under the assumption of a fixed design (numerator), and an unbalanced design with Poisson distribution (denominator). The abscissa represents the parameter λ of the Poisson distribution.*

3 Maximum likelihood estimation

To estimate model (5)-(7), we propose a maximum likelihood estimator combined with the Kalman filter to recover the underlying state variables.

The composite error term $\eta_{it} = u_{it} + \beta_t$ can be obtained as $\eta_{it} = Y_{it} - X'_{it}\hat{\alpha}_{GLS}$. One can write the model (5) as

$$Y_{it} = X'_{it}\alpha + \eta_{it}, \quad t = 1, \ldots, T; i = 1, \ldots n_t \tag{11}$$

The joint model (5)-(7) then reads compactly

$$\eta_t = a_t\beta_t + u_t \tag{12}$$

$$\beta_t = \phi\beta_{t-1} + \xi_t, \tag{13}$$

where $\eta_t = (\eta_{1t}, \ldots, \eta_{n_t,t})'$. This linear Gaussian state space representation (12)-(13) allows us to estimate the underlying β_t, for given parameter estimates, using the Kalman filter. This will

be shown in the following. Note that

$$
\begin{aligned}
(\beta_t|\eta_1,\ldots,\eta_{t-1}) &\sim N(\beta_{t|t-1},\sigma_\beta(t|t-1)) && (14)\\
(\beta_t|\eta_1,\ldots,\eta_t) &\sim N(\beta_{t|t},\sigma_\beta(t|t)) && (15)\\
(\eta_t|\eta_1,\ldots,\eta_{t-1}) &\sim N(\eta_{t|t-1},\Sigma_\eta(t|t-1)) && (16)
\end{aligned}
$$

For a given set of parameters, the conditional means and variances can be obtained using the following Kalman recursions:

1. Prediction step ($t = 1,\ldots,T$)

$$
\begin{aligned}
\beta_{t|t-1} &= \phi\beta_{t-1|t-1} && (17)\\
\sigma^2_\beta(t|t-1) &= \phi^2\sigma^2_\beta(t-1|t-1)+\sigma^2_\xi && (18)\\
\eta_{t|t-1} &= a_t\beta_{t|t-1} && (19)\\
\Sigma_\eta(t|t-1) &= a_t\sigma^2_\beta(t|t-1)a'_t+\sigma^2_u I_{n_t} && (20)
\end{aligned}
$$

2. Correction step ($t = 1,\ldots,T$)

$$
\begin{aligned}
\beta_{t|t} &= \beta_{t|t-1}+\sigma^2_\beta(t|t-1)a'_t\Sigma^{-1}_\eta(t|t-1)(\eta_t-\eta_{t|t-1}) && (21)\\
\sigma^2_\beta(t|t) &= \sigma^2_\beta(t|t-1)-\sigma^4_\beta(t|t-1)a'_t\Sigma^{-1}_\eta(t|t-1)a_t && (22)
\end{aligned}
$$

3. Smoothing step ($t = T-1, T-2,\ldots,1$)

   To estimate the underlying state β_t, one uses the full sample information ($t = 1,\ldots,T$).

$$
\begin{aligned}
\beta_{t|T} &= \beta_{t|t}+\phi\frac{\sigma^2_\beta(t|t)}{\sigma^2_\beta(t+1|t)}\left\{\beta_{t+1|T}-\beta_{t+1|t}\right\} && (23)\\
\sigma^2_\beta(t|T) &= \sigma^2_\beta(t|t)+\phi^2\frac{\sigma^4_\beta(t|t)}{\sigma^4_\beta(t+1|t)}\left\{\sigma^2_\beta(t+1|T)-\sigma^2_\beta(t+1|t)\right\} && (24)
\end{aligned}
$$

Parameter estimation can be achieved in an efficient and straightforward way by maximum likelihood. Denote the parameter vector by $\theta = (\sigma^2_\xi, \sigma^2_u)$ and define the parameter space $\Theta = \{\theta : \sigma^2_\xi > 0, \sigma^2_u > 0\}$. If stationarity is imposed on the AR(1) model in (7), that is, $|\phi| < 1$,

then ϕ could be included in θ and be jointly estimated with σ_ξ^2 and σ_u^2. We do not discuss this possibility further, however, since we want explicitly to allow for the unit root case, $\phi = 1$.

Let $e_t(\theta) = \eta_t - \eta_{t|t-1}$ and $\Sigma_t(\theta) = \Sigma_\eta(t|t-1)$. Then, the log-likelihood, up to an additive constant, can be written as

$$L(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left\{\ln(|\Sigma_t(\theta)|) + e_t(\theta)'\Sigma_t(\theta)^{-1}e_t(\theta)\right\} \tag{25}$$

and the maximum likelihood estimator is defined as

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L(\theta),$$

with parameter space $\Theta = \mathcal{R}_+^2$. The maximization problem has no analytical solution, but numerical methods can be used conveniently. In large dimensions, computational problems may arise due to the optimization of a function that involves frequent calculation of the determinant and inverse of high dimensional matrices. We can exploit however the particular structure of Σ_t to obtain explicit formulas that largely facilitate the optimization. It can easily be shown that $|\Sigma_t| = \sigma_u^{2(n_t-1)}(n_t\zeta_{t-1} + \sigma_u^2)$ and $\Sigma_t^{-1} = (n_t\zeta_{t-1} + \sigma_u^2)^{-1}a_t a_t'/n_t + (I_{n_t} - a_t a_t'/n_t)/\sigma_u^2$, where $\zeta_t = \phi^2\sigma_\beta^2(t|t) + \sigma_\xi^2$. Using these expressions in (25) reduces computational costs substantially.

The term σ_ξ, the estimated standard deviation of ξ, corresponds to the volatility of returns of the underlying portfolio.

In order to obtain asymptotic theory, we need the following additional assumption.

(A4) Assume that θ_0, the true parameter vector, is an interior point of Θ.

Proposition 3 *Under (A1'), (A2), (A3) and (A4), the MLE of θ is consistent and asymptotically normally distributed,*

$$\sqrt{T}(\hat{\theta} - \theta_0) \to_d N\left(0, \lim_{T\to\infty}\left(\frac{I(\theta)}{T}\right)^{-1}\right)$$

where

$$\begin{aligned} I(\theta) &= -E\left[\frac{\partial^2 \ln L(\theta)}{\partial\theta\partial\theta'}\right] \\ &= \frac{1}{2}\sum_{t=1}^{T}\left\{\frac{\partial vec(\Sigma_t)'}{\partial\theta}(\Sigma_t^{-1}\otimes\Sigma_t^{-1})\frac{\partial vec(\Sigma_t)}{\partial\theta'} + 2E\left(\frac{\partial e_t'}{\partial\theta}\Sigma_t^{-1}\frac{\partial e_t}{\partial\theta'}\right)\right\} \end{aligned}$$

Analytical expressions for the derivatives used to calculate $I(\theta)$ are provided in Appendix B.

Finally, for the special case $\phi = 0$, we obtain $\Sigma_t = \sigma_\xi^2 a_t a_t' + \sigma_u^2 I_{n_t}$, $e_t(\theta) = \eta_t$, $\partial e_t(\theta)/\partial\theta = 0$, and $\partial \text{vec}(\Sigma_t)/\partial\theta = (\text{vec}(I_{n_t})', a_t')'$. Straightforward calculations show that, for this special case, $I(\theta)$ reduces to

$$I(\theta) = \frac{T}{2\sigma_u^4} \begin{pmatrix} N - 1 + (N\frac{\sigma_\xi^2}{\sigma_u^2} + 1)^{-2} & N(N\frac{\sigma_\xi^2}{\sigma_u^2} + 1)^{-2} \\ N(N\frac{\sigma_\xi^2}{\sigma_u^2} + 1)^{-2} & N^2(N\frac{\sigma_\xi^2}{\sigma_u^2} + 1)^{-2} \end{pmatrix},$$

yielding the asymptotic covariance matrix

$$\lim_{T\to\infty} \left(\frac{I(\theta)}{T}\right)^{-1} = 2\sigma_u^4 \begin{pmatrix} \frac{1}{N-1} & -\frac{1}{N(N-1)} \\ -\frac{1}{N(N-1)} & \left(\frac{\sigma_\xi^2}{\sigma_u^2} + \frac{1}{N}\right)^2 + \frac{1}{N^2(N-1)} \end{pmatrix} \tag{26}$$

Comparing (26) with (2), we see that for the special case $\phi = 0$, the MLE estimator has the same asymptotic distribution as the OLS estimator and, hence, both estimators are asymptotically equivalent. If $\phi \neq 0$, however, the estimators are different. In the following we discuss their efficiency.

We consider several scenarios in order to compare the efficiency of the OLS and maximum likelihood estimators of volatility. Since log prices are usually assumed to follow a random walk, we set $\phi = 1$. Moreover, we assume that η_{it} are observed directly, in order to focus on the estimation of θ without needing to estimate α. It may be expected that MLE of θ is even more efficient relative to OLS if α is estimated jointly with θ.

