Heterogeneous impact of soil contamination on farmland prices in the Belgian Campine region : evidence from unconditional quantile regressions

Reference:
Peeters Ludo, Schreurs Eloi, Van Passel Steven.- Heterogeneous impact of soil contamination on farmland prices in the Belgian Campine region : evidence from unconditional quantile regressions
Environmental and resource economics - ISSN 0924-6460 - 66:1(2017), p. 135-168
Full text (Publishers DOI): http://dx.doi.org/doi:10.1007/S10640-015-9945-6

Title

Heterogeneous Impact of Soil Contamination on Farmland Prices in the Belgian Campine Region: Evidence from Unconditional Quantile Regressions

This paper has not been submitted elsewhere in identical or similar form, nor will it be during the first three months after its submission to the Publisher.

Abstract

We estimate a hedonic-pricing model using geo-coded farmland-transaction data from the Campine region, situated in the north-east of Belgium. Unlike previous hedonic studies, we use the method of unconditional quantile regression (Firpo, Fortin, and Lemieux 2009). An important advantage of this new method over the traditional conditional quantile regression (Koenker and Bassett 1978) is that it allows for the estimation of potentially heterogeneous effects of cadmium pollution along the entire (unconditional) distribution of farmland prices. Using a threshold specification of the hedonic-pricing model, we find evidence of a U-shaped valuation pattern, where cadmium pollution of the soil has a negative and significant impact on prices only in the middle range of the distribution, insofar as cadmium concentrations are above the regulatory standard of 2 parts per million for agricultural land. Results obtained from a probit model to classify land plots into different price segments further suggest that the heterogeneous impact of soil pollution on price can be directly related to the variety of amenities that farmland provides.

Key words

Hedonic analysis
Cadmium pollution
Farmland prices
Unconditional quantile regression
Non-uniform valuation

1. Introduction

Soil pollution in the form of heavy-metal contamination is one of several major pollution problems in today's world, causing severe threats to humans and the environment. In many countries, especially those experiencing rapid industrialization such as, for example, China (Hu et al. 2014), the problem of heavy-metal polluted soils is mounting to alarming proportions. Despite the growing concern about this environmental hazard, remarkably little attention has been devoted in the environmental-economics literature to the impact of soil pollution on local land markets and, more precisely, on local farmland prices and associated welfare implications.

The lack of hedonic studies dealing with soil pollution is striking. The purpose of the present paper is to start filling this gap. Specifically, we examine whether heavy-metal pollution plays a significant role in determining farmland prices in the Campine region, situated in the north-east of Belgium. This region is widely known for its burdensome legacy of airborne cadmium (Cd) pollution due to zinc-smelting/-alloying activities of the local non-ferrous metal industry in the past (e.g., Van Meirvenne and Goovaerts 2001). We specify a hedonic land-pricing model to estimate the implicit price of Cd pollution, which is a useful indicator of the value that people attach to this potentially hazardous land attribute. The estimation of our model uses data from more than 500 geo-coded sales transactions in a specific area of the Campine region, concluded during the period 2004–2011.

Estimation of the implicit price of Cd pollution is challenging, especially if one is willing to allow for a *non-uniform* valuation of land characteristics, which may arise due to the variety of amenities that farmland provides; e.g., agricultural production, rural living, open space, and net asset value to land developers (Ma and Swinton 2012). Purchasers of farmland may value the Cd contamination of their newly acquired land differently, for various reasons. If such is the case, conventional mean regression using ordinary least squares (OLS), which has long been a standard tool of hedonic valuation, is not appropriate due to its singular focus on the effect of Cd pollution on the conditional *mean* of the price distribution, at the expense of exploring the segmentation that may exist in the local land market. To overcome this obvious limitation, we use the new method of *unconditional* quantile regression (UQR), recently introduced by Firpo, Fortin, and Lemieux (2009; FFL hereafter). To the best of our knowledge, the present hedonic study is the very first one using UQR, so we consider this a key contribution of this paper.

The use of UQR opens up new avenues for research. In the present case, UQR allows us to address the issue of potentially heterogeneous effects of Cd pollution on the prices of farmland. That is, UQR enables us to assess the impact of Cd pollution at different quantiles of the *unconditional* (or marginal) distribution of the farmland prices—i.e., the price distribution for the whole (sample) population of farmland sales transactions concluded in the local area. Using a threshold specification of the hedonic-pricing model, our application of UQR provides evidence of a U-shaped valuation pattern, where Cd pollution tends to have a negative and significant impact on farmland prices only in the middle range of the price distribution—insofar as Cd levels are above the regulatory standard for agricultural land prevailing in the region. Moreover, results from an examination of the information conveyed by the distribution of farmland prices suggest that the varying impact of Cd contamination basically arises from the self-selection of purchasers into specific price segments for reasons related to the variety of goods and services provided by farmland.

The rest of the paper is organized as follows. Section 2 provides relevant background information which may help the reader in appreciating the empirical results presented later in the paper. Section 3 sets out the estimation methods used, with a focus on the UQR method. Section 4 describes the study area and presents the baseline specification of the hedonic-pricing model, along with an overview of the data and the variables included in the empirical model.

Section 5 starts with a discussion of the main results, followed by a robustness check of the empirical findings. Section 6 summarizes the results and formulates some concluding remarks.

2. Background

2.1 Measures of Soil Pollution

The standard approach in the hedonic literature is to take several measures of distance or adjacency to the pollution source (toxic site) as a proxy for environmental quality. Despite its widespread use, it remains unclear whether distance to the pollution source is a valid measure of what participants in the land market are actually aware of and/or care about (e.g., Guignet 2013).

For the present study, we were in a fortunate position to have had access to objective (science-based) measures of Cd pollution. More specifically, we used the information available from previous research on heavy-metal polluted soil in the Campine region, which was based on a large number of field measurements of Cd pollution (Schreurs et al. 2011). Using GIS techniques, we then assigned spatially-interpolated levels of Cd concentration to each geo-coded farmland plot comprised in our data set. We recognize that difficulties may arise by using objective rather than subjective pollution measures, given that heavy-metal pollution of the soil is not visible, does not emit odor, and does not have any other characteristic that easily identifies its presence (e.g., Boyle and Kiel 2001). But support comes from previous hedonic research in which it was found that objective and subjective measures of pollution tend to converge in validity (Baranzini et al. 2010), and in some instances objective measures would even outperform subjective ones (Poor et al. 2001).

From another perspective, most buyers of farmland in the present case are local farmers and other individuals who are fairly knowledgeable about the problem of heavy-metal pollution in the study area. The public's awareness of the soil pollution stems from numerous media reports released in the past on the alarming results of a series of clinical studies conducted in the area (e.g., Hogervorst et al. 2007; Nawrot et al. 2008), along with several large-scale action plans and information campaigns launched by the Flemish government, including the Cadmium Action Plan and the (cross-border) BeNeKempen project.[1] Moreover, if buyers of farmland bid down prices for those plots whose actual (objective) Cd contamination is higher, then, intuitively, it must be that these buyers are aware of the spatial distribution of Cd-polluted soils and can easily localize the plots of land that are heavily polluted and those that are not. Unfortunately, we are not able to test these conjectures as we have no direct (survey-based) evidence at our disposal of what participants in the local land market know in reality.

Finally, the use of objectively measured levels of Cd contamination is also consistent with policy considerations, and therefore enables us to directly involve environmental regulations in the hedonic-pricing analysis. To target public-health concerns, a threshold level for the Cd pollution of agricultural land is set at 2 parts per million (ppm) by the Public Waste Agency of Flanders (OVAM), which is the regulatory institute that is responsible for waste management and soil remediation in Flanders (Van Meirvenne and Meklit 2010).[2] In the empirical analysis that follows, we show that this threshold value plays a critical role in the valuation of Cd contamination of farmland. An important implication of this environmental regulation is that farmers buying heavily-polluted land (i.e., land carrying Cd levels above the alert value of 2 ppm) face agricultural land-use restrictions. That is, the newly acquired land is not the best and is much less versatile, as it cannot be used for crop/vegetable production. So it can be expected that this limitation translates into lower farmland prices (see also, e.g., Lizin et al. 2015).[3] In the case of non-agricultural land uses, environmental regulations are less stringent. For example, the threshold values set by OVAM for residential and recreational land uses are 6 ppm and 9.5 ppm, respectively. These values are far outside the empirical range of our data sample.

