EXTENDED ABSTRACT

Title: A VaR-based model for air pollution control in big cities. The case of PM10 in the city of Madrid

Authors and e-mail of all: José-María Montero (jose.mlorenzo@uclm.es)[1] ; Gema Fernández-Avilés[1] (gema.faviles@uclm.es); Lidia Sanchis-Marco[2] (lidia.sanchis@uclm.es)

Department: (1) DECOADE; (2) Finance

University: University of Castilla-La Mancha, Spain

Subject area: Sostenibilidad, Medio Ambiente y Recursos Naturales

Abstract: *(minimum1500 words)*

Outdoor air pollution is one of the main problems affecting human health in urban areas all around the world. Therefore, it is no surprise that air pollution control is currently a major concern for citizens. Although emissions of most air pollutants have decreased substantially over the past decades, their concentrations still exceed the legal limits in most countries, indicating that air pollution control continues to be a challenge for modern societies. Every year, more than 4.2 million people suffer an early death because of outdoor air pollution (WHO, 2016). The main culprits are ozone (O3), nitrogen dioxide (NOx) and, most crucially, fine particles or particulate matter (PM) with a diameter of 10 micrometres or less (PM10).

In response to the adverse consequences of outdoor air pollution for humans and the environment, an extensive body of legislation has been developed, establishing health-based standards and objectives for the most harmful atmospheric pollutants. If set limits are exceeded, environmental authorities are required to adopt air quality plans detailing measures to keep the exceedance period as short as possible. Consequently, the prediction of exceedances of the legal standards has become a critical task for environmental authorities, especially in mega-cities, to prevent a wide range of adverse outcomes. Based on the existing air pollution monitoring systems, a number of methods have been put to work to provide a way of identifying when an environmental system is getting "out-of-control", and thus enable the application of appropriate remedial measures. Most of these attempts have focused on

mathematical models (either determinist or statistical, but usually statistical) aiming at one of the two following objectives: (i) The short-term prediction of the concentrations of a number of air pollutants especially harmful for human health (most of the literature); (ii) Predicting the violation of the legal limits established for such pollutants (here, the literature is much scarcer).

This paper is part of the current literature that focuses on the early warning of the violation of the legal limits established for the most harmful air pollutants, instead of on the short-term prediction of the value of their concentrations, and given the limitations above-mentioned of the deterministic approach, our methodological proposal is based on statistical methods.

We export the financial rudiments associated with risk measurement to the air pollution control arena. First, we use the financial concept of Value-at-Risk (VaR) to evaluate the risk of an environmental system (more specifically, an air pollution system) experiencing an extreme relative increase that, depending on the current state of such a system, could lead it to become "out-of-control" (to exceed the legal limits). Second, we do not use indirect methods to estimate the VaR, as usual in financial literature[1]. Instead, we use quantile regression (QR) for VaR estimation, which is a distribution-free direct method, and avoids the limitations of the indirect procedures. Specifically, we use the Conditional Autoregressive Value at Risk (CAViaR) approach by Engle and Manganelli (2004). Third, we extend CAViaR with certain observable endogenous variables related to meteorological conditions (although predictability is not necessarily limited to them) because they have been proved to have valuable predictive power to forecast the right tail of the conditional air pollution concentration distribution, which makes them useful for risk management in a pollution control framework (some recent examples are Ravindra et al., 2019; Feng et al., 2019; Załuska and Gładyszewska-Fiedoruk, 2020; Liu et al., 2020; Haddad and Vizakos, 2021).

In non-technical words, our proposal is to provide the citizens and environmental authorities of a territory with the probability, p, that the relative upside variation of the next-day concentrations of an air pollutant (in our case PM10) will equal or exceed a specific extreme value (this value is the p%-VaR value). More interestingly, citizens and environmental authorities can be provided with a set of p%-VaR values. This information is extremely useful for the environmental authorities to adopt the necessary measures to prevent damages for human health, fauna, and vegetation, and especially to alert of imminent exceedances of the legal limits set by environmental legislation.

For this purpose, we do not rely on the indirect methods usually employed in the VaR forecasting literature. Given the natural methodological link between VaR forecasting and the QR approach (VaR is a quantile), we propose a direct method, CAViaR, which considers the quantile from a conditional perspective: that is, the quantile is seen as a latent autoregressive process that may also depend on exogenous covariates and exhibit nonlinearities in parameters. Specifically, in the CAViaR approach the past behaviour of the VaR and the absolute value of the returns (in our case pollution "returns") is transmitted to the response variable via an ARMA-type function. We take advantage of the possibility of considering exogenous covariates in the CAViaR specification to include key meteorological variables that have been proved to be good predictors for air pollutant concentrations. Given the computational

[1] We are referring to the most popular parametric methods (like EWMA or GARCH-type models) nonparametric density estimation methods, and semiparametric approaches as Extreme Value Theory and Modelling based procedures, among others.

challenge of estimating a CAViaR model with several exogenous covariates, we have used a meteorological conditions index as the only exogenous covariate.