To further simplify the analysis, note that only the ratio of σ_u^2 and σ_ξ^2 is of interest, since the scaling of the data η_{it} is irrelevant. Hence, we set σ_u to one without loss of generality. We assume a balanced panel with $N = 5, 10, 20$ and 50 observations per period. Define the asymptotic relative efficiency as

$$\lim_{T\to\infty} \frac{\text{Var}(\hat{\theta}_{MLE})}{\text{Var}(\hat{\theta}_{OLS})}$$

which, if MLE is more efficient than OLS, is a number between 0 and 1. Table 1 reports the asymptotic relative efficiencies.

Note that in all situations, the OLS estimator of σ_u^2 is almost as efficient as the ML estimator. However, this is not the case for our parameter of interest, the variance of index returns, σ_ξ^2. Here, the efficiency loss of OLS is remarkable in cases where σ_ξ is small, even if N is large. Figure 2 depicts the relative efficiencies of the estimator of σ_ξ^2. Clearly, for σ_ξ^2 close enough to zero, the relative efficiency is arbitrarily small no matter how large N. This motivates the ML estimator, knowing that small values of σ_ξ are empirically relevant as we will see in Section 5.

Table 1: Asymptotic relative efficiency of $\hat{\theta}_{OLS}$ w.r.t. $\hat{\theta}_{MLE}$.

σ_ξ^2	$N=5$		$N=10$		$N=20$		$N=50$	
	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_u^2$	$\hat{\sigma}_\xi^2$	$\hat{\sigma}_u^2$
0.1	0.0225	0.9341	0.0588	0.9912	0.1213	0.9993	0.2725	1.0000
0.2	0.1162	0.9880	0.2258	0.9998	0.3880	0.9998	0.6617	0.9998
0.3	0.2481	0.9998	0.4175	0.9991	0.6241	0.9992	0.8538	0.9997
0.4	0.3806	0.9975	0.5801	0.9973	0.7728	0.9987	0.9272	0.9997
0.5	0.4958	0.9919	0.6973	0.9956	0.8543	0.9984	0.9558	0.9996
0.6	0.5881	0.9865	0.7749	0.9944	0.8975	0.9981	0.9680	0.9996
0.7	0.6582	0.9819	0.8241	0.9935	0.9208	0.9980	0.9738	0.9996
0.8	0.7098	0.9782	0.8550	0.9928	0.9337	0.9979	0.9766	0.9996
0.9	0.7469	0.9752	0.8745	0.9923	0.9411	0.9978	0.9782	0.9996
1	0.7733	0.9729	0.8870	0.9919	0.9455	0.9977	0.9791	0.9996

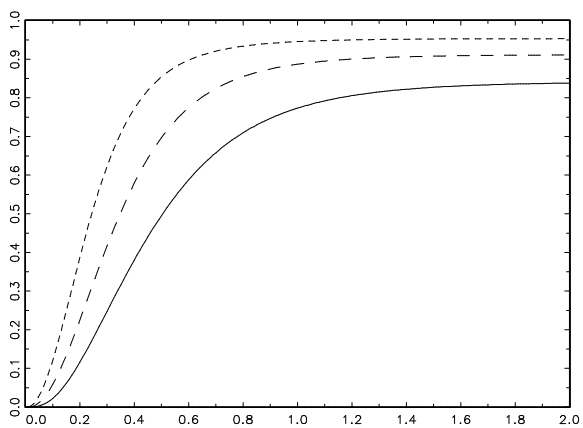

Figure 2: *Asymptotic relative efficiency of the estimator of σ_ξ^2 using OLS versus MLE. The value of σ_ξ^2 is on the abscissa, σ_u^2 and ϕ are fixed at 1. The curves are for $N=5$ (solid), $N=10$ (long dashed) and $N=20$ (short dashed).*

4 Model extensions

In this section we will discuss three possible extensions of the model: First, the inclusion of a drift term in the random walk characterizing market prices. Second, the possibility of autocorrelation in returns. And finally, allowing for time-varying volatility.

4.1 Non-zero mean of returns

Instead of assuming a random walk with mean zero for β_t, we could add a constant drift parameter γ and replace (13) by $\beta_t = \gamma + \phi\beta_{t-1} + \xi_t$. The only change in the Kalman filter would be in equation (17), which would be replaced by $\beta_{t|t-1} = \gamma + \phi\beta_{t-1|t-1}$. The drift γ would have to be estimated by MLE, jointly with σ_u and σ_ξ. Alternatively, one could detrend the data in a first step and instead of (11) estimate $Y_{it} = X'_{it}\alpha + \gamma t + \eta_{it}$ by OLS. The composite error $\eta_{it} = \beta_t + u_{it}$ would then have, by construction, mean zero without linear time trend. Returns would be estimated by adding the OLS estimate of γ to the residuals $\hat{\xi}_t$. This latter procedure would be convenient but less efficient than the former.

Rather than explicitly modelling non-zero means of returns, it should be noted that the Kalman filter of the model without drift at least partially captures a potential non-zero mean of returns, which would end up in a non-zero mean of residuals $\hat{\xi}_t$. To see this, consider the updating equation for β_t, (21). If the Kalman filter without drift is used but the true model contains a drift, then the prediction error $\eta_t - \eta_{t|t-1}$ is equal to $\gamma + u_t$. Straightforward calculations show that the second term on the right hand side of (21) would be given by

$$\frac{n_t\sigma_\beta^2(t|t-1)\gamma + a'_t u_t}{n_t\sigma_\beta^2(t|t-1) + \sigma_u^2}$$

which, conditional on n_t and letting n_t increase, converges to γ in probability. Hence, (21) corrects the predicted β_t by the neglected γ, if the cross-section information is sufficiently large. For the estimated β_t it therefore does not make a difference whether or not a trend is included. An explicit estimation of γ would have the advantage of possible inference concerning the drift term, but it does not matter for the subsequent modelling and estimation of volatility.

4.2 Autocorrelation of returns

Markets for heterogenous goods may deliver returns that are serially correlated. For real estate markets, this has been motivated by Schulz and Werwatz (2004). It is possible to extend our basic model to account for serial correlation. Consider, for example, the random walk $\beta_t = \beta_{t-1} + \xi_t$, where now ξ_t itself follows an AR(1) model, $\xi_t = \rho\xi_{t-1} + v_t$, with $|\rho| < 1$ and v_t white noise. This

can be written as an AR(2) model with parameter constraints, i.e., $\beta_t = (1+\rho)\beta_{t-1} - \rho\beta_{t-2} + v_t$. We can then define a new state vector $(\beta_t, \beta_{t-1})'$ and a transition equation

$$\begin{pmatrix} \beta_t \\ \beta_{t-1} \end{pmatrix} = \begin{pmatrix} 1+\rho & -\rho \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_{t-1} \\ \beta_{t-2} \end{pmatrix} + \begin{pmatrix} v_t \\ 0 \end{pmatrix}$$

The Kalman filter equations can then be extended easily to this case. The parameter ρ could be estimated jointly with the other model parameters by maximum likelihood.

In the empirical part, we will estimate the model without autocorrelation of returns, and then test for residual autocorrelation using standard Portmanteau-type tests.

4.3 Time-varying volatility

With the enormous experience on time-varying volatility in financial and other markets, it seems doubtful that markets with heterogenous goods have constant volatility. Having information on possibly time-varying volatility, for example by rejecting the hypothesis of an absence of structural breaks, one may want to generalize the above model to allow for time-varying volatility. It is a priori difficult to guess which pattern volatility may follow. One could assume, as Hodgson and Vorkink (2004) suggest, that returns follow a GARCH type process, as it has been standard for financial markets. There are however three drawbacks of this approach. First, data sets of heterogenous markets typically have a much smaller time dimension, which renders estimation imprecise and highly dependent on starting values. Second, due to the high degree of time-aggregation, short term fluctuations of volatility may have been averaged out such that GARCH effects become insignificant, as it is also the case in Hodgson and Vorkink (2004). Third, estimation of the GARCH part could only feasibly be done in a second step, having estimated first the index returns, e.g. by OLS. This two-step procedure is inefficient, and it would be desirable to develop a framework where the model components can be estimated in one step.

In the following, we propose a nonparametric extension of the model presented in Section 3, letting both market and idiosyncratic volatility be unknown functions of time that can be estimated with nonparametric methods. The approach is similar in spirit to the estimation of long-run trends of volatility in financial markets, as in the spline GARCH model of Engle and Rangel (2008).

We can regard $\theta = (\sigma_u^2, \sigma_\xi^2)'$ as a smooth function of time, $\theta(\tau)$, and obtain an estimate thereof via the local maximum likelihood approach, which has been discussed in a unified framework by Fan, Farmen and Gijbels (1998). In the following we apply their main ideas to our problem.