2.2 Exurbanization and Distance Measures

Exurban development, sometimes designated as "rural sprawl" (Ma and Swinton 2012), is prevalent is the Campine region and elsewhere in Belgium, particularly in Flanders. As an outcome of this pattern of rural-urban development, developed space is always nearby because of the scattered housing (i.e., houses spread out over an extended area) and the archetypical "ribbon development" (i.e., houses along both sides of the main provincial roads). The consequence of these tight rural-urban interfaces is that every reasonably accessible patch of land is a potential building site, where all farmland potentially qualifies for being put to a different future use other than agricultural production.

Due to the fact that the landscape of the Campine region is predominantly characterized by low-density residential development and a high level of spatial fragmentation of the land in conjunction with significant land-use diversity, the development potential of a given farmland plot cannot possibly be associated in a meaningful way with its mere distance to the nearest urban (local town) center. Therefore, the use of conventional distance-to-city-center measures as a proxy for the urban-development potential of farmland (Plantinga et al. 2002; Livanis et al. 2006; Guiling et al. 2009) is likely to fail in the present case, as is confirmed later in the paper.

Besides, three other reasons can easily be advanced for *not* using distance measures. First, the use of distance to the nearest town center is considered inappropriate because such an approach imposes a uniform morphology on the area. As a simple example, take two land plots located at equal distance from the nearest town center. When using a distance-to-city-center measure, the two plots are viewed as being subject to the same extent of urban-development pressure, regardless of the topographical and zoning or land-use differences that may exist in the direct vicinity of the two plots. Second, since our data set covers sales transactions concluded over a period of several years, the area surrounding a given farmland plot observed in, say, 2005 is likely to be different from the one in, say, 2011, while the distance to the nearest town center obviously remains the same. Third, the distance to the nearest central business district (CBD), which is a widely-used measure in hedonic-pricing studies dealing with the impact of "urban sprawl", is unlikely to exert a perceptible influence on farmland prices in our case, since the nearest CBD, in the city of Antwerp, is located at a distance of 70 km, on average, which amounts to about an hour's drive by car.[4]

Finally, given the relatively low residential mobility typically observed in Belgium, it is felt realistic to assume that urban-development pressure on agricultural land primarily comes from the expansion of local towns to accommodate increasing housing needs (due to local demand for rural living) and commercial and other development needs (due to local demand for amenities associated with urban life). Accordingly, we use a more natural way to control for urban-development pressure, namely by incorporating the "address density" within a small radius around each land plot.

2.3 Exurbanization, Land-Use Conversions, and Speculative Purchases

Exurbanization with rural housing is also an increasingly important driver of changes in the demand for agricultural land and potential land-use conversions. Farmland markets are therefore increasingly subject to competitive pressures due to the expansion needs of local towns. Furthermore, agricultural zoning and land-use regulations in Belgium (Flanders) are not always strictly enforced. Finally, prices of developable land in our study area are, on average, 35 times higher than those of agricultural land, thus potentially giving rise to purchases for speculative purposes by non-agricultural investors.

As a result, farmland prices may far exceed the capitalization of agricultural rents (Capozza and Helsley 1989; Deaton and Vyn 2010), such that any attempt to conduct a sensible hedonic analysis of farmland prices calls for the inclusion of appropriate controls describing non-agricultural amenities of farmland. If purchasers of farmland have reasons other than agricultural production for buying the land, Cd pollution is likely to be traded-off for other characteristics of the land deemed attractive or convenient for the intended purposes (Cho et al. 2011; Cohen et al. 2015), which may be different from agricultural production (e.g., residential or recreational purposes). Therefore, it is possible that non-agricultural amenities can be found to be more important determinants of farmland prices (see also Uematsu et al. 2013, for the case of farmland; Snyder et al. 2007, for the case of forestland).

2.4 Arm's Length Transactions and Grey Market

An important characteristic of sales transactions in Belgium requires some more clarification. We refer to the fact that most transactions in the local land market in the Campine region (in our data sample about 80% of all transactions) are realized through *private treaty* rather than through *public auction*. Private sales are common in Belgium, but these transactions are not necessarily—and mostly not—of the non-arm's length type involving lower prices for relatives and friends (Huang et al. 2006).

Excessively high transaction costs related to the purchase of agricultural land (e.g., registration tax, notary fee, etc.) have led to the emergence of a lively grey market, which provides strong incentives to buyers of land to pay part of the price to the seller "in an envelope", and to refrain from paying any taxes on this side payment. It was estimated in a previous study that, on average, the envelope payment (which is approximately equal to the mean difference between auction-sale prices and private-sale prices) is about 20% of the full (or actual) purchase price (Ciaian et al. 2012, p. 8).

3. Unconditional Quantile Regression Method

To estimate the impact of soil pollution on the prices of farmland, we perform a hedonic-valuation analysis. A stylized form of the hedonic-pricing model is as follows:

$$\log Price = \alpha + \beta(Pollution) + \gamma(Covariates) + \varepsilon, \tag{1}$$

where $\log Price$ is the natural log of the price per square meter (m^2) realized for farmland, *Pollution* is some measure of Cd contamination, *Covariates* comprises a set of controls, and ε is the error term.

To estimate the hedonic-pricing model in Eq. (1), we use three different methods: (i) ordinary least squares (OLS), (ii) conditional quantile regression (CQR), and (iii) unconditional quantile regression (UQR). Many researchers have found quantile-regression methods useful because they provide richer evidence than OLS by allowing for heterogeneous effects and estimating the distributional impacts of the explanatory variables. However, standard CQR provides estimates of the relationship between the covariates and the conditional rather than the unconditional distribution of the outcome variable. Our focus is, therefore, on the new method of UQR, recently introduced by FFL (2009), while OLS and CQR are used mainly for the sake of comparison with the results obtained using UQR.

3.1 From OLS over CQR to UQR

OLS is a widely-used tool for conducting hedonic-pricing studies. The traditional linear regression shows how the conditional *mean* of the dependent variable Y (farmland prices, in our case) responds to a change in an explanatory variable X (a farmland characteristic, in our case), other things being equal. An appealing feature of OLS, through the law of iterated expectations, is that it provides consistent estimates of the effect of X on the *unconditional* population mean outcome of Y. However, since OLS regression focuses only on the mean, it is not well suited to explain the changes in the *overall distribution* of the outcome variable.

To overcome this limitation, researchers typically resort to the use of the CQR method, developed by Koenker and Bassett (1978), to examine how implicit prices vary across the *conditional* distribution of property prices.[5] This is not a trivial case, as property characteristics may be valued differently at different points of the conditional price distribution. However, despite the fact that the use of CQR has become widespread in applied economics by now, its implications are not always fully understood. Accordingly, misleading interpretations of the CQR results easily arise. Basically, the problem with CQR stems from the fact that it is not able to provide consistent estimate of the *population* quantiles, since the law of iterated expectations does not apply. Specifically, the CQR coefficient on a covariate X represents its effect on the τth quantile of the *conditional* distribution of Y--or the “within-group” effect conditional on, say, the mean values of all other covariates included in the model. The problem is even worsened by the fact that this asymmetry implies that the interpretation of the CQR coefficients changes whenever a different set of covariates is included in the hedonic-pricing model.

Given the shortcomings of both OLS and CQR, this study uses the new method of UQR, recently introduced by FFL (2009).[6] What is particularly interesting about UQR is that it also obeys the law of iterated expectation, so that it allows us to *directly* assess the impact of an explanatory variable X (level of Cd pollution) on the τth quantile of the *unconditional* distribution, $F_Y(\cdot)$, of the outcome variable Y (farmland prices); that is, $dF_Y(y)/dX$, all else held constant and irrespective of the set of covariates included. Control variables are conditioned on for identification but *without* altering the interpretation of the estimates. This is an undeniable advantage of UQR over CQR, as the estimated effects on Y of incremental changes in X can be *directly* interpreted; i.e., adding control variables or changing the set of control variables does *not* alter the interpretation of the parameter of interest.[7] Clearly, this property of UQR (which it shares with OLS) is critically important when it comes to understanding the impact of the level of soil pollution in, say, the lower or the upper tail of the unconditional distribution of farmland prices. As a result of that, one should also be cautious when comparing the results obtained using CQR and UQR, since they estimate distinct objects of interest that are not directly comparable. Thus, apart from differences that occur in the *magnitude* of the coefficients estimated using CQR and UQR, the *interpretation* of the coefficients is entirely different as well (Maclean et al. 2014, p. 203).