As far as we know, this the first time that an extended CAViaR strategy is used in the literature of air pollution control, and specifically as a warning mechanism of exceedances of the legal limits set for harmful pollutants concentrations (in our case PM10 concentrations) in the (very) short term.

We apply the methodology proposed in the city of Madrid, Spain. Specifically, we use it for one-day-ahead VaR forecasting at the 99%-quantile, that is, for the forecasting of the VaR value that there is a probability of 1% that it will be equalled or exceeded (or a probability of 99% that it will not be equalled or exceeded). For this purpose, we use the daily PM10 registers recorded by the air pollution monitoring stations of the city in the period January 2011-December 2019.

We use a SAV-CAVIaR, as this specification outperforms the others in VaR forecasting:

$$VaR_{\lambda,t} = \beta_{\lambda,0} + \beta_{\lambda,1} VaR_{\lambda,t-1} + \beta_{\lambda,2} \left| r_{t-1} \right| + \beta_{\lambda}^{*} \ f\left(X_{t-1} \right) + e_{\lambda,t}, \tag{7}$$

with X_t being a predictive time t variable, other than returns and $f(\cdot)$ a function of X_t-variables.

In the SAV-CAVIaR model the lagged returns are considered in absolute value. The reason behind this decision is that the VaR is expected to increase as r_{t-1} becomes very negative, because one bad day makes the probability of the next somewhat greater. However, it might be that very good days also increase VaR, as would be the case for volatility models. Hence, VaR could depend symmetrically on $\left| r_{t-1} \right|$.

The functional form of model (7) attempts to parsimoniously exploit the additional information conveyed by both the past of the conditional quantile and X_t. In other words, the main purpose of the autoregressive structure is to ensure that the dynamics of the conditional quantile change smoothly over time. Since VaR dynamics are highly persistent, the lag of the VaR process could also be seen as an instrumental variable that proxies the true latent process. Similarly, $\left| r_{t-1} \right|$ is a natural proxy for the unobservable volatility of returns. Since it introduces a source of (stochastic) short-term variation related to the arrival of news in the pollution market, this process is expected to be a major driver in any market risk measure. At this point, the similarities between the basic structure of the CAViaR model under the restriction and the class of GARCH models widely used to characterize volatility are fully evident. The existing literature in the finance framework has not yet discussed which variables should be included in such an analysis.[2] The central strategy we adopt consists of individually analysing the principal components of a number of key meteorological variables which are accepted as being related to the PM10 concentration level. In this article, framed in the area of air pollution control, we use as additional variables typical meteorological variables that have been proved to be highly related to the level of PM10 concentrations; their dynamics have been captured by a principal components indicator of such meteorological conditions (see more details in data section). It is of note that, although the results in a parametric modelling might be sensitive to the choice of the proxy selected to capture the dynamics of meteorological conditions, a robust picture is expected to emerge for a wide range of proxy variables.

The reason behind MCI is that the non-linear estimation of the SAV-CAViaR model extended with more than one X-variables is a computational challenge, quite time demanding and highly unstable for extreme quantiles, which are the ones we are interested in. In addition, the high correlation among meteorological variables could lead to multicollinearity problems.

[2]There is scarce literature on this matter. Rubia and Sanchis-Marco (2013) analyze the stock market forecasting ability of liquidity and trading activity variables using different SAV-CAViaR models extended with one microstructure covariate selected from the most relevant ones.

The model resulting from extending the basic SAV-CAViaR specification with the lags of a single predictor can be seen as a low-order individual autoregressive distributed lag model for the conditional quantile. We name this model the extended SAV-CAViaR model. Table 1 shows the backtesting results for the extended Caviar model and four competing alternatives for the monitoring stations operating in the city of Madrid. These competing alternatives include the traditional CAViaR specification as well as the EWMA model (RiskMetrics), the Gaussian GARCH(1,1) model, and a model that combines GARCH estimation with the block-maxima approach of the EVT&M.