The local likelihood estimator is defined as

$$\widehat{\theta}(\tau) = \text{argmax}_\theta\{L(\theta \mid \tau)\},$$

where $\theta = \theta(\tau)$ and

$$L(\theta \mid \tau) = -\frac{1}{2}\sum_{t=1}^{T}\left\{\ln(|\Sigma_t(\theta)|) + e_t(\theta)'\Sigma_t(\theta)^{-1}e_t(\theta)\right\} K\left(\frac{t-\tau}{h}\right) \tag{27}$$

which gives estimates of time-varying idiosyncratic and market volatility. This approach fits locally a constant to the unknown volatilities, weighted by a kernel function K and bandwidth h. One could extend this approach to local polynomial fitting, often giving more precise estimates especially at the boundaries of the support. Furthermore, one can estimate bias and variance of the estimator by fitting locally a polynomial of higher order. For the local constant estimator it suffices to fit, in a second step, a local linear model where the term $\Sigma_t(\theta(\tau))$ in (27) is replaced by $\Sigma_t(\theta(\tau) + \theta'(\tau)(t-\tau))$, where $\theta'(\tau)$ is the first derivative of θ evaluated at τ. Similarly, $e_t(\theta(\tau))$ is replaced by $e_t(\theta(\tau)+\theta'(\tau)(t-\tau))$. The resulting local likelihood function is more precise than (27) and permits to obtain bias and variance estimates. The pilot bandwidth to compute the local linear estimator can be chosen according to the extended residual squares criterion, and the optimal bandwidth for the local constant estimator minimizes the estimated mean squared error, integrated over time.

Finally, pointwise confidence intervals can be obtained by invoking asymptotic normality of the local constant likelihood estimator and using the estimates of its bias and variance.

5 Volatility of the art market

A set of data provided by Artnet AG[1] and Tutela Capital S.A.[2] is used to illustrate the methodology. It concerns artworks sold at auction between January 2001 and December 2011 and consists of 11'521 paintings made by 40 artists who had the biggest volume of sales at auction in 2008 and 2009[3].

[1] A provider of data related to art. `www.artnet.com`

[2] A company specialized in managing art as an asset class. `www.tutelacapital.com`

[3] These artists are Jean-Michel Basquiat (1960-1988), George Braque (1882-1963), Alexander Calder (1898-1976), Mark Chagall (1887-1985), Edgar Degas (1834-1917), Kees van Dongen (1877-1968), Raoul Dufy (1877-1953), Max Ernst (1891-1976), Lucio Fontana (1899-1968), Sam Francis (1923-1994), Paul Gauguin (1848-1903), Childe Hassam (1859-1935), Damien Hirst (1965-), Alexej von Jawlensky (1864-1941), Wassily Kandisky (1866-1944), Ernst Ludwig Kirchner (1880-1938), Paul Klee (1879-1940), Willem de Kooning (1904-1997), Yayoi Kusama (1929-), Rene Magritte (1898-1967), Henri Matisse (1869-1954), Joan Miro (1893-1983), Claude Monet (1840-1926), Henry Moore (1831-1895), Edvard Munch (1863-1944), Emil Nolde (1867-1956), Pablo Picasso (1881-1973),

First, logged prices are regressed on available characteristics using ordinary least squares (OLS) without time dummies. The explanatory variables are the artist's name (40 levels), the medium used by the artist (35 levels), the height and width of the artwork in cm, the nationality of the artist (14 levels), the estimated date when the artwork was realized, the auction house where the sale took place (97 levels), whether the price in the database includes the buyer's premium or not, and the country in which the sale happened. We applied three methods to select variables: stepwise forward, stepwise backward and autometrics[4], all with a 5% significance level. The backward selection kept 119 variables in the model, the forward selection 102, and the autometrics procedure kept 111 variables. 89 variables are common to the forward and backward selection, 80 variables are common to the forward and autometrics selection procedures, while 90 variables are common to the backward and autometrics procedures. Results of estimated returns and volatilities are robust to the choice of the selection procedure, and we therefore only report the results for the autometrics procedure. The adjusted R^2 for all three selected models is about 60%. The final estimation results are summarized in Tables 5 to 9 in Appendix D. To economize on space, only the OLS estimates are reported, the GLS estimates being very close to these. For the selected model, we also calculated the fixed effects OLS estimator, i.e., the OLS estimator of the model including time dummies. The Hausman test in (6) takes the value 2.28, which is insignificant at 1%, hence supporting our assumption of exogeneity of X. Furthermore, as indicated by the variance inflation factors given in Table 9, the final model does not encounter problems due to multicollinearity.

The estimated β_t are computed using the fixed effects (OLS) and MLE estimators. Figure 3 plots the index on a semi-annual basis with both methodologies, while Figure 4 depicts the corresponding returns. Apart from the last semester, where less observations were available, both estimates are almost indistinguishable. This reflects the fact that both estimators are consistent under exogeneity of X, and we are having several hundreds of observations per period.

The mean of estimated returns, $T^{-1}\sum_{t=1}^{T}\hat{\xi}_t$, is 0.0713 for OLS and 0.0666 for MLE, corresponding to annualized returns of about 14% (OLS) and 13% (MLE), substantially higher than the mean annualized returns for the S&P 500 over the same period (about 0% annual return).

The pattern of the estimated index and its returns is remarkable. Negative returns of 2008 to 2009 reflect the direct impact of the banking crisis on the art market. Several concurring factors contribute to explain the drop in prices. First, a negative shock on demand for art as a consumption good may have hit the auction market, as the number of ultra-high net worth

Camille Pissarro (1831-1903), Richard Prince (1949-), Pierre-Auguste Renoir (1841-1919), Gerhard Richter (1932-1984), Mark Rothko (1903-1970), Egon Schiele (1890-1918), Alfred Sisley (1839-1899), Henri de Toulouse-Lautrec (1864-1901), Maurice de Vlaminck (1876-1958), Edouard Vuillard (1868-1940), Andy Warhol (1928-1987).

[4]See appendix C for an explanation of the autometrics procedure.

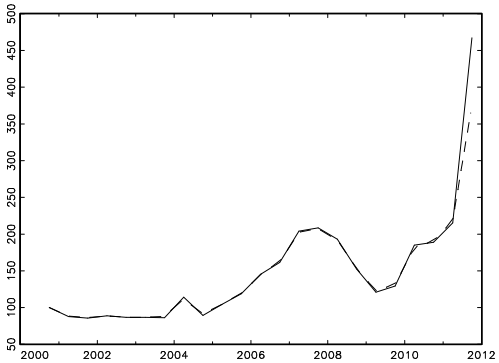

Figure 3: Semi-annual price index for blue chip artists. The index is set to 100 in the second semester of 2000. The solid line corresponds to the fixed effects estimator (4), the dashed one to the smoothed estimator $\beta_{t|T}$ of the Kalman filter using MLE.

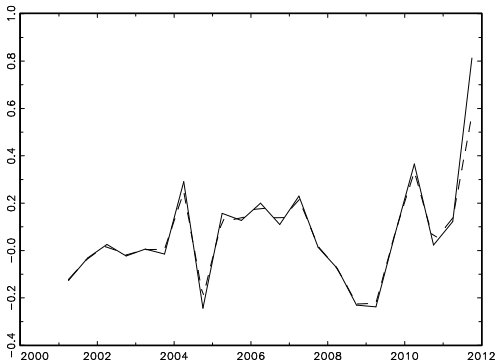

Figure 4: Semi-annual returns for blue chip artists. The solid line corresponds to the OLS estimator $\hat{\beta}_t - \hat{\beta}_{t-1}$, the dashed one to the estimator $\beta_{t|T} - \beta_{t-1|T}$ of the Kalman filter using MLE.

individuals dropped in 2008 and 2009. Goetzmann et al. (2010) showed the positive relationship between top-income and art prices. Second, another negative shock on demand came from banks and financial institutions that may have frozen their acquisitions. Indeed, at a time of corporate jets and bonuses being under scrutiny in government-controlled banks, acquisition or sponsoring of luxury goods such as artworks by bailed-out financial institutions may have been seen as an undesirable spending. Third, risks of large liquidation of banking collection may have put pressure on prices as failed and bailed-out banks were expected to sell their collections. After Lehman Brothers in 2008, RBS and Bank of Ireland announced that they would liquidate parts of their collection in 2009 and 2010. Although these collections were considered relatively small (less than 50 million USD), risks of a major collection being liquidated did exist and could have materialized. UBS in particular, host of a very large art collection (more than 35'000 artworks according to their website) suffered a massive loss in 2008. The bank is well known for sponsoring many art-related activities (including the world-class fair Art Basel) but precipitately shut down in April 2009 its unit that advised wealthy Swiss clients on acquiring art. This decision may have been seen by art market participants as a move towards divestment from the art market.

In 2011, a different pattern took place with large, positive returns. As the European debt crisis spread, fearful investors may have become eager to diversify their portfolios in safe-haven assets such as investment grade art. The safe haven properties of artworks (especially blue chips artworks) may have surpassed their consumption properties, triggering a rally similar to the one experienced by gold. Oosterlinck (2010) showed that in the specific case of World War II (an example of a major global crisis), art outperformed all other asset classes except gold, as one of a few viable investment goods in a highly uncertain environment. On the other hand, however, the positive spike in the second semester of 2011 may have been distorted by extreme events in the market at that time. Sotheby's November auctions of contemporary art saw artworks by Gerhard Richter and by Sam Francis be hammered at extraordinary price levels for these artists. Only 71 artworks are used to estimate returns for the second semester of 2011, of which 12 are by Francis and Richter. Further studies with more complete data sets will have to investigate whether the recent surge is genuine or not.