In sum, UQR allows us to study the distributional impact of an explanatory variable X on the outcome of interest Y at the *population* level (Borah et al. 2013, p. 25) and thus to go beyond the *within-group* effects. UQR allows us to answer questions such as the following: “Suppose that every member (land plot) of the population would experience the same exogenous change in one of its characteristics (say, a 0.1-ppm or 1% reduction of Cd concentration), how would this affect the median price or the price at any quantile of the overall price distribution, holding everything else constant?”. To illustrate the differences between UQR and CQR, and the potential pitfalls in interpreting their results, a simple numerical example with a simulation is provided in Appendix A of the paper.

3.2 RIF-OLS Regression

To implement the UQR method, we make use of the *influence function* (IF), which is widely used in the robust-estimation literature. The IF for a distributional statistic (or functional), such as a quantile of the distribution of Y, refers to the influence of a single observation on that particular quantile.

From FFL (2009) we learn that the IF of the τth unconditional quantile is defined as

$$\mathrm{IF}(Y; q_\tau) = \frac{\tau - I(Y \leq q_\tau)}{f_Y(q_\tau)}, \tag{2}$$

where $I(\cdot)$ is an indicator function taking the value one if $Y \leq q_\tau$ and zero otherwise, and $f_Y(q_\tau)$ is the probability density function of Y evaluated at q_τ. The *re-centered influence function* (RIF) adds back the τth quantile, such that

$$\mathrm{RIF}(Y; q_\tau) = q_\tau + \mathrm{IF}(Y; q_\tau), \tag{3}$$

where the *mean* of $\mathrm{RIF}(Y; q_\tau)$ is equal to the actual τ-quantile. That is,

$$E[\mathrm{RIF}(Y; q_\tau)] = q_\tau. \tag{4}$$

FFL (2009) derive the *RIF regression* model by specifying the conditional expectation of the $\mathrm{RIF}(Y; q_\tau)$ in Eq. (4) as a function of the explanatory variables X, given by

$$E[\mathrm{RIF}(Y; q_\tau)|X] = X'\beta_\tau, \tag{5}$$

which is viewed as the unconditional quantile regression (UQR). Because $E_X E[\mathrm{RIF}(Y; q_\tau)|X] = q_\tau$, by virtue of the definition of the RIF, $E_X[\mathrm{d}q_\tau/\mathrm{d}X]$ is the marginal effect of a small change (or location shift) in the distribution of a *continuous* covariate X on the τth unconditional quantile of the outcome variable Y, holding everything else constant. FFL (2009) called this effect the *unconditional quantile partial effect* on the τth quantile, $\mathrm{UQPE}(\tau)$, which is given by

$$\beta_\tau \equiv \mathrm{UQPE}(\tau) = \frac{1}{f_Y(q_\tau)} E\left[\frac{\mathrm{d}\Pr(Y > q_\tau|X)}{\mathrm{d}X}\right], \tag{6}$$

where $E[\mathrm{d}\Pr(Y > q_\tau|X)/\mathrm{d}X]$ is the *average* derivative of the probability that Y is greater than q_τ given the covariates X. Conversely, the estimate of the $\mathrm{UQPE}(\tau)$ for a *binary* covariate X can be obtained as

$$\beta_\tau \equiv \mathrm{UQPE}(\tau) = E[\Pr(Y > q_\tau|X = 1)] - E[\Pr(Y > q_\tau|X = 0)], \tag{7}$$

where the $\mathrm{UQPE}(\tau)$ is to be interpreted as the impact on the τth quantile of an incremental (say, a 1-percentage point) change in the probability $\Pr(X = 1)$.

The UQR method involves a *two-step* procedure which is easy to implement. For a specific quantile τ, the first step is to estimate the RIF of the τth quantile of Y in accordance with Eqs. (2)–(3). In practice, a feasible RIF can be obtained as $\widehat{\mathrm{RIF}}(Y; \hat{q}_\tau) = \hat{q}_\tau + [\tau - I(Y \leq \hat{q}_\tau)]/\hat{f}_Y(\hat{q}_\tau)$, where $\hat{q}_\tau$ is the τth sample quantile of the unconditional distribution of Y, and $\hat{f}_Y(\cdot)$ is estimated using a non-parametric kernel estimator. The second step of the procedure is to run an OLS regression, the so-called *RIF-OLS regression*, of the $\widehat{\mathrm{RIF}}(Y; \hat{q}_\tau)$ on the covariates X, where it is assumed that the quantiles of Y are a linear function of the covariates X.

4. Data and Variables

The data used for the present hedonic study were made available by the Belgian Land Registry Office. This unique dataset contains 599 sales records of farmland plots in our study area covering 14 municipalities in the Campine region, shown in Fig. 1. The (unrepeated) sales transactions were concluded during the period April 2004 to December 2011. All sales records contain detailed cadastral information, including sales price, date and type of transaction, size of farmland plot, presence of built structures on the property, and the type of land use prior to the sales transaction. Even though all properties are labeled "agricultural land" in the land register, several plots in our sample (almost 20%) were not located in an agricultural-zoning area at the time of the transaction.[8]

Insert **Fig. 1** about here

The information available from the land register allowed us to geo-code all farmland plots (using ArcGIS software, release 10.0).[9] The geo-coded information thus obtained was then used to assign scientifically measured Cd-pollution levels to each farmland plot by using spatial-interpolation techniques (kriging). It further allowed us to construct various spatial variables that can be entered into the hedonic-pricing model, including, among others, the address (or housing) density within a 1-km buffer around each land plot, distance of each land plot to the nearest urban (local town) center, and distance to the Dutch border. Finally, the GIS information allowed us to construct spatiotemporal lags of the realized sales prices (discussed further in Section 4.4).

The dependent variable of our hedonic-pricing model measures the prices of farmland, expressed in euros/m^2. We converted the nominal sales prices into real prices, expressed in constant 2011 euros (based on the monthly Consumer Price Index). It is important to note that the dependent price variable does not always represents the full transaction price due to the existence of a grey market (see Section 2.4). The covariates and other controls entered in the hedonic-pricing model include: (i) structural property characteristics, along with the level of Cd pollution; (ii) neighborhood characteristics; (iii) municipality and year dummies, and (iv) spatiotemporal lag of the dependent price variable. Table 1 lists all the variables and provides basic descriptive statistics.

Insert **Table 1** about here

We now proceed with a description of all the variables included in the hedonic-pricing model. We begin with an account of the environmental-hazard variable, namely the objectively determined level of Cd pollution assigned to each farmland plot, and the form in which this measure is incorporated in the empirical model. Next, we provide an overview of the set of covariates (other land-plot characteristics) and some additional controls.

4.1 Environmental Variable: Cd Pollution

Our measure of cadmium (Cd) pollution of the farmland is derived from a *trend surface* which was estimated, through spatial interpolation, on the basis of about 12,000 actual field measurements in the study area (Schreurs et al. 2011). GIS information was used to assign "the best estimate" of the level of Cd pollution to each geo-coded farmland plot in our sample. The objective pollution values assigned to each plot are best regarded as proxies for the buyers' and sellers' perceptions of the pollution. Said differently, prevailing perceptions of pollution are largely consistent with the "actual" levels of Cd contamination. A contour map of the soil pollution in the study area is shown in Fig. 2.

Insert **Fig. 2** about here

To model the relationship between prices and pollution levels, we considered several ways of specifying the Cd-pollution term, but eventually settled for a *threshold* specification which enforces the effect of Cd to be zero at levels below a critical threshold value, θ, and non-zero and monotonically increasing (in absolute size) at levels above the cutoff value. Accordingly, the effect of Cd pollution is switched on, so to speak, once the threshold is passed. This brings us to the following form of the RIF-OLS regression model for the τ-th quantile (using notation in a "loose" way, for ease of presentation):

$$E[\mathrm{RIF}(\log Price_i\,; q_\tau)] = \alpha_\tau + \beta_\tau \left\{ I(Cd_i > \theta) \times \left[\log\left(\frac{Cd_i}{\theta}\right)\right]^2 \right\} + \gamma_\tau(Covariates), \quad (8)$$

where $Price_i$ is the price/m^2 of farmland, β_τ measures the τth quantile effect of Cd pollution, $I(\cdot)$ is an indicator function which is equal to 1 if $Cd_i > \theta$, and 0 otherwise, and $\tau \in [0, 1]$. The threshold value θ for agricultural land is set at 2 ppm by OVAM.