Table 1

Backtesting results for PM10 returns: Extended CAViaR model vs. competing models

Cuatro Caminos							
VaR Model	Λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.4%	0.9257(0.3450)	0.2760(0.5993)	1.2017(0.5506)	1.4680(0.9616)	31.0420(0.0000)
GARCH	1%	2.1%	2.3530(0.1250)	0.3922(0.5311)	2.7452(0.2377)	2.2807(0.8922)	30.6789(0.0000)
EVT&M-BM	1%	1.4%	0.3530(0.4904)	0.0722(0.6711)	0.4252(0.6094)	1.8887(0.9322)	18.0499(0.0000)
CAVIaR	1%	1.3%	0.2169(0.6414)	0.0647(0.7993)	0.2816(0.8653)	1.0989(0.9815)	10.6789(0.0048)
Extended CAVIaR	1%	0.9%	0.2031(0.6615)	0.0363(0.8489)	0.2394(0.8704)	1.1569(0.9890)	4.1702(0.1243)
Escuelas Aguirre							
VaR Model	Λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.6%	0.6632(0.3315)	0.3082(0.5788)	0.9714(0.6153)	2.7371(0.8411)	29.0430(0.0000)
GARCH	1%	2.5%	0.7213(0.4305)	0.2642(0.6218)	0.9855(0.6535)	2.3608(0.8837)	31.5429(0.0000)
EVT&M-BM	1%	1.4%	0.2510(0.5924)	0.1722(0.5751)	0.4232(0.7991)	2.8987(0.9010)	16.4221(0.0008)
CAVIaR	1%	1.3%	0.2596(0.6414)	0.0447(0.6993)	0.3043(0.8504)	2.2125(0.8992)	11.4073(0.0033)
Extended CAVIaR	1%	0.9%	0.2169(0.6414)	0.0647(0.7993)	0.2816(0.8704)	0.3476(0.9992)	14.8308(0.0860)
Mendez Álvaro							
VaR Model	Λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.8%	0.7187(0.3966)	0.3278(0.5670)	0.9260(0.6294)	2.4446(0.8631)	26.0461(0.0000)
GARCH	1%	2.4%	0.1899(0.6630)	0.3841(0.5575)	0.5740(0.8437)	2.7788(0.8526)	24.6488(0.0000)
EVT&-BM	1%	1.8%	0.3528(0.5712)	0.1277(0.6115)	0.4805(0.6511)	1.8997(0.9190)	19.0529(0.0000)
CAVIaR	1%	1.3%	1.5383(0.2149)	0.2607(0.6096)	1.8112(0.4043)	1.4219(0.9243)	12.7340(0.0017)
Extended CAVIaR	1%	0.8%	0.0000(1.0000)	0.1012(0.7504)	0.1012(0.9506)	1.9002(0.9286)	1.6631(0.4354)
Paseo de la Castellana							
VaR Model	Λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	1.9%	0.9890(0.3092)	0.2760(0.5993)	1.2650(0.4806)	0.6555(0.9954)	11.9220(0.0024)
GARCH	1%	1.8%	2.6126(0.1060)	0.4321(0.5109)	3.0447(0.1077)	12.6878(0.0483)	11.0413(0.0092)
EVT&M-BM	1%	1.6%	0.3550(0.5224)	0.1789(0.5915)	0.5339(0.8001)	2.9897(0.8620)	10.6619(0.0085)
CAVIaR	1%	1.2%	1.3948(0.3289)	0.1012(0.7504)	1.4960(0.3906)	2.2153(0.8989)	1.9795(0.3717)
Extended CAVIaR	1%	0.9%	0.7187(0.3966)	0.1992(0.6553)	0.9260(0.7694)	2.2001(0.9059)	1.2847(0.5261)
Plaza Castilla							
VaR Model	Λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.1%	0.9431(0.3315)	0.3082(0.5788)	1.2513(0.5153)	1.1875(0.9775)	16.398(0.0000)
GARCH	1%	1.9%	0.8814(0.2114)	0.2275(0.5983)	1.1089(0.6704)	0.9131(0.9887)	10.987(0.0012)
EVT&M-BM	1%	1.7%	0.4550(0.4914)	0.1252(0.6514)	0.5802(0.8311)	1.9897(0.8990)	9.0929(0.0010)
CAVIaR	1%	1.3%	0.3890(0.5015)	0.0363(0.8489)	0.9714(0.7153)	1.2185(0.9760)	8.0982(0.0174)
Extended CAVIaR	1%	0.8%	0.1530(0.6540)	0.0161(0.8990)	0.1998(0.9682)	0.9087(0.9893)	4.6034(0.1001)
Moratalaz							