We now turn to the volatility estimation. Table 2 reports the estimation results for the full sample. As expected, market volatility is much lower than idiosyncratic volatility, but the OLS estimate of market volatility is about 8% higher than the corresponding MLE estimate. It is likely that OLS, being less efficient than MLE and not taking into account the time variation of β_t, overestimates market volatility. In order to see whether our distributional assumptions of the error terms are reasonable, we show in Table 3 summary statistics of the estimated residuals. The Jarque-Bera normality test does not reject normality for ξ_t, it does so however for u_{it},

mainly due to the high kurtosis. This is similar to financial markets, where leptokurtosis is often still present in residuals, even after standardizing with volatility estimates. In our case, the non-normality of u_{it} implies that the Kalman filter used in MLE is not fully efficient. Even though we do not expect major gains in efficiency using more general filtering algorithms, this may be a line of future research.

Table 4 reports empirical autocorrelations $\hat{\rho}(h)$ of estimated residuals $\hat{\xi}_t$ and portmanteau statistics of order h, $Q(h) = T^2 \sum_{i=1}^{h} (T-i)^{-1} \hat{\rho}(h)^2$. Under H_0 of white noise, $Q(h)$ has an asymptotic χ^2 distribution with h degrees of freedom. The empirical p-values indicate that we do not reject the null, which confirms our modelling approach.

In order to gauge parameter stability, we estimate the model additionally for two subsamples, results of which are reported in Table 2. Obviously, both estimated idiosyncratic and market volatilities are lower in the first subsample than in the second. A formal test of parameter constancy, $H_0 : \sigma^2_{u,1} = \sigma^2_{u,2}, \sigma^2_{\xi,1} = \sigma^2_{\xi,2}$, is the likelihood ratio test. Let L^*_i denote the log-likelihood of the ith subsample. Then, the LR statistic is given by $LR = 2(L^*_1 + L^*_2 - L)$ and has under the null an asymptotic χ^2 distribution with two degrees of freedom. In our case, the LR statistic takes the value 7.7626, the corresponding p-value is 0.0206, and parameter stability is rejected. We therefore turn to extensions of the basic model allowing for time-varying volatilities.

		σ_u	σ_ξ	log likelihood
Full sample	OLS	1.2332 (0.0962)	0.0471 (0.0132)	
	MLE	1.2062 (0.0399)	0.0438 (0.0207)	-6873.8531
01/2001-	OLS	1.1870 (0.0970)	0.0162 (0.0118)	
01/2006	MLE	1.1660 (0.0690)	0.0206 (0.0111)	-2694.3886
02/2006-	OLS	1.2756 (0.0962)	0.0793 (0.0178)	
02/2011	MLE	1.2332 (0.0465)	0.0637 (0.0439)	-4175.5832

Table 2: *Parameter estimates of the static model using OLS and MLE. The first column reports the mean of estimated returns $\hat{\xi}_t$. Asymptotic standard errors are given in parentheses.*

We estimate a model with smoothly time-varying idiosyncratic and market volatilities using the local likelihood estimator of Section 4.3 with Gaussian kernel and bandwidth chosen as the minimizer of the estimated mean integrated squared error. Figure 5 depicts the estimate of idiosyncratic volatility, $\sigma_u(\tau)$, which shows an increasing trend in the second part of the sample, but overall the variation seems minor considering the scale of the estimate. The increase from 2006 to 2011 is about 20%. Pointwise 95% confidence bands are slightly wider around 2005 due to less available data in that year.

		mean	std.dev.	skewness	kurtosis	JB
ξ_t	OLS	0.0713	0.2316	1.2740	5.4149	11.29
	MLE	0.0666	0.1861	0.5858	3.5420	1.52
u_{it}	OLS	0.0000	1.0970	-0.2806	8.1747	13007.02
	MLE	0.0585	1.1112	-0.3156	8.2852	13187.51

Table 3: *Summary statistics for $\hat{\xi}_t$ and $\hat{u}_{it}$ in the constant volatility model. JB is the Jarque-Bera test statistic, which under normality has an asymptotic χ^2_2 distribution.*

h	1	2	3	4
ACF(h)	0.1837	-0.0543	-0.3112	-0.1550
$Q(h)$	0.8120	0.8864	3.4487	4.1180
p-value	0.3675	0.6420	0.3275	0.3903

Table 4: *Autocorrelation function of order h of residuals $\hat{\xi}_t$, corresponding Portmanteau statistics $Q(h)$ and p-values.*

Figure 6 shows the local likelihood estimate of market volatility, $\sigma_\xi(\tau)$. Recall from Table 2 that the constant likelihood estimate is 0.0438. We now see a much smaller estimate at the beginning of the sample, around 2002 and 2003, which then increases and remains at about the same level of 0.04 between 2005 and 2008. After the financial crisis 2008/09, the estimate of market volatility increased further and reached levels around 0.2 in 2011. The wide 95% pointwise confidence bands at the end of the sample indicate however that this estimate is highly imprecise.

6 Conclusion

The widespread use of the hedonic regression methodology in the economics of heterogeneous goods has led academics and business practitioners to devise risk metrics from price indices as if they were directly measured. We have shown that the standard deviation of estimated returns overestimates market volatility and needs to be corrected by taking into account the idiosyncratic volatility. We have further shown that in a framework where the market index follows a random walk, or a stationary autoregressive process, important efficiency gains of the volatility estimator can be obtained by using maximum likelihood in combination with the Kalman filter. As an extension, we propose a nonparametric approach to allow for time-varying volatility.

The application to a blue chips art market has shown that returns declined during the

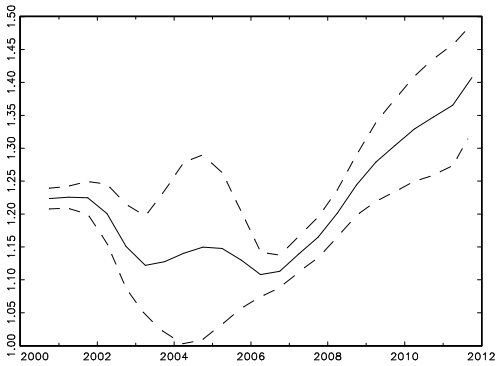

Figure 5: Idiosyncratic volatility of the blue chips art market, estimated by local maximum likelihood.

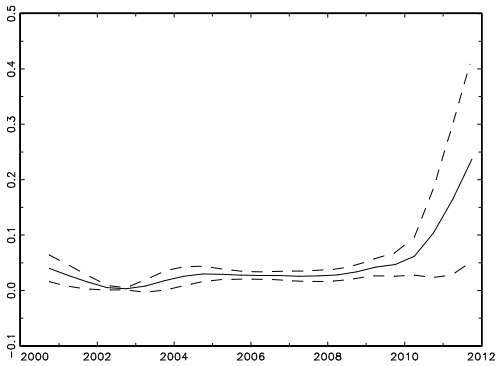

Figure 6: Market volatility of the blue chips art market, estimated by local maximum likelihood.

financial crisis 2008/09 but increased during the recent European debt crisis. We may suspect art to be considered as a safe haven in crisis times, but our dataset needs to be augmented to confirm this for the recent debt crisis. One of the reasons why prices fell in 2008/09 was that important collections of failed banks such as Lehman Brothers had to be sold, putting pressure on supply, while demand from high net worth individuals was dropping. On the other hand, market volatility has in both crises increased substantially, similar to the financial markets. In future work, one may model explicitly time-varying correlations between art, financial and other assets to gauge the diversification benefits of including alternative assets in the portfolio. The modelling framework developed in this paper naturally permits to include other assets and estimate time-varying correlations, which is not feasible in classical OLS estimation.

On the econometrics side, future work may consider more general filters than the Kalman filter to accomodate departures from normality in the error terms. Further efficiency gains may be expected.

Appendix A: Proofs

Proof of Proposition 1. The first part of the proposition concerns the estimator of σ_u^2. We have $\hat{\beta}_t = \beta_t + \bar{u}_t$, with $\bar{u}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} u_{it}$, and hence $Y_{it} - X_{it}'\alpha - \hat{\beta}_t = u_{it} - \bar{u}_t$. Consider the expression

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}(Y_{it} - X_{it}'\alpha - \hat{\beta}_t)^2 \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}(u_{it} - \bar{u}_t)^2 \\
= \; & \sigma_u^2 + \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}(u_{it}^2 - \sigma_u^2) - \frac{2}{T}\sum_{t=1}^{T}\frac{1}{n_t^2}\sum_{i=1}^{n_t}u_{it}^2 - \frac{2}{T}\sum_{t=1}^{T}\frac{1}{n_t^2}\sum_{i=1}^{n_t}\sum_{j\neq i}u_{it}u_{jt} + \frac{1}{T}\sum_{t=1}^{T}\bar{u}_t^2
\end{aligned}$$

The second and fourth terms are the means of independent r.v. with mean zero and finite variance by Assumption (A1). Hence, the Chebychev weak law of large numbers applies to these terms, which are $O_p(T^{-1/2})$. In the same vein, $\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t^2}\sum_{i=1}^{n_t}u_{it}^2 = \sigma_u^2\mathrm{E}[n_t^{-1}] + O_p(T^{-1/2})$ and $\frac{1}{T}\sum_{t=1}^{T}\bar{u}_t^2 = \sigma_u^2\mathrm{E}[n_t^{-1}] + O_p(T^{-1/2})$. Thus, we have $\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}(Y_{it} - \hat{\beta}_t)^2 = \sigma_u^2(1 - \mathrm{E}[n_t^{-1}]) + O_p(T^{-1/2})$. It immediately follows that $\hat{\sigma}_u^2 = \sigma_u^2 + O_p(T^{-1/2})$, as stated.