The specification choice in Eq. (8), involving a two-part transformation of the Cd-contamination measure, is motivated by three considerations. First, the threshold-indicator design (the first part) implies two "regimes" associated with distinct valuations of Cd pollution. The underlying idea is as follows. At low levels of Cd pollution (Cd $\leq$ 2 ppm), purchasers may, on average, not worry about soil pollution. Since Cd levels are below the alert value, there is nothing to really worry about as the land is declared safe, and, therefore, farmland prices remain essentially unaffected by small variations in contamination levels. At high levels of Cd pollution (Cd $>$ 2 ppm), however, buyers may begin to worry about soil pollution and consider the land potentially hazardous, so that prices become more sensitive to increasing Cd levels (see also, e.g., Johnson and Chess 2003; Alberini 2007; Guignet 2012). Moreover, the threshold specification is consistent with the observation that objective (science-based) measures of pollution are convergent in validity with the public's subjective perceptions of pollution (Poor et al. 2001), especially for moderate to high pollution levels (Baranzini et al. 2010).

Second, the quadratic form of the proportional-deviation measure (the second part) is intended to rule out the occurrence of a sharp discontinuity or "break" at the threshold value of 2 ppm, where the impact of Cd pollution is assumed to become effective. So the specification allows for a timid start beyond the Cd level of 2 ppm. Evidently, such a "smooth-transition" type of model is more plausible, since it is highly improbable that buyers are fully knowledgeable about the threshold point (most likely they have only fuzzy knowledge about it) and would massively bid down prices immediately beyond the level of 2 ppm (see also, e.g., Chay and Greenstone 2005, p. 392).

A third reason for adopting the baseline specification in Eq. (8) is that it gives rise to a "natural" concave-convex (or reverse-logistic type) price response at Cd levels greater than 2 ppm, as will be illustrated in Section 5.2 below, where it can be shown that the inflection point of the price-response curve occurs where

$$2\beta_\tau \left[\log\left(\frac{Cd_i}{\theta}\right)\right]^2 - \log\left(\frac{Cd_i}{\theta}\right) + 1 = 0; \quad (9)$$

that is, at

$$Cd_i = 2\exp\left(\frac{1-\sqrt{1-8\beta_\tau}}{4\beta_\tau}\right), \quad \text{for } Cd_i > \theta.$$

From Eq. (8), the implied elasticity of the farmland price with respect to the level of Cd pollution at the τth quantile, η_τ, can be computed as

$$\eta_\tau \equiv \frac{\Delta_\tau \log Price}{\Delta \log Cd} = I(Cd_{\mathrm{i}} > \theta) \times 2\beta_\tau \log\left(\frac{Cd_i}{\theta}\right). \tag{10}$$

This expression indicates that η_τ is a *non-constant* elasticity at any given τ-quantile, which increases (in absolute value) in a non-linear fashion with a growing proportionate deviation of the actual Cd level from the regulatory standard of $\theta = 2$ ppm.[10]

4.2 Covariates: Other Farmland Characteristics

This section provides a brief overview of the covariates (other farmland characteristics) and additional controls incorporated in the hedonic-pricing model.[11] First, a binary variable (*public*) is included to indicate whether the sales transaction was concluded through public auction (*public* = 1) or private treaty (*public* = 0). This indicator variable is intended to control for the fact that the dependent variable does not always represent the full price of the farmland plot but only the price *net* of the (hidden) side payment in the case of private-sale transactions (see Section 2.4). The size of the land plot is also considered an important determinant of the price (see, e.g., Cavailhés and Wavresky 2003). A small number of farmland plots have buildings (farmhouses, barns, or other buildings) attached to them. So we included a binary variable to indicate whether a land plot comprises built structures (*structure* = 1) or is vacant (*structure* = 0).[12]

Next, we included additional binary variables to indicate (i) the farmland's use prior to the sales transaction (*pasture*), and (ii) the type of zoning of the area in which the farmland plot is located (either *residential* or *natureforest*). In the first case, the included category is pasture land (*pasture* = 1), covering permanent and temporal grassland as well as meadowland, while the omitted (reference) category has been designated as "arable and other land" (*pasture* = 0). The reason for the latter is that in several instances our data set does not allow us to identify the specific types of usage of the land. In the second case, the included categories are residential zoning (*residential* = 1) and nature/forest zoning (*natureforest* = 1), while the omitted category is agricultural zoning (*residential* = *natureforest* = 0).

To capture the future development potential of farmland plots, we further include the local *address density* as a proxy for the pressure exerted by increasing urbanization on agricultural land in our study area. This variable is defined as the number of addresses (expressed in thousands) found within a one-km radius around each farmland plot.[13] We prefer the use of the address-density variable over the use of a distance-based measure, for the reasons discussed above, in Section 2.2.

Finally, we include a number of spatial variables in the hedonic-pricing model. First, we enter the distance of each farmland plot to the Dutch border to capture the existence of a "border effect" caused by the fact that many Dutch people want to buy land just across the border in Belgium, given that real-estate prices are lower in Belgium. Second, we include a spatiotemporal lag of the dependent price variable, the specification of which will be described below.

4.3 Municipality and Year Dummies

To control for unobserved local-area and time effects, we also incorporate municipality and year dummies in the hedonic-pricing model, consistent with common practice in hedonic studies. First, municipality dummies have been included to capture slowly-changing (if not entirely time-invariant) attributes of the municipality in which a farmland plot happens to be located, such as accessibility (access to express highways, public transportation services, etc.), and different municipal planning, land-use (conversion vs. conservation), and zoning policies. Many of these attributes are likely to be closely related to the "rural character" or "agricultural connection" of a municipality; that is, whether the local land market in a municipality is still, say, predominantly rural, semi-rural, or rural-urban.

Second, year dummies have been included to represent various common factors such as changing interest rates, changes in agricultural support measures (e.g., subsidies), adjustments to environmental and land regulations, and so on. However, unreported analyses suggested that the year dummies are principally reflective of the time-series variation in the prices of developable land, averaged over the municipalities in our study area.[14]

4.4 Spatiotemporal Lag of Sales Prices

Lastly, we include spatiotemporal lags of local farmland prices, following (with some modifications) Patton and McErlean (2003) and Maddison (2009). This additional variable is defined as a spatially weighted average of realized farmland prices in the recent past within some predefined neighborhood (distance range), and is intended to account for the fact that agricultural land prices are not solely determined by the characteristics of the land but tend to reflect also average local farmland prices. It is should be emphasized, though, that we do not use a symmetric weights matrix (as is typically used in standard spatial-lag models), because a symmetric matrix would be over-connected as information about future prices cannot possibly travel backwards in time. Thus, all observations have been *pooled* while taking into account the *temporal* dimension of the data (see also Dubé and Legros 2014).

In constructing the spatiotemporal lag of farmland prices, two specific issues arise: (i) how to define the relevant "neighborhood"; and (ii) how to determine the "recent past". To address the first issue, we conducted a preliminary covariance analysis, as explained in Appendix B of the paper. Based on this analysis, we decided to account only for the within-municipality spatiotemporal price effects, while ignoring covariances among prices of farmland plots located in different municipalities. As regards the second issue, we have chosen a 12-month time window.[15]

The specifics of the construction of this variable are as follows. The spatiotemporal weights matrix $\mathbf{W}$ is obtained from the Hadamard (element-by-element) multiplication of the spatial weights matrix $\mathbf{S}$ and the temporal weights matrix $\mathbf{T}$; that is, $\mathbf{W} = \mathbf{S} \odot \mathbf{T} = \{s_{ij} \cdot t_{ij}\} = \{w_{ij}\}$, where $\odot$ is the Hadamard-multiplication operator. Our measure of proximity is based on the distances between the farmland plots in our sample. Specifically, the matrix $\mathbf{S}$ contains *inverse* distances, $s_{ij} = 1/d_{ij}$ for land plots $i \neq j$, with diagonal elements, s_{ii}, set to zero (note that the use of inverse distances implies a 50% decay of the spatial weight at a distance of 2 km). If sales are ordered by their time of occurrence (from the most recent sales transaction to the first one in the observation period), the matrix $\mathbf{T}$, with elements t_{ij}, is specified as the Hadamard product of three matrices: (i) an upper-triangular matrix of ones; (ii) a matrix containing blocks of ones to identify farmland plots sold on an earlier date within the 12-month time window (where the ones are replaced by zeroes for those transactions that took place on the very same day), and (iii) a matrix of interspersed ones to identify within-municipality sales. The resulting hybrid weights matrix $\mathbf{W}$ was row-standardized for the purpose of computing the spatiotemporal lags of the within-municipality farmland prices.[16]

5. Estimation Results

This section presents the estimation results. A summary of the results is provided in Table 2. Since the dependent variable is measured in logs, we look at proportionate changes (or percentage changes in the case of elasticities) in the real price per square meter of farmland.