VaR Model	Λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.5%	0.9349(0.3248)	0.276(0.5993)	1.2012(0.5506)	1.8739(0.9309)	20.2441(0.0000)
GARCH	1%	2.3%	1.3490(0.1983)	0.376(0.4993)	1.7250(0.7406)	1.4968(0.9428)	21.0426 0.0000)
EVT&M-BM	1%	2.1%	0.4102(0.5264)	0.2989(0.5404)	0.7091(0.8111)	1.9897(0.8990)	17.9929(0.0000)
CAVIaR	1%	1.3%	1.5383(0.2149)	0.2607(0.6096)	1.7990(0.7243)	2.9355(0.8169)	13.5398(0.0011)
Extended CAVIaR	1%	0.7%	0.2169(0.6414)	0.0647(0.7993)	0.2816(0.9590)	1.6468(0.9492)	6.8176(0.0445)

Cont. Table 1

Backtesting results for PM10 returns: Extended CAViaR model vs. competing models.

Vallecas							
VaR Model	λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	1.6%	0.9239(0.3340)	0.276(0.5993)	1.1999(0.5606)	5.1803(0.5209)	10.0461(0.0099)
GARCH	1%	1.5%	0.2169(0.6414)	0.2775(0.5983)	0.4944(0.6704)	1.308(0.9145)	9.0491(0.0121)
EVT&M-BM	1%	1.4%	0.3901(0.5378)	0.2799(0.5624)	0.6700(0.7111)	1.9697(0.9091)	8.0529(0.0208)
CAVIaR	1%	1.2%	0.1899(0.6630)	0.1461(0.7023)	0.6060(0.7427)	2.2065(0.8998)	1.4028(0.4959)
Extended CAVIaR	1%	0.9%	0.1499(0.6830)	0.1061(0.7323)	0.2560(0.9627)	1.8548(0.9326)	1.3330(0.5134)
Sanchinarro							
VaR Model	λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.3%	0.8169(0.2414)	0.2775(0.5983)	1.0944(0.6004)	1.1751(0.9781)	30.0707(0.0000)
GARCH	1%	2.4%	0.7187(0.3966)	0.3278(0.5670)	1.0465(0.6294)	13.0398(0.0424)	29.0739(0.0000)
EVT&M-BM	1%	1.9%	0.4511(0.5018)	0.3010(0.5324)	0.7521(0.6511)	2.0997(0.8651)	27.9529(0.0000)
CAVIaR	1%	1.6%	0.5383(0.5149)	0.2607(0.6096)	0.7990(0.6043)	1.9797(0.8867)	20.6081(0.0000)
Extended CAVIaR	1%	0.7%	0.4187(0.5996)	0.1992(0.6553)	0.6179(0.7800)	1.0931(0.9815)	7.0280(0.0298)
Tres Olivos							
VaR Model	λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.4%	0.2169(0.6514)	0.2675(0.6043)	0.4844(0.5704)	1.3051(0.9581)	32.0707(0.0000)
GARCH	1%	2.2%	0.7249(0.3890)	0.2760(0.5993)	1.0009(0.7006)	1.4599(0.9083)	25.0739(0.0000)
EVT&M-BM	1%	1.7%	0.4411(0.5198)	0.2910(0.5414)	0.7321(0.6711)	2.0187(0.8821)	23.9529(0.0000)
CAVIaR	1%	1.3%	1.5383(0.2149)	0.2607(0.6096)	1.7990(0.4043)	2.1016(0.8797)	13.0988(0.0014)
Extended CAVIaR	1%	0.7%	0.7187(0.3966)	0.1992(0.6553)	0.9179(0.7229)	1.2339(0.9695)	12.1087(0.0596)
Urb. Emb. Barajas							
VaR Model	λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.9%	0.7187(0.3966)	0.3278(0.5670)	1.0465(0.6294)	8.8440(0.1825)	98.2092(0.0000)
GARCH	1%	2.8%	0.2169(0.6414)	0.2169(0.5983)	0.4338(0.6470)	0.4501(0.9884)	97.2020(0.0000)
EVT&M-BM	1%	1.7%	0.5411(0.4218)	0.3710(0.5012)	0.9121(0.7311)	2.9027(0.8712)	95.9393(0.0000)
CAVIaR	1%	1.4%	0.7187(0.3966)	0.1992(0.6553)	0.9179(0.7229)	2.0755(0.8894)	89.0205(0.0000)
Extended CAVIaR	1%	0.8%	0.1894(0.6540)	0.4091(0.5224)	0.5985(0.8390)	0.4153(0.9887)	6.0305(0.0490)
Casa de Campo							
VaR Model	λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.8%	0.9431(0.3315)	0.3082(0.5788)	1.2513(0.6153)	0.8606(0.9903)	26.7134(0.0000)
GARCH	1%	2.7%	0.2169(0.6514)	0.2775(0.5983)	0.4944(0.6704)	1.0894(0.9820)	25.6403(0.0000)
EVT&M-BM	1%	2.3%	0.5611(0.4008)	0.3910(0.4891)	0.9521(0.7010)	2.9927(0.8612)	24.6910(0.0000)
CAVIaR	1%	1.6%	2.6126(0.1060)	0.3307(0.5653)	2.9433(0.2277)	3.1681(0.7256)	19.0881(0.0000)
Extended CAVIaR	1%	0.8%	0.1891(0.6699)	0.2661(0.6013)	0.4551(0.8793)	1.0430(0.9857)	6.2392(0.0442)
Farolillo							
VaR Model	λ	Exc (%)	LRUC	LRIND	LRCC	DQ	VQR
EWMA	1%	2.6%	1.8112(0.4043)	0.3742(0.5407)	2.1854(0.2843)	12.1087(0.0596)	23.0304(0.0000)
GARCH	1%	2.5%	0.3849(0.6479)	0.2460(0.5293)	0.6309(0.7606)	2.2867(0.8915)	22.1087(0.0000)
EVT&M-BM	1%	1.8%	0.5911(0.37808)	0.4310(0.4592)	1.0221(0.5710)	3.0127(0.8212)	18.0910(0.0000)
CAVIaR	1%	1.4%	3.9136(0.0479)	1.7512(0.1857)	5.6648(0.0583)	3.7087(0.7025)	11.6144(0.0030)