The second part of the proposition concerns the estimator of σ_ξ^2. Note that $\hat{\xi}_t = \xi_t + \bar{u}_t - \phi\bar{u}_{t-1}$, where we denote $\bar{u}_t = \frac{1}{n_t}\sum_{i=1}^{n_t}u_{it}$, with $\bar{u}_t \sim N(0, \sigma_u^2/n_t)$. Consider the naive estimator $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{\xi}_t - \frac{1}{T}\sum_{j=1}^{T}\hat{\xi}_j)^2 = \frac{1}{T}\sum_{t=1}^{T}\xi_t^2 + \frac{1}{T}\sum_{t=1}^{T}\bar{u}_t^2 + \frac{\phi^2}{T}\sum_{t=2}^{T}\bar{u}_{t-1}^2 + O_p(T^{-1/2})$.

Since $\frac{1}{T}\sum_{t=1}^{T}\bar{u}_t^2 \to_p \sigma_u^2 \mathrm{E}[n_t^{-1}]$, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{\xi}_t - \frac{1}{T}\sum_{j=1}^{T}\hat{\xi}_j)^2 = \sigma_\xi^2 + (1+\phi^2)\sigma_u^2\mathrm{E}[n_t^{-1}] + O_p(T^{-1/2})$$

and it follows using the first part of the proposition that the estimator $\hat{\sigma}_\xi^2$ is $\sqrt{T}$-consistent, as stated. Q.E.D.

Proof of Proposition 2.

Asymptotic normality follows from an application of the Liapounov central limit theorem. By Proposition 1, the estimators are asymptotically unbiased. It remains to compute the asymptotic variance. Define the information set generated by the number of observations by $\mathcal{N}_T = \sigma(n_1, n_2, \ldots, n_T)$. Then, we have the variance decomposition $\mathrm{Var}(\sqrt{T}\hat{\sigma}_u^2) = \mathrm{Var}(\mathrm{E}[\sqrt{T}\hat{\sigma}_u^2 \mid \mathcal{N}_T]) + \mathrm{E}[\mathrm{Var}(\sqrt{T}\hat{\sigma}_u^2)|\mathcal{N}_T]$. The first term is zero since $\mathrm{E}[\hat{\sigma}_u^2 \mid \mathcal{N}_T] = \sigma_u^2$. The second term is

$$\mathrm{E}[\mathrm{Var}(\sqrt{T}\hat{\sigma}_u^2)|\mathcal{N}_T] = 2\sigma_u^4 \frac{T^{-1}\sum_{t=1}^{T}(n_t-1)/n_t^2}{\left\{T^{-1}\sum_{t=1}^{T}(n_t-1)/n_t\right\}^2},$$

which delivers $\Sigma_{uu,T}$. Next, the covariance is given by

$$\mathrm{Cov}(\sqrt{T}\hat{\sigma}_u^2, \sqrt{T}\hat{\sigma}_\xi^2) = \mathrm{Cov}(\sqrt{T}\hat{\sigma}_u^2, \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\xi}_t^2) - (1+\phi^2)\mathrm{Cov}(\sqrt{T}\hat{\sigma}_u^2, \hat{\sigma}_u^2\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n_t})$$

It is straightforward to show that the first term on the right hand side is zero. The covariance in the second term is given by

$$\mathrm{Cov}(\sqrt{T}\hat{\sigma}_u^2, \hat{\sigma}_u^2\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n_t}) = \mathrm{E}[\mathrm{Var}(\sqrt{T}\hat{\sigma}_u^2|\mathcal{N}_T)\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}] = 2\sigma_u^4\mathrm{E}\frac{T^{-1}\sum_{t=1}^{T}(n_t-1)/n_t^2}{\left\{T^{-1}\sum_{t=1}^{T}(n_t-1)/n_t\right\}^2}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t},$$

which gives $\Sigma_{uv,T}$.

We now prove the expression for the asymptotic variance of $\hat{\sigma}_\xi^2$. First note that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{\xi}_t - \frac{1}{T}\sum_{j=1}^{T}\hat{\xi}_j)^2 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\xi}_t^2 - \sqrt{T}(\frac{1}{T}\sum_{j=1}^{T}\hat{\xi}_j^2)^2$, where the second term on the right hand side is $O_p(T^{-1/2})$ and, hence, asymptotically negligible. We have

$$\lim_{T\to\infty}\mathrm{Var}(\sqrt{T}\hat{\sigma}_\xi^2) = \lim_{T\to\infty}\mathrm{Var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\xi}_t^2) + (1+\phi^2)^2\lim_{T\to\infty}\mathrm{Var}(\hat{\sigma}_u^2\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n_t}) \tag{28}$$

since again the covariance term disappears. Note first that $\hat{\xi}_t^2 = \xi_t^2 + \bar{u}_t^2 + \phi^2\bar{u}_{t-1}^2 + 2\xi_t\bar{u}_t -$

$2\phi\xi_t\bar{u}_{t-1} - 2\phi\bar{u}_t\bar{u}_{t-1}$. All terms are mutually orthogonal and, by assumption, $\xi_t \sim N(0, \sigma_\xi^2)$ and $\bar{u}_t \sim N(0, \sigma_u^2/n_t)$. Hence, $\mathrm{Var}(\hat{\xi}_t^2|\mathcal{N}_T) = 2\sigma_\xi^4 + 2\sigma_u^4/n_t^2 + 2\phi^4\sigma_u^4/n_{t-1}^2 + 4\sigma_\xi^2\sigma_u^2/n_t + 4\phi^2\sigma_\xi^2\sigma_u^2/n_{t-1} + 4\phi^2\sigma_u^4/(n_t n_{t-1}) = 2\sigma_u^4(\sigma_\xi^2/\sigma_u^2 + 1/n_t + \phi^2/n_{t-1})^2$. Next, $\mathrm{Cov}(\hat{\xi}_t^2, \hat{\xi}_{t-1}^2|\mathcal{N}_T) = 2\sigma_u^4\phi^2/n_{t-1}^2$, and $\mathrm{Cov}(\hat{\xi}_t^2, \hat{\xi}_{t-\tau}^2|\mathcal{N}_T) = 0$, $|\tau| \geq 2$. Therefore, we have

$$\mathrm{Var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\xi}_t^2|\mathcal{N}_T) = 2\sigma_u^4\frac{1}{T}\sum_{t=2}^{T}\left\{\left(\frac{\sigma_\xi^2}{\sigma_u^2} + \frac{1}{n_t} + \frac{\phi^2}{n_{t-1}}\right)^2 + \frac{2\phi^2}{n_{t-1}^2}\right\}$$

Furthermore, $\mathrm{Var}(\mathrm{E}[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\xi}_t^2|\mathcal{N}_T]) = \mathrm{Var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\sigma_\xi^2 + \sigma_u^2/n_t + \phi^2\sigma_u^2/n_{t-1})) = \sigma_u^4(1 + \phi^2)^2\mathrm{Var}(1/n_t)$. Hence,

$$\mathrm{Var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\xi}_t^2) = 2\sigma_u^4\frac{1}{T}\sum_{t=2}^{T}\mathrm{E}\left\{\left(\frac{\sigma_\xi^2}{\sigma_u^2} + \frac{1}{n_t} + \frac{\phi^2}{n_{t-1}}\right)^2 + \frac{2\phi^2}{n_{t-1}^2}\right\} + \sigma_u^4(1+\phi^2)^2\mathrm{Var}(1/n_t)$$

Finally, we have for the variance in the second term of (28):

$$\mathrm{Var}(\hat{\sigma}_u^2\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n_t}|\mathcal{N}_T) = 2\sigma_u^4\frac{\frac{1}{T}\sum_{t=1}^{T}\frac{n_t-1}{n_t^2}}{\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{n_t-1}{n_t}\right\}^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\right)^2$$

and $\mathrm{Var}(\mathrm{E}[\hat{\sigma}_u^2\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n_t}|\mathcal{N}_T]) = \sigma_u^4\mathrm{Var}(1/n_t)$. Putting the pieces of (28) together, we obtain the stated result for $\lim_{T\to\infty}\mathrm{Var}(\sqrt{T}\hat{\sigma}_\xi^2)$. Q.E.D.

Proof of Proposition 3.