Insert **Table 2** about here

In the discussion that follows, our focus is on the results obtained using UQR. In contrasting the UQR results with those returned by OLS and CQR, it should be remembered that the estimates are not directly comparable. The estimated coefficients on the other covariates will be discussed only briefly.

5.1 Impact of Cd Pollution

The OLS estimate of the coefficient on the Cd-pollution term, reported in column 1 of Table 2, predicts a downward (-0.378) but statistically insignificant ($p = 0.512$) change in the mean of the distribution of farmland prices. This initial finding also masks the heterogeneity of the effects across the quantiles of the distribution of the farmland prices. The results obtained using CQR (columns 2–4 of Table 2) and UQR (columns 5–7 of Table 2) do reveal such heterogeneity, but the results from the two quantile-regression estimators display strongly diverging patterns of the quantile effects.

A more detailed picture is provided by the quantile plots shown in Fig. 3, where the estimated coefficients on the Cd term are plotted against the quantiles of the price distribution. Looking at the CQR results, it can be seen that the estimated coefficient on the Cd-pollution term decreases from a value slightly above zero at the 10th percentile (0.343) to a negative value at the 30th percentile (-1.011) and then remains fairly stable all the way through to the 90th percentile (-0.928) of the *conditional* price distribution, while actually being statistically indistinguishable from zero.

Insert **Fig. 3** about here

In contrast, the UQR estimates of the Cd coefficient appear to display a U-shaped pattern of the estimated pollution coefficients along the quantiles of the *unconditional* distribution of farmland prices. Specifically, the UQR effect of Cd pollution is negative and significant *only* for the (logged) prices in the middle range of the distribution, yielding a coefficient at the median of -1.381 (which is significant at the 1% level). This result suggests that the unconditional price distribution becomes more (less) positively skewed with higher (lower) levels of Cd pollution, holding everything else constant. In contrast, the effects of Cd pollution at the lower and upper tails of the price distribution, while being quite important in size, are imprecisely estimated to such an extent that they are not significantly different from zero. By uncovering the non-uniform price response to changes in the level of Cd pollution, the advantages from using UQR become utterly clear, while a singular focus on the mean of the price distribution using conventional OLS would not have revealed the “dip” in the quantile plot at the median. Also, using CQR does not lead to conclusions different from the one that can be made on the basis of the OLS estimate.[17]

Although a U-shaped pattern of the quantile effects is discernable from Fig. 3, we note that the null hypothesis of pair-wise equality of the coefficients at the 90th percentile and the median of the price distribution could not be rejected by a Wald-type *F* test (see at bottom of Table 2).[18] However, it should be noted that the Wald-type *F* test is a rather "conservative" test, such that it is not uncommon to find insignificant coefficient heterogeneity in hedonic-pricing studies using quantile regression (see also, e.g., Gamper-Rabindran and Timmins 2013).[19] Conversely, the pair-wise equality of the estimated coefficients at the 10th percentile and the median is strongly rejected at the 1% level of significance.

While we find a negative and significant effect of Cd pollution on farmland prices at the median, consistent with a-priori expectations, the insignificant results obtained at the "extremes" of the price distribution call for a closer inspection. To improve our understanding of why farmland plots in the lower and upper 20% tails of the distribution are virtually unaffected by variations in the level of Cd pollution, we conducted a more detailed investigation of the bundles of land characteristics found in different price segments. We regressed an indicator for being a farmland plot in one of the tails or the middle range of the price distribution using a probit model (see Maclean et al. 2014, p. 202, for a similar idea, applied in a completely different setting).[20] The probit estimates, summarized in Table 3, suggest that several plot characteristics are strikingly different depending on whether the land plot belongs to the lower tail, the middle range, or the upper tail of the price distribution. Importantly, the results of the probit analysis may provide some priors about why Cd pollution affects farmland prices differently in distinct segments of the distribution. The following findings stand out. Land plots in the lower 20% tail tend to be used as pasture land or to be located in nature/forest-zoning areas, where buyers care less about the level of Cd pollution (perhaps some buyers may simply use the acquired land plot for the purpose of manure disposal). Land plots in the 60% middle range are significantly larger in size, so these plots are likely to be more suitable for continued farming, and buyers care more about soil pollution. Finally, land plots in the upper 20% tail are smaller in size, surrounded by a larger number of residential dwellings (higher address density), and tend to be located in residential-zoning areas. These land plots are also more likely to be sold through public auction, and tend to be located much closer to the national border. Our conjecture, therefore, is that these farmland plots were primarily sold to Dutch purchasers and used mainly for residential purposes.

The above results suggest that buyers who sort themselves into the lower and higher ends of the price distribution tend to place less or no value on soil pollution. Although the estimates at the lower extreme of the price distribution are statistically indistinguishable from zero, their counter-intuitive signs and large magnitudes remain somewhat disturbing. The unexpected results could be indicative of a positive correlation that exists between Cd levels and some unobservable characteristics that make the land plots involved particularly attractive, hence rendering pollution an endogenous attribute (where a naïve interpretation would erroneously conclude that buyers derive utility from pollution). One possible explanation, in the spirit of Cotteleer et al. (2008), could be that a government agency (for instance, the Agency for Nature and Forest in Flanders) buys farmland for nature and landscape preservation, or that individuals buy the land for the natural-amenity services it provides. Thus, prices at the lower end of the distribution might just have been driven up by strong demand shocks related to the farmland's non-agricultural amenities (e.g., open space and recreational opportunities).[21]

Insert **Table 3** about here

We calculated the implied *elasticities* of farmland prices with respect to Cd pollution, as defined in Eq. (10), evaluated at selected levels of Cd contamination. Our findings suggest that the elasticities at the median of the price distribution range from −0.095, at Cd = 2.07 ppm, to −2.082, at Cd = 4.25 ppm, and are significantly different from zero (at the 1% level). The elasticities evaluated at Cd levels of 3 ppm and 4 ppm are −1.120 (with 95% confidence interval (−0.313, −1.926)) and −1.914 (with 95% CI (−0.536, −3.293)), respectively.[22]

To inform our intuition about the implications of the estimated coefficients for the price *level*, we simulated farmland prices for Cd levels above the threshold value of 2 ppm by using Eq. (8) in its exponentiated form, $\exp(X_i'\beta_\tau)$. Evaluated at the means of the continuous variables, and assuming a zero value for all the binary indicators, we find that a discrete change in the distribution of the soil pollution such that the average level of Cd concentration increases from 2 ppm to 3 ppm results in a price depreciation of 20.3% (with 95% CI (6.1%, 32.4%)), for $\hat{\beta}_{0.50} = -1.381$. If Cd levels change to an average level of 4 ppm, as compared to 2 ppm, the implied price reduction would be 48.5% (with 95% CI (16.8%, 68.1%)). If we apply these percent changes to the *median* price in the estimation sample (2.37 euro/m^2), the above shifts in Cd level would lead to price discounts of 0.48 euro/m^2 (with 95% CI (0.14, 0.77 euro/m^2)) and 1.15 euro/m^2 (with 95% CI (0.40, 1.61 euro/m^2)), respectively, if farmland prices are uniformly converted into constant 2011 euros. We further note that these findings are broadly in line with estimates reported in Howland (2000), Jackson (2002), and Alberini (2007).

5.2 Relationships with Other Farmland Characteristics

While our primary interest lies with the impact of soil pollution on farmland prices, several reasonable associations between farmland prices and other characteristics of the land plots seem to emerge from the estimations as well. The quantile plots, showing the estimated coefficients on those characteristics returned by OLS, CQR, and UQR, are presented in Fig. 4.