Extended CAVIaR	1%	0.8%	0.3671(0.6589)	0.2024(0.5510)	0.5695(0.8500)	2.1061 (0.9194)	6.5789(0.0373)

*The column Exc. shows the estimated frequency of empirical exceptions, while the columns LRUC, LRIND, LRCC, DQ and VQR show the values of the test statistics for the respective tests (*p*-values in parentheses).

The out-of-sample (as well the in-sample) results for the estimation and forecasting of the one-day-ahead VaR at the 99%-quantile, in the sites where the twelve PM10 monitoring stations are operating in the city of Madrid, indicate that the extension of the standard CAViaR model with the abovementioned meteorological index notably improve the accuracy of the forecasts. In addition, our extended CAViaR proposal clearly outperforms traditional CAViaR and other competing popular parametric and semi-parametric forecasting alternatives not only in finance but also in a large number of disciplines (including air pollution). Importantly, the heavier the traffic in the urban area, the more the extended CAViaR strategy outperforms the alternatives.

Therefore, we contribute a high-performance strategy to the literature on air pollution control—a strategy that has yielded very good results in finance and that has been exported to the air pollution control arena, extended with a meteorological index which captures the impact of the meteorological conditions, which in turn have been proved to be relevant drivers of air pollutant concentrations. This is an important finding of undoubted relevance for municipal forecasting teams, municipal authorities, and policy makers in charge of environmental issues.

References:

Engle, R. and Manganelli, S. 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. J. Bus. Econ. Stat. 22(4), 367-381.

Feng, H., Zou, B., Wang, J., Gu, X. 2019. Dominant variables of global air pollution-climate interaction: Geographic insight. Ecol. Indic. 99, 251-260.

Haddad, K., Vizakos, N. 2021. Air quality pollutants and their relationship with meteorological variables in four suburbs of Greater Sydney, Australia. Air Qual. Atmos. Health 14, 55–67.

Liu, Y., Zhou, Y., Lu, J. 2020. Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. Sci. Rep. 10, 14518.

Ravindra, K., Rattan, P., Mor, S., Aggarwal, A.N. 2019. Generalized additive models: Building evidence of air pollution, climate change and human health. Environ. Int. 132, 104987.

Rubia, A., Sanchis-Marco, L. 2013. On Downside Risk Predictability through Liquidity and Trading Activity: A Dynamic Quantile Approach. International Journal of Forecasting, 29, 202-219

WHO. 2016. https://www.who.int/health-topics/air-pollution#tab=tab_2. Accessed March 16, 2021.

Załuska, M., Gładyszewska-Fiedoruk, K. 2020. Regression Model of PM2.5 Concentration in a Single-Family House. Sustainability 12, 5952; doi: 10.3390/su12155952.

Keywords: *Air pollution control; PM10; Meteorological conditions; Value-at-Risk; CAViaR; Non-linear quantile regression.*

JEL codes: Q53, G11