Since ϕ is known but possibly equal to one, in which case $\{\beta_t\}$ would be non-stationary, classical theory on the estimation of state space models does not directly apply. As noted by Pagan (1980), however, the theory remains valid in the unit root case if the model is locally asymptotically identified in the sense of Kohn (1978). A necessary and sufficient condition for θ_0 to be locally asymptotically identified is that $\lim_{T\to\infty}(I(\theta_0)/T)^{-1}$ be non-singular. In our model we can directly check for this condition to hold. The sum in $I(\theta)$ contains two terms, the second of which, $2\mathrm{E}\left(\frac{\partial e_t'}{\partial\theta}\Sigma_t^{-1}\frac{\partial e_t}{\partial\theta'}\right)$ is positive semi-definite. It suffices to show that the first term, $\frac{\partial\mathrm{vec}(\Sigma_t)'}{\partial\theta}(\Sigma_t^{-1}\otimes\Sigma_t^{-1})\frac{\partial\mathrm{vec}(\Sigma_t)}{\partial\theta'}$, is positive definite. Similar to Proposition 1 of Pagan (1980), we have

$$\frac{\partial\mathrm{vec}(\Sigma_t)'}{\partial\theta}(\Sigma_t^{-1}\otimes\Sigma_t^{-1})\frac{\partial\mathrm{vec}(\Sigma_t)}{\partial\theta'} \geq \lambda_{min}^2(\Sigma_t^{-1})\Omega_t$$

where $\Omega_t = \frac{\partial\mathrm{vec}(\Sigma_t)'}{\partial\theta}\frac{\partial\mathrm{vec}(\Sigma_t)}{\partial\theta'}$, $\lambda_{min}(\cdot)$ denotes the smallest eigenvalue, and where $\geq$ means that the left hand side matrix minus the right hand side matrix is p.s.d. Due to the particular structure of our model, we have $\lambda_{min} = \sigma_u^2 > 0$, by assumption. Hence, it suffices to show that

Ω_t is p.d. Using the expressions in Appendix B, and denoting $X_t = \partial\sigma_t^2/\partial\sigma_u^2$ and $Y_t = \partial\sigma_t^2/\partial\sigma_\xi^2$, we can write

$$\Omega_t = n_t \begin{pmatrix} n_t X_t^2 + 2X_t + 1 & (Y_t+1)(n_t X_t + 1) \\ (Y_t+1)(n_t X_t + 1) & n_t (Y_t+1)^2 \end{pmatrix}$$

It follows immediately that $|\Omega_t| > 0$ if and only if $n_t > 1$, which happens with positive probability by assumption. Hence, $\lim_{T\to\infty} T^{-1}\sum_{t=1}^T \Omega_t$ is positive definite almost surely, which implies that $\lim_{T\to\infty}(I(\theta_0)/T)^{-1}$ is positive definite almost surely. This shows asymptotic local identifiability of the model.

The remaining conditions of Theorem 4 of Pagan (1980) hold trivially, as a_t is uniformly bounded from above and non-stochastic, the state space form is uniformly completely observable and uniformly completely controllable. Finally, for asymptotic normality we need that θ_0 is an interior point of Θ. Then, consistency and asymptotic normality follow by Theorem 4 of Pagan (1980). The form of the information matrix is standard for state space models, see e.g. Lütkepohl (1993, p.437).

Appendix B: Information matrix

To simplify the notation, we denote $\beta_t := \beta_{t|t}$ and $\sigma_t^2 := \sigma_\beta^2(t|t)$. The model in (5)-(7) can then be written as

$$\begin{aligned} e_t &= \eta_t - a_t\phi\beta_{t-1} \\ \beta_t &= \phi\beta_{t-1} + \zeta_{t-1}a_t'\Sigma_t^{-1}e_t \\ \Sigma_t &= a_t\zeta_{t-1}a_t' + \sigma_u^2 I_{n_t} \\ \sigma_t^2 &= \zeta_{t-1} - \zeta_{t-1}^2 a_t'\Sigma_t^{-1}a_t \end{aligned}$$

where $\zeta_t = \phi^2\sigma_t^2 + \sigma_\xi^2$. Let $\theta = (\sigma_u^2, \sigma_\xi^2)'$ and $\delta = (1,0)'$. Then,

$$\begin{aligned} \frac{\partial e_t}{\partial\theta} &= -a_t\phi\frac{\partial\beta_{t-1}}{\partial\theta} \\ \frac{\partial\beta_t}{\partial\theta} &= \phi\frac{\partial\beta_{t-1}}{\partial\theta} + \frac{\partial\zeta_{t-1}}{\partial\theta}a_t'\Sigma_t^{-1}e_t - \zeta_{t-1}\left\{\frac{\partial\mathrm{vec}(\Sigma_t)'}{\partial\theta}(\Sigma_t^{-1}e_t\otimes\Sigma_t^{-1}a_t) - a_t'\Sigma_t^{-1}\frac{\partial e_t}{\partial\theta}\right\} \\ \frac{\partial\sigma_t^2}{\partial\theta} &= \frac{\partial\zeta_{t-1}}{\partial\theta}\left(1 - 2\zeta_{t-1}a_t'\Sigma_t^{-1}a_t\right) + \zeta_{t-1}^2\frac{\partial\mathrm{vec}(\Sigma_t)'}{\partial\theta}(\Sigma_t^{-1}a_t\otimes\Sigma_t^{-1}a_t) \end{aligned}$$

$$\frac{\partial\mathrm{vec}(\Sigma_t)}{\partial\theta'} = (a_t\otimes a_t)\frac{\partial\zeta_{t-1}}{\partial\theta'} + \mathrm{vec}(I_{n_t})\delta'$$

Some expressions can be simplified by observing that, due to the particular structure of the model, $a_t'\Sigma_t^{-1}a_t = n_t/(n_t\zeta_{t-1} + \sigma_u^2)$. For example, this leads to

$$\frac{\partial\sigma_t^2}{\partial\theta} = \frac{\partial\zeta_{t-1}}{\partial\theta}\frac{\sigma_u^4}{(n_t\zeta_{t-1} + \sigma_u^2)^2} + a_t\Sigma_t^{-2}a_t\zeta_{t-1}^2\delta$$

Appendix C: Autometrics

Autometrics (Doornik, 2009) is an implementation of the general-to-specific modelling approach (see e.g. Hendry and Krolzig, 1999) based on several steps: First, a General Unrestricted Model (GUM) is defined. This set constitutes the base prior to model reduction. Second, each insignificant variable at a 5% confidence level offers a reduction path and each sub-path offers other reduction paths. If there are k insignificant variables in the GUM, then 2^k possible models need to be considered. For each sub-path, an F-test is made on the removed variables. At this step, variables meeting the test may be reintroduced in the model even if they are not significant in the final model. Several diagnostics are made, including a Chow test, in order to check the model's stability. If the tests fail, another reduction path is selected. As the amount of paths is considerable, the Autometrics algorithm uses rules to select which models are worth going through the selection procedure: it automatically discards sub-paths of failed deletion, eliminates bunches of variables instead of taking them off one-by-one, and, once a variable is deemed "highly insignificant", all paths containing this variable are discarded (at the risk of eliminating viable solutions).

Appendix D: Tables

Table 5: Regression autometrics I

	Estimate	Std. Error	t-statistic	$Pr(>\|t\|)$	
Intercept	17.28	1.31	13.22	2.00E-16	***
Artists					
Basquiat	1.63	0.09	18.31	2.00E-16	***
Braque	- 0.22	0.10	- 2.26	2.36E-02	*
Chagall	0.46	0.08	5.44	5.57E-08	***
Dufy	- 0.85	0.08	- 10.93	2.00E-16	***
Ernst	- 0.90	0.08	- 10.62	2.00E-16	***
Gauguin	1.13	0.15	7.72	1.27E-14	***
Hassam	1.28	0.13	9.66	2.00E-16	***
Judd	0.75	0.27	2.81	4.92E-03	**
Kandinsky	0.86	0.15	5.73	1.03E-08	***
Kooning	1.69	0.10	16.70	2.00E-16	***
Matisse	1.20	0.11	10.98	2.00E-16	***
Monet	1.22	0.09	14.19	2.00E-16	***
Moore	- 2.14	0.20	- 10.71	2.00E-16	***
Picasso	0.92	0.07	12.44	2.00E-16	***
Pissarro	0.55	0.09	6.31	2.96E-10	***
Prince	1.06	0.10	10.27	2.00E-16	***
Richter	- 1.11	0.07	- 15.00	2.00E-16	***
Rothko	2.82	0.14	20.00	2.00E-16	***
Sisley	0.64	0.10	6.12	9.65E-10	***
Vlaminck	- 1.31	0.06	- 20.65	2.00E-16	***
Vuillard	- 0.69	0.09	- 7.65	2.10E-14	***
Warhol	1.80	0.06	28.53	2.00E-16	***
ZaoWouKi	- 0.96	0.10	- 9.43	2.00E-16	***
Nationality of the artist					
U.S.A.	- 2.32	0.08	- 28.26	2.00E-16	***
Italy	- 0.43	0.09	- 4.53	6.01E-06	***
Great Britain	- 1.68	0.13	- 13.35	2.00E-16	***
The Netherlands	- 0.75	0.08	- 9.12	2.00E-16	***
France	- 0.59	0.07	- 8.50	2.00E-16	***
Japan	- 2.20	0.09	- 23.91	2.00E-16	***
Russia	- 0.42	0.09	- 4.91	9.46E-07	***