Insert **Fig. 4** about here

We begin by looking at the coefficients on the farmland characteristics entered as binary variables in the hedonic-pricing model. (Remember that the UQR estimate of the coefficient on a binary variable X represents the effect of a 1 percentage-point change in the share of "successes", or $\Delta X = 0.01$.) To the extent that a causal interpretation of the estimates is valid, our results suggest that 1 percentage-point growth in the share of public sales leads to a 0.57% and 0.83% increase of the 80th and 95th percentiles of the price distribution, respectively (the rebound at the 95th percentile after the lower value of 0.21% at the 90th percentile has is not shown in panel a of Fig. 4). By contrast, the quantiles in the lower 30% tail of the price distribution hardly change.[23] The overall picture of a (quasi) monotonically increasing effect of the share of public sales across the quantiles of the price distribution is consistent with our finding that this type of sales is predominantly associated with the higher segments of the price distribution (see column 3 of Table 3).

The presence of built structures on a land plot has no statistically significant impact on farmland prices (panel b). This insignificant result could be due to the fact that in many instances the presence of a built structure is a liability when it must be removed before the sales transaction and potential development of the land can take place (e.g., Howland 2000). Pasture land sells at prices significantly lower than arable and other land, where the UQR estimates at all quantiles of the price distribution are close to the OLS estimate, thus suggesting a pure location shift (panel c). Moreover, as expected, land plots located in a residential-zoning area sell at significantly higher prices than those located in agricultural-zoning areas, though this finding holds for plots in the upper price range only (panel d). Conversely, land plots located in nature- or forest-zoning areas (*natureforest*) sell at much lower prices than those in agricultural-zoning areas (panel e).

We now turn to the coefficients on the farmland characteristics measured as continuous variables. (Remember that the UQR estimate of the coefficient on a continuous variable represents the effect of a location shift of the entire distribution of this variable.) First, the relationship between the price and the size of a land plot is found to be negative and significant only in the upper end of the price distribution (panel f of Fig. 4). This could be indicative of the fact that a larger size in the high-price segment of the land market may not be in agreement with the intended uses that buyers had in mind at the time of the purchase. The positive value of smaller land plots, occasionally called "plattage value", is known to be prevalent in (ex)urbanized areas, where farmland is likely to have more development potential. This phenomenon typically arises from the fact that the costs of subdividing larger land plots into smaller ones can be avoided (e.g., Chicoine 1981; Colwell and Munneke 1997; Ecker and Isakson 2005). Second, the address density within a 1-km buffer around each plot of land has a sizeable positive effect on its price, particularly for those plots in the higher price segments of the market (panel g). Third, land plots that are closer to the Dutch border fetch much higher prices, as expected (panel h). Finally, the coefficient on the spatiotemporal lag of the sales price is significant only at the 60th percentile of the price distribution; the overall picture, though, appears to be one of a pure location shift of the price distribution due to spatiotemporal price spillovers, which is teasing the borderline of statistical significance at the 10% level (panel i).

Finally, the municipality dummies are generally found to be significant along the quantiles of the price distribution, while the year dummies are found to be significant only at the median of the price distribution. To the extent that the year dummies are reflective of the time-series variation in the average prices of developable land in the study area (see footnote 14), this finding suggests that only the median of the farmland prices is particularly sensitive to "spillovers" from the prices prevailing in the local market of developable land. Evidently, more work is needed to understand the mechanisms that are at work.

5.3 Comparison of Alternative Cd-Pollution Terms

As already mentioned above, there are several ways to specify the Cd-pollution term in the hedonic-pricing model. Therefore, we examine the sensitivity of the estimated Cd-pollution impact on prices of farmland obtained from five alternative specifications, including our baseline specification.

We first look at two other *threshold* specifications used to ensure that the Cd-contamination level becomes effective only beyond the regulatory ceiling of 2 ppm, namely: (i) a threshold specification with a *linear* (rather than quadratic) form of the deviation measure, $\log(\mathrm{Cd}/2)$; and (ii) a simple binary indicator, which is equal to one if Cd > 2 ppm, and zero otherwise. Next, we consider two *continuous* specifications, namely: (i) a quadratic form in the (untransformed) level of Cd pollution (Cd and Cd^2); and (ii) a quadratic form in the logged levels of Cd pollution ($\log \mathrm{Cd}$ and $(\log \mathrm{Cd})^2$). The estimated coefficients and implied elasticities obtained from using the various specifications, for the median of the price distribution, are presented in Table 4 (where, for the sake of comparison, column 1 presents the outcomes for our baseline specification). The implied elasticities (if any) have been evaluated at selected levels of Cd contamination.

Insert **Table 4** about here

The threshold specification with a linear form of the proportionate-deviation measure (column 2 of Table 4) returns a *constant* elasticity of -0.560 (significant at the 5% level). The specification with the binary indicator (column 3) gives rise to a sudden (downward) "break" at Cd = 2 ppm (insignificant at conventional levels), and no elasticities are involved here. The estimates for the two quadratic forms (columns 4 and 5 of Table 4) are somewhat more difficult to interpret at first sight. Examination of the estimates returned by the quadratic form in Cd levels (column 4) reveals that the implied elasticities have the unexpected sign at levels of Cd contamination below 1.43 ppm. Beyond that specific Cd level the elasticities become negative, and the values they take are remarkably close to the ones obtained from our baseline specification (column 1). However, the elasticities obtained from this quadratic-form specification are statistically insignificant at *all* evaluation points. A broadly similar, though somewhat less pronounced, picture emerges for the elasticities returned by the quadratic form in the *logged* levels of Cd contamination (column 5); they become negative (but only marginally significant) beyond a Cd level of 0.83 ppm. An interesting point to notice here is that imposing the zero-impact restriction at Cd levels below the threshold of 2 ppm (columns 1–2) increases the efficiency of the estimated elasticities to a considerable extent.

To further illustrate the superiority of our baseline specification of the Cd-pollution term, we plot the simulated price responses for each of the alternative specifications against the level of Cd pollution in Fig. 5, where the numbers displayed between parentheses in the interior part of the graph match with the column numbers of Table 4. It can be seen that all specifications roughly point in the same direction, but do so to varying degrees. The attractive features of our baseline specification, as defined in Eq. (8), becomes clearly visible from the pattern displayed by curve (1) in Fig. 5. Specifically, the negative impact of Cd pollution becomes effective only with a timid start at 2 ppm, and we see no abrupt and exponentially-declining price response immediately above 2 ppm, as featured by curve (2), nor a highly implausible break at 2 ppm, as displayed by curve (3). Second, curve (1) exhibits a natural (mild) concave-convex response, where the point of inflection occurs at a Cd level of 3.13 ppm, according to Eq. (9). Finally, unlike curve (1), the continuous specifications, represented by the curves (4) and (5), display counter-intuitive price responses at Cd levels below 2 ppm.

Insert **Fig. 5** about here

5.4 Robustness Checks

To conclude the empirical section, we check the robustness of the empirical findings to alternative model specifications by experimenting with different sets of covariates and controls. By doing so, we take advantage of an appealing feature of UQR, relative to CQR, that altering the set of covariates and controls does not change the estimated object, such that comparisons across alternative specifications are informative as changes in parameter estimates can be attributed solely to changes in identification of the parameters of interest (see also Maclean et al. 2014, p. 204). The results of the robustness checks are summarized in Table 5 (where the first column repeats the results reported in column 6 of Table 2, to ease comparison).

Insert **Table 5** about here

In column 2 of Table 5, we show the results obtained from incorporating the (log) distance *(*in km) of each farmland plot to the nearest urban center (town centroid), as a potentially valid alternative to the address-density variable, to capture the influence of local development pressures on farmland prices. The estimated coefficient on the distance variable is negative (−0.089) and significant at the 5% level, while the estimated coefficient on the Cd-pollution term (−1.328) hardly changes in magnitude or statistical significance.

Next, in column 3 of Table 5, we add the (logged) distance in km of each farmland plot to the nearest pollution source (former zinc-smelting site) to our pollution term, which is consistent with the popular distance-based strategy of incorporating environmental disamenities in hedonic-pricing analyses. The estimated coefficient on the distance variable (−0.038) is insignificant (besides having an unexpected sign), however, while the coefficient on the Cd-pollution term (−1.421) hardly changes in magnitude and significance. So the omitted-variable bias that may arise by neglecting the proximity of the land plot to the nearest pollution source (toxic site), as pointed out by Leggett and Bockstael (2000), does not seem to occur in the present case. Thus, incorporating the distance-to-source variable does not matter (which is not surprising, since we are not dealing with residential land prices), while removing the Cd-pollution term, in turn, was not found to improve the estimated distance coefficient.