Table 6: Regression autometrics II

	Estimate	Std. Error	t-statistic	$Pr(>\|t\|)$	
Date of creation of the artwork					
WorkDate	- 0.00	0.00	- 4.99	6.11E-07	***
Medium used					
acrylic	0.55	0.06	9.30	2.00E-16	***
autograph	- 2.72	0.42	- 6.52	7.26E-11	***
Brush	- 0.82	0.14	- 5.74	9.79E-09	***
butterflies and enamel on canvas	1.33	0.15	9.04	2.00E-16	***
Casein	2.40	0.37	6.49	8.85E-11	***
color	- 2.04	0.20	- 10.07	2.00E-16	***
flies and resin on canvas	1.35	0.45	3.00	2.74E-03	**
gloss and household paint	1.19	0.19	6.22	5.30E-10	***
handcolored	- 3.28	0.44	- 7.40	1.42E-13	***
House Paint	0.78	0.14	5.51	3.70E-08	***
inkjet print and acrylic	2.11	0.34	6.28	3.46E-10	***
Lacquer	0.93	0.34	2.73	6.40E-03	**
latex paint	1.43	0.42	3.41	6.60E-04	***
litograph	- 2.52	0.54	- 4.64	3.60E-06	***
Oil	0.90	0.05	16.42	2.00E-16	***
Oilstick	- 0.80	0.21	- 3.89	1.03E-04	***
paint	0.44	0.14	3.23	1.23E-03	**
paste	0.79	0.28	2.85	4.36E-03	**
peinture la colle	1.23	0.36	3.46	5.41E-04	***
Silk	- 0.29	0.09	- 3.30	9.60E-04	***
spray	0.67	0.29	2.34	1.94E-02	*
synthentic polymer paint	0.61	0.08	7.43	1.14E-13	***
Tapestry	- 3.74	0.20	- 18.69	2.00E-16	***
Tempera	0.45	0.14	3.27	1.09E-03	**
water paint	0.81	0.10	7.89	3.35E-15	***
Size					
Height	0.01	0.00	24.27	2.00E-16	***
Width	0.01	0.00	25.95	2.00E-16	***

Table 7: Regression autometrics III: Auction house variables

	Estimate	Std. Error	t-statistic	$Pr(>\|t\|)$	
artcurial	0.43	0.11	3.90	9.77E-05	***
Blomqvist	- 3.30	0.59	- 5.56	2.79E-08	***
BloomsburyAuctionsNewYork	- 5.22	1.16	- 4.50	6.80E-06	***
BruunRasmussenBredgadeCopenhagen	1.93	0.52	3.68	2.35E-04	***
BukowskisStockholm	0.80	0.29	2.80	5.17E-03	**
ChristiesAmsterdam	1.86	0.59	3.15	1.67E-03	**
ChristiesDubai	1.08	0.37	2.94	3.33E-03	**
ChristiesEast	- 1.33	0.37	- 3.60	3.23E-04	***
ChristiesLondon	1.08	0.05	20.30	2.00E-16	***
ChristiesMilan	1.00	0.33	3.05	2.31E-03	**
ChristiesNewYork	0.49	0.06	8.37	2.00E-16	***
ChristiesParis	0.91	0.09	9.61	2.00E-16	***
ChristiesTaipei	- 1.51	0.46	- 3.28	1.03E-03	**
ClaudeAguttes	0.55	0.19	2.87	4.10E-03	**
Dorotheum	1.54	0.70	2.19	2.86E-02	*
DuMouchellesFineArtsAuctioneers	- 3.68	0.38	- 9.68	2.00E-16	***
FarsettiArte	1.41	0.33	4.30	1.76E-05	***
FinarteMilan	1.64	0.65	2.54	1.11E-02	*
GalerieKornfeld	0.78	0.11	7.26	4.28E-13	***
HauswedellandNolte	- 0.77	0.21	- 3.66	2.52E-04	***
imKinsky	1.92	0.73	2.64	8.43E-03	**
KAuction	1.10	0.25	4.36	1.33E-05	***
KettererKunstMunchen	1.14	0.21	5.39	7.13E-08	***
KollerAuktionenAG	0.57	0.20	2.84	4.55E-03	**
Lempertz	0.76	0.18	4.29	1.78E-05	***
MatsartAuctioneersandAppraisers	0.63	0.24	2.66	7.82E-03	**
MillonandCornettedeSaintCyr	- 1.16	0.27	- 4.38	1.22E-05	***
OsenatScp	1.60	0.48	3.31	9.29E-04	***
PandolfiniCasaDAste	1.88	0.86	2.18	2.91E-02	*
PhillipsdePuryandCompanyLondon	1.31	0.13	9.77	2.00E-16	***
PiasaArtcurial	0.72	0.34	2.11	3.50E-02	*
PolyInternationalAuctionCoLtd	1.67	0.37	4.47	7.80E-06	***
PorroandC	1.51	0.41	3.69	2.25E-04	***
SeoulAuction	1.64	0.20	8.15	4.11E-16	***
SothebysAmsterdam	1.42	0.62	2.30	2.14E-02	*
SothebysArcade	- 1.57	0.39	- 4.03	5.72E-05	***
SothebysLondon	1.09	0.05	20.92	2.00E-16	***
SothebysMilan	1.45	0.30	4.88	1.08E-06	***
SothebysNewYork	0.44	0.06	7.48	8.00E-14	***
SothebysParis	1.10	0.14	7.97	1.76E-15	***
VanHamKunstauktionen	0.78	0.25	3.15	1.66E-03	**
VillaGrisebachAuktionenGmbH	1.14	0.17	6.61	4.03E-11	***

Table 8: Regression autometrics IV

	Estimate	Std. Error	t-statistic	$Pr(>\|t\|)$	
fees charged by auction house					
HAMMER	- 0.16	0.05	- 3.65	2.65E-04	***
Country of sale					
Austria	- 1.37	0.67	- 2.04	4.18E-02	*
Denmark	- 2.22	0.49	- 4.58	4.72E-06	***
Germany	- 0.90	0.15	- 6.17	7.21E-10	***
Hong Kong	1.19	0.12	9.98	2.00E-16	***
Italy	- 0.72	0.27	- 2.66	7.75E-03	**
Japan	0.71	0.08	8.64	2.00E-16	***
Norway	1.92	0.46	4.21	2.52E-05	***
Taiwan	1.31	0.16	8.10	6.19E-16	***
The Netherlands	- 1.88	0.53	- 3.58	3.45E-04	***
U.S.A.	0.61	0.07	9.13	2.00E-16	***

Table 9: Regression autometrics: diagnostics

Number of selected variables	111
R^2	61%
Adjusted R^2	60%
Maximum variance inflation factor	16.10
Mean variance inflation factor	2.70
Proportion of p-values $< 1\%$	93%
Standard deviation of residuals	1.15
Skewness of residuals	- 0.23
Kurtosis of residuals	7.28
p-value Jarque Bera normality test	$<$ 2.2e-16

References

Arellano, M., 2003. *Panel Data Econometrics*. Oxford University Press.

Baumol, W., 1986. Unnatural Value: Or Art Investment as Floating Crap Game. American Economic Review, American Economic Association, vol. 76(2), pages 10-14.

Bocart, F., Oosterlinck, K., 2011. Discoveries of fakes: Their impact on the art market. *Economics Letters* 113(2), 99-102.

Bocart, F., Hafner, C., 2011. Econometric analysis of volatile art markets. *Computational Statistics and Data Analysis* doi:10.1016/j.csda.2011.10.019

Case, K., Shiller, R., 1987. Prices of single-family homes since 1970: new indexes for four cities.*New England Economic Review* Federal Reserve Bank of Boston, 45-56.

Chanel, O., Gerard-Varet, L., Ginsburgh, V., 1996. The relevance of hedonic price indices: the case of paintings. *Journal of Cultural Economics* 20, 1-24.

Collins, A., Scorcu, A.E., Zanola, R., 2009. Reconsidering hedonic art price indexes. *Economics Letters, Elsevier* 104(2), 57-60.

Combris, P., Lecocq S., Visser M., 1997. Estimation of a Hedonic Price Equation for Bordeaux Wine: Does Quality Matter? *The Economic Journal* 107(441) 390-402.

de la Barre M., Docclo S., Ginsburgh V., 1994. Returns of Impressionist, Modern and Contemporary European Paintings 1962-1991. *Annales d'economie et de statistique* 35, 143-181.

Doornik J., 2009. Autometrics. In Castle, J., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Dorsey R., Hu H., Mayer W., Wang H., 2010. Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle, *Journal of Housing Economics*. 19 7593

Duan, N., 1983. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association* 78, 605-610.

Engle, R.F., Rangel, G., 2008. The spline GARCH model for low frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21, 1187-1222.

Fan, J., Farmen, M., Gijbels, I., 1998. Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society, Series B*, 60, 591-608.

Fogarty, J., 2006. The return to Australian fine wine. *European Review of Agricultural Economics* 33(4) 542-561.

Frey, B., Eichenberger, R., 1995. On the Return of Art Investment Return Analyses. *Journal of Cultural Economics* 19, 207-220.

Ginsburgh, V., Mei, J., Moses, M., 2006. On The Computation of Price Indices, in: Ginsburgh, V., Throsby, D. (Eds), *Handbook of the economics of art and culture*, Elsevier, pp. 948-979.

Goetzmann, W., 1992. The Accuracy of Real Estate Indices: Repeat Sale Estimators. *Journal of Real Estate Finance and Economics* 5, 5-53.

Goetzmann, W., 1993. Accounting for Taste: Art and the Financial Markets Over Three Centuries. *American Economic Review* 83, 1370-1376.

Gouriéroux, C., Laferrere, A., 2009, Managing hedonic housing price indexes: The French experience. *Journal of Housing Economics* 18, 206-213.

Graddy, K., Margolis, P., 2011. Fiddling with Value: Violins as an Investment? *Economic Inquiry* 49(4), 1083-1097.

Greenleaf, E. A., Rao, A. G., Sinha, A. R., 1993. Guarantees in Auctions: The Auction House As Negotiator and Managerial Decision Maker. *Management Science* 39, 1130-1145.

Hansen, J., 2009. Australian House Prices: A Comparison of Hedonic and Repeat-Sales Measures *The Economic Record* 85(269), 132145

Hendry, D., Krolzig H.-M., 1999. Improving on 'Data mining reconsidered'by KD Hoover and SJ Perez. *The Econometrics Journal* 2(2), 202-219.

Hodgson, D., Vorkink, K., 2004. Asset Pricing Theory and the Valuation of Canadian Paintings. *The Canadian Journal of Economics* 37, 629-655.

Jones, A.M., Zanola, R., 2010. Retransformation bias in the adjacent price index. *POLIS Working Papers* n178.

Kohn, R., Local and global identification and strong consistency in time series models, *Journal of Econometrics* 8, 269-294.

Lütkepohl, H., 1993. Introduction to Multiple Time Series Analysis. Springer Verlag, Berlin.

McAndrew, C., Thompson, R., 2007. The collateral value of fine art. *Journal of Banking and Finance* 31, 589-607.

Mei, J., Moses, M., 2002. Art as an Investment and the Underperformance of Masterpieces. *The American Economic Review* 92, 1656-1668.

Pagan, A., 1980. Some identification and estimation results for regression models with stochastically varying coefficients, *Journal of Econometrics*, 13, 341-363.

Pesando, J., 1993. Art as an Investment: The Market for Modern Prints. *American Economic Review* 83, 1075-1089.

Renneboog, L., Spaenjers, C., 2010. Buying Beauty: On Prices and Returns in the Art Market. *Tilburg University, Center for Economic Research Discussion Paper* 2009-15.

Schulz, R. and Werwatz, A., 2004. A state space model for Berlin house prices: Estimation and economic interpretation. *Journal of Real Estate Finance and Economics*, 28, 37-57.

Scorcu, A.E., Zanola, R., 2010. The Right Price for Art Collectibles. A Quantile Hedonic Regression Investigation of Picasso Paintings. *Working Paper Series Rimini Centre for Economic Analysis* 01_10.

Seckin, A., Atukeren, E., 2006. Art and the Economy: A First Look at the Market for Paintings in Turkey. *Economics Bulletin* 26, 1-13.

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "HMM in dynamic HAC models" by Wolfgang Karl Härdle, Ostap Okhrin and Weining Wang, January 2012.
002 "Dynamic Activity Analysis Model Based Win-Win Development Forecasting Under the Environmental Regulation in China" by Shiyi Chen and Wolfgang Karl Härdle, January 2012.
003 "A Donsker Theorem for Lévy Measures" by Richard Nickl and Markus Reiß, January 2012.
004 "Computational Statistics (Journal)" by Wolfgang Karl Härdle, Yuichi Mori and Jürgen Symanzik, January 2012.
005 "Implementing quotas in university admissions: An experimental analysis" by Sebastian Braun, Nadja Dwenger, Dorothea Kübler and Alexander Westkamp, January 2012.
006 "Quantile Regression in Risk Calibration" by Shih-Kang Chao, Wolfgang Karl Härdle and Weining Wang, January 2012.
007 "Total Work and Gender: Facts and Possible Explanations" by Michael Burda, Daniel S. Hamermesh and Philippe Weil, February 2012.
008 "Does Basel II Pillar 3 Risk Exposure Data help to Identify Risky Banks?" by Ralf Sabiwalsky, February 2012.
009 "Comparability Effects of Mandatory IFRS Adoption" by Stefano Cascino and Joachim Gassen, February 2012.
010 "Fair Value Reclassifications of Financial Assets during the Financial Crisis" by Jannis Bischof, Ulf Brüggemann and Holger Daske, February 2012.
011 "Intended and unintended consequences of mandatory IFRS adoption: A review of extant evidence and suggestions for future research" by Ulf Brüggemann, Jörg-Markus Hitz and Thorsten Sellhorn, February 2012.
012 "Confidence sets in nonparametric calibration of exponential Lévy models" by Jakob Söhl, February 2012.
013 "The Polarization of Employment in German Local Labor Markets" by Charlotte Senftleben and Hanna Wielandt, February 2012.
014 "On the Dark Side of the Market: Identifying and Analyzing Hidden Order Placements" by Nikolaus Hautsch and Ruihong Huang, February 2012.
015 "Existence and Uniqueness of Perturbation Solutions to DSGE Models" by Hong Lan and Alexander Meyer-Gohde, February 2012.
016 "Nonparametric adaptive estimation of linear functionals for low frequency observed Lévy processes" by Johanna Kappus, February 2012.
017 "Option calibration of exponential Lévy models: Implementation and empirical results" by Jakob Söhl und Mathias Trabs, February 2012.
018 "Managerial Overconfidence and Corporate Risk Management" by Tim R. Adam, Chitru S. Fernando and Evgenia Golubeva, February 2012.
019 "Why Do Firms Engage in Selective Hedging?" by Tim R. Adam, Chitru S. Fernando and Jesus M. Salas, February 2012.
020 "A Slab in the Face: Building Quality and Neighborhood Effects" by Rainer Schulz and Martin Wersing, February 2012.
021 "A Strategy Perspective on the Performance Relevance of the CFO" by Andreas Venus and Andreas Engelen, February 2012.
022 "Assessing the Anchoring of Inflation Expectations" by Till Strohsal and Lars Winkelmann, February 2012.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Hidden Liquidity: Determinants and Impact" by Gökhan Cebiroglu and Ulrich Horst, March 2012.
024 "Bye Bye, G.I. - The Impact of the U.S. Military Drawdown on Local German Labor Markets" by Jan Peter aus dem Moore and Alexandra Spitz-Oener, March 2012.
025 "Is socially responsible investing just screening? Evidence from mutual funds" by Markus Hirschberger, Ralph E. Steuer, Sebastian Utz and Maximilian Wimmer, March 2012.
026 "Explaining regional unemployment differences in Germany: a spatial panel data analysis" by Franziska Lottmann, March 2012.
027 "Forecast based Pricing of Weather Derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Matthias Ritter, March 2012.
028 "Does umbrella branding really work? Investigating cross-category brand loyalty" by Nadja Silberhorn and Lutz Hildebrandt, April 2012.
029 "Statistical Modelling of Temperature Risk" by Zografia Anastasiadou, and Brenda López-Cabrera, April 2012.
030 "Support Vector Machines with Evolutionary Feature Selection for Default Prediction" by Wolfgang Karl Härdle, Dedy Dwi Prastyo and Christian Hafner, April 2012.
031 "Local Adaptive Multiplicative Error Models for High-Frequency Forecasts" by Wolfgang Karl Härdle, Nikolaus Hautsch and Andrija Mihoci, April 2012.
032 "Copula Dynamics in CDOs." by Barbara Choroś-Tomczyk, Wolfgang Karl Härdle and Ludger Overbeck, May 2012.
033 "Simultaneous Statistical Inference in Dynamic Factor Models" by Thorsten Dickhaus, May 2012.
034 "Realized Copula" by Matthias R. Fengler and Ostap Okhrin, Mai 2012.
035 "Correlated Trades and Herd Behavior in the Stock Market" by Simon Jurkatis, Stephanie Kremer and Dieter Nautz, May 2012
036 "Hierarchical Archimedean Copulae: The HAC Package" by Ostap Okhrin and Alexander Ristig, May 2012.
037 "Do Japanese Stock Prices Reflect Macro Fundamentals?" by Wenjuan Chen and Anton Velinov, May 2012.
038 "The Aging Investor: Insights from Neuroeconomics" by Peter N. C. Mohr and Hauke R. Heekeren, May 2012.
039 "Volatility of price indices for heterogeneous goods" by Fabian Y.R.P. Bocart and Christian M. Hafner, May 2012.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".