Furthermore, we examine whether results change by omitting either the municipality dummies, or the year dummies, or both. First, we drop the municipality dummies. The results are shown in column 4 of Table 5, where the (absolute) size of the estimated coefficient on the Cd-pollution term becomes somewhat smaller (−1.172). Second, we delete the year dummies. The results are shown in column 5 of Table 5. The alteration of the Cd-pollution coefficient is negligible (−1.394). Third, we remove both municipality and year dummies. The results, in column 6 of Table 5, indicate that the estimated effect of Cd pollution on farmland prices decreases a bit more in this case (−1.098).

Finally, the previous specification with no municipality and year dummies is augmented with two time-variant characteristics defined at the municipality level, namely: (i) the (one-year lagged) five-year moving averages of (CPI deflated) prices per square meter of developable land in a municipality, following Geniaux et al. (2011); and (ii) the area of cultivated land as a proportion of the municipal area.[24] The results are provided in column 7 of Table 5. Concerning prices of developable land, we find that the estimated coefficient implies a significant "cross-elasticity" of about 0.4, which highlights the close link that exists between the local markets for agricultural and developable land. Accordingly, buyers in the market may adjust their maximum bid for agricultural land not only depending on its potential for continued farming, but also—and perhaps primarily—on the basis of anticipated future land-use conversions (e.g., Snyder et al. 2007). With regard to the proportion of cultivated land, the estimated coefficient is also positive and strongly significant. The appreciation of the median price could signify a positive valuation of land that is likely to be more suitable for agricultural uses, at least to the extent that differences in the proportions of cultivated land across municipalities represent differences in agricultural productivity. The estimated impact of Cd contamination on farmland prices appears to be somewhat less pronounced for this specification (−0.987).

In sum, we find that the estimated price discount due to increasing Cd pollution at the median of the price distribution is quite robust to changes in model specification, with estimates staying fairly close to the value returned by the baseline model, especially if municipality dummies are included. In fact, inspection of the standard errors reveals that the estimated impacts of Cd pollution on farmland prices are not significantly different from one another across model specifications.

6. Concluding Remarks

The purpose of this paper was to provide evidence of how cadmium (Cd) pollution of the soil affects farmland prices in a small area of the Campine region in Belgium. To estimate the hedonic (implicit) price of Cd contamination, we used the new method of unconditional quantile regression (UQR), introduced by Firpo, Fortin, and Lemieux (2009). The present study benefited from three important advantages of UQR over conventional ordinary least squares (OLS) and standard conditional quantile regression (CQR). First, UQR goes beyond the impact of Cd on just the mean of the price distribution. Second, UQR estimates the heterogeneous impact of Cd pollution at different quantiles of the entire distribution of farmland prices, thus for all sales transactions in the sample population. Third, UQR results are much easier to interpret. That is, unlike CQR results, they have a direct and more intuitive interpretation, namely as the impact of Cd pollution along the overall distribution of farmland prices, conditional on other covariates. Evidently, these are important features that are of great relevance for the proper evaluation of the cost-benefit performance of potential future soil remediation and soil-management policies and regulations. To the best of our knowledge, this study is the first in which UQR has been applied to hedonic valuation.

The UQR estimations provided convincing evidence of a U-shaped pattern of the influence of soil pollution on farmland prices, where an increasing level of Cd pollution has a negative and significant impact only in the middle range of the price distribution—insofar as Cd contamination levels are above the regulatory standard of 2 ppm imposed by the Flemish government, thus at pollution levels considered hazardous for agricultural production. It was found that the elasticities of the farmland price with respect to the level of Cd concentration at the median of the distribution are −1.1 and −1.9 (and strongly significant), when evaluated at Cd levels of 3 ppm and 4 ppm, respectively. Conversely, the elasticities at the lower and upper ends of the price distribution turned out to be not significantly different from zero, at all levels of Cd pollution. The non-uniform valuation of Cd contamination that emerged can essentially be traced back to the different amenities provided by farmland, namely agricultural production (with significant impact of Cd) as opposed to open space/recreation and rural living (with no significant impact of Cd).

Like all empirical studies, this study has also limitations. We faced a number of practical problems that typically arise when applying hedonic-valuation methods to assess the impacts of environmental disamenities. These problems include, among other things: (i) small sample size, (ii) difficult-to-measure pollutants (often not observable for buyers), and (iii) non-existence of property-specific pollution measurements (see Guignet 2012, p. 54). In this respect, our study is not different from other hedonic studies found in the environmental-economics literature. However, despite these limitations, we were able to obtain statistically significant and substantively meaningful results. One such result suggests that soil-remediation measures would raise the median price of farmland, rendering the price distribution more symmetric and, therefore, generating potential welfare gains for certain groups of participants in the local land market. Of course, prices may not always rebound after a cleanup of hazardous land, considering that an environmental stigma of the public's perception may linger for quite some time after the removal of the heavy-metal pollutants (e.g., Dale et al. 1999; Gregory and Satterfield 2002; Messer et al. 2006). Yet it is hardly imaginable that the stigma effect would persist indefinitely (e.g., Thanos et al. 2014).

Finally, our empirical case has demonstrated the flexibility and usefulness of UQR, as well as its advantages over conventional mean regression (OLS) and standard CQR. We also feel that we have convincingly shown that OLS leads to conclusions that are vastly oversimplified and, hence, that OLS is not really suitable as an analytical tool of hedonic valuation. Given the (relative) ease with which UQR is implemented in practice, and the credible results returned by this new estimation method, we recommend its use to anyone involved in the hedonic valuation of environmental (dis)amenities.

Notes

[1] Synthesized information (written in Dutch) on the Cadmium Action Plan and the BeNeKempen project is online available at http://www.milieurapport.be/Upload/main/miradata/MIRA-T/02_themas/02_15/Synthesetekst_MiraT2006-08Def.pdf (MIRA–Milieurapport Vlaanderen/Flanders Environmental Report).

[2] Cadmium has been identified as the trace element with the highest bio-accumulation index in green plants, causing severe health risks (e.g., kidney damage, bone decalcification, and cancer). As a result of that, farmers are not allowed to grow crops/vegetables for human consumption on heavily contaminated land.

[3] To verify if farmers are compliant with existing food regulations, the Federal Agency for the Safety of the Food Chain (FASFC), an executive body with jurisdiction over the entire Belgian territory, collects a large number of food samplings and performs different kinds of controls and inspections. To enforce compliance with the law, non-compliance is penalized, and transgressors even risk legal prosecution. OVAM takes action only in the case of land that is directly in danger; i.e., land that is potentially hazardous due to former polluting activities on-site.

[4] The average (Google-Maps) distances between the 14 municipalities in our study area and the cities of Brussels and Antwerp are 94.7 km (min = 68.1, max = 117.0) and 71.5 km (min = 56.0, max = 90.9), respectively.

[5] CQR has been well documented elsewhere (Koenker and Bassett 1978; Koenker and Hallock 2001). Therefore, it will not be further discussed here. An example of the application of CQR to the hedonic valuation of farmland characteristics is Uematsu et al. (2013). Their study is closely linked to ours through its focus on the contribution of natural-amenity attributes to the price of farmland.

[6] We just give a brief account of UQR. Readers unfamiliar with UQR and who wish to learn more about it are directed to the *Econometrica* paper by FFL (2009). Some good expositions of UQR can be found in Borah and Basu (2013) and Fournier and Koske (2012).

[7] This property does not hold for CQR, for the simple reason that, say, observations at the top of a conditional distribution may be at the bottom of the unconditional distribution.

[8] In the present study, sales transactions are only those transfers of farmland where a landowner sells a plot of land to a buyer at an agreed price. Transactions in the rental market have been excluded from our analysis, since rental prices are regulated by the Flemish government; that is, they are not allowed to exceed a legal maximum level, where the latter is below the market price. Moreover, there is legislation to protect the tenant (e.g., in the form of long-term contracts), and the tenant has a pre-emptive right to buy the tenanted land if it is for sale. Thus, a farmer with a written lease cannot easily be evicted.

[9] The GIS files are polygons of plot boundaries. Given the generally small size of the farmland plots in our sample (see summary statistics in Table 1 below), our geo-referenced variables are expected to be quite robust.

[10] In Section 5.3 we compare the performance of the baseline model specification with some other, more commonly-used specifications.

[11] These covariates have been included as control variables, and are primarily aimed at helping the identification of the Cd parameter of interest. So the coefficients on the former do not necessarily have a “causal” interpretation.

[12] Evidently, built structures vary tremendously in quality, size, age, etc. Unfortunately, our data set contains only information about the presence or otherwise of built structures.

[13] One could, of course, argue that the choice of the one-km distance is arbitrary, but we consider this radius as reasonable, given the “peculiar” landscape features of our study area. If we had chosen, say, a larger distance, the local nature of the urban-development pressure might have been lost. Also, the use of distance-buffer variables is not uncommon in the literature. For instance, Alberini (2007) used a variable defined as the percentage of land slated for residential use within 1,500 m of the parcel.

[14] Specifically, it was found that the correlation between the time-series variation of the average real prices per square meter of developable land (averaged over the 14 municipalities of the study area) and the estimated year dummies in our baseline hedonic-pricing model is 0.96.

[15] Our choice of the length of the time period is partly motivated by the modest size of our data set. We choose a 12-month time-window, which implies that we lose the first 64 observations (or 11% of the original sample). If we had chosen 24 months, we would have already lost 160 observations (or 27% of the original sample).

[16] The Euclidian distances (in km) between the farmland plots in our sample (the elements in the distance matrix **D**) have been calculated on the basis of the so-called Lambert (x, y) coordinates. The mean (median) distance between the farmland plots is 15.3 km (14.3 km). The minimum distance is only 9 m, while the maximum distance is 49.6 km. The largest minimum distance is 4.1 km (i.e., there is one plot that is located at 4.1 km from its nearest neighbor), while the smallest maximum distance is 25.7 km (i.e., there is one plot that is located within a 25.7-km distance of every other plot).

[17] It should be noted, in passing, that the UQR estimate (for the median of the price distribution) is two and four times larger in absolute size than its CQR (for the median) and OLS (for the mean) counterparts, respectively.

[18] To obtain the full covariance matrix of the estimated coefficients, necessary to implement the Wald-type F test, we had to perform a prior bootstrap estimation (based on 500 replications).

[19] For example, although the quantile estimate at the 80th percentile ($\hat{\beta}_{0.80} = -0.236$) lies outside the 95% CI of the estimate at the median ($\hat{\beta}_{0.50} = -1.381$), as can be seen from Fig. 3, the null hypothesis of their equality could not be rejected at conventional levels of significance ($F_{\hat{\beta}_{0.80}-\hat{\beta}_{0.50}} = 1.38, p = 0.241$).

[20] It is important to note that a probit analysis as conducted here is meaningful only in the context of UQR (thus not for CQR), as it provides useful information on the observable association between the characteristics of the land plots in the sample population and their position in the unconditional price distribution.

[21] While being an intriguing issue, any further investigation of it to make stronger claims is beyond the scope of the present paper, as our data set contains no information on the buyers' identity and/or personal characteristics and preferences as, e.g., in Kostov (2010) and Curtiss et al. (2013).

[22] A more extensive overview of the estimated elasticities is provided below in Table 4 of Section 5.3, where our baseline elasticities (evaluated at the median of the price distribution) will be compared with those returned by some alternative specifications of the Cd-pollution term.

[23] The OLS estimate of the coefficient on the *public* dummy is 0.209 (see column 1 of Table 2). This means that, on average, the public-sale (auction) price is estimated to be about 23.2% $= (e^{0.209} - 1) \times 100$ higher than the private-sale price. Or, the price stated to the government (full price net of side payment) is, on average, about 18.8% lower than the full transaction price. This percentage is close to the 20% mentioned in Ciaian et al. (2012, p. 8).

[24] Municipal-level data on average prices of developable land and areas of cultivated land were available from the Flemish portal of local statistics, http://aps.vlaanderen.be/lokaal/lokale_statistieken.htm.

Acknowledgements

The authors thank the co-editor, Phoebe Koundouri, and two anonymous reviewers for their valuable comments and suggestions on earlier versions of the paper. A special word of thanks is directed to one of the reviewers for providing exceptionally detailed and critical comments that led to a significant improvement of the paper. The authors also thank the participants of the 6th Jean Paelinck Seminar (Madrid 2013) and the 14th EAAE Congress (Ljubljana 2014) for helpful remarks. The first author is grateful to Yves Surry for sharing his views on issues related to quantile regression at an early stage of the paper.

References

Alberini A (2007) Determinants and effects on property values of participation in voluntary cleanup programs: The case of Colorado. Contemporary Economic Policy 25(3):415–432

Baranzini A, Schaerer C, Thalmann P (2010) Using measured instead of perceived noise in hedonic models. Transportation Research, Part D: Transport and Environment 15:473–482

Barrios T, Diamond R, Imbens GW, Kolesár M (2012) Clustering, spatial correlations, and randomization inference. Journal of the American Statistical Association 107(498):578–591

Borah BJ, Basu A (2013) Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. Health Economics 22(9):1052–1070

Borah BJ, Naessens J, Olsen K, Shah N (2013) Explaining obesity- and smoking-related healthcare costs through unconditional quantile regression. Journal of Health Economics and Outcomes Research 1(1):23–41

Boyle M, Kiel KA (2001) A survey of house price hedonic studies of the impact of environmental externalities. Journal of Real Estate Literature 9(2):117–144

Capozza DR, Helsley RW (1989) The fundamentals of land prices and urban growth. Journal of Urban Economics 26(3):295–306

Cavailhès J, Wavresky P (2003) Urban influences on periurban farmland prices. European Review of Agricultural Economics 30(3):333–357

Chay KY, Greenstone M (2005). Does air quality matter? Evidence from the housing market. Journal of Political Economy 113(2):376–424

Chicoine DL (1981) Farmland values at the urban fringe: An analysis of sale prices. Land Economics 57(3):353–362

Cho SH, Roberts R, Kim S (2011) Negative externalities on property values resulting from water impairment: The case of the Pigeon River watershed. Ecological Economics 70:2390–2399

Ciaian P, Kancs d'A, Swinnen J, Van Herck K, Vranken L (2012) Institutional factors affecting agricultural land markets. Factor Markets Working Paper No. 16 (Feb. 2012)

Cohen J, Blinn CE, Boyle KJ, Holmes TP, Moeltner K (2015) Hedonic valuation with translating amenities: Mountain Pine Beetles and host trees in the Colorado Front Range. Environmental and Resource Economics. DOI 10.1007/s10640-014-9856-y

Colwell PF, Munneke HJ (1997) The structure of urban land prices. Journal of Urban Economics 41(3):321–336

Cotteleer G, Gardebroek C, Luijt J (2008) Market power in a GIS-based hedonic price model of local farmland markets. Land Economics 84(4):573–592

Curtiss J, Jelínek L, Hruška M, Medonos T, Vilhelm V (2013) The effect of heterogeneous buyers on agricultural land prices: The case of the Czech land market. German Journal of Agricultural Economics 62(2):116–133

Dale L, Murdoch JC, Thayer MA, Waddell PA (1999) Do property values rebound from environmental stigmas? Evidence from Dallas. Land Economics 72(2):311–326

Deaton BJ, Vyn RJ (2010) The effect of strict agricultural zoning on agricultural land values: The case of Ontario's greenbelt. American Journal of Agricultural Economics 92(4):941–955

Dubé J, Legros D (2014) Spatial econometrics and the hedonic pricing model: What about the temporal dimension? Journal of Property Research 31(4):333–359

Ecker MD, Isakson HR (2005) A unified convex-concave model of urban land values. Regional Science and Urban Economics 35(3):265–277

Firpo S, Fortin NM, Lemieux T (2009) Unconditional quantile regressions. Econometrica 77(3):953–973

Fournier J, Koske I (2012) Less income inequality and more growth–Are they compatible? Part 7. The drivers of labour earnings inequality–An analysis based on conditional and unconditional quantile regressions. OECD Economics Department Working Papers No. 930, OECD Publishing

Gamper-Rabindran S, Timmins C (2013) Does cleanup of hazardous waste sites raise housing values? Evidence of spatially localized benefits. Journal of Environmental Economics and Management 65:345–360

Geniaux G, Ay J-S, Napoléone C (2011) A spatial hedonic approach on land use change anticipations. Journal of Regional Science 51(5):967–986

Gregory RS, Satterfield TA (2002) Beyond perception: The experience of risk and stigma in community contexts. Risk Analysis 22(2):347–358

Guignet D (2012) The impacts of pollution and exposure pathways on home values: A stated preference analysis. Ecological Economics 82(1):53–63

Guignet D (2013) What do property values really tell us? A hedonic study of underground storage tanks. Land Economics 89(2):211–226

Guiling P, Brorsen BW, Doye D (2009) Effect of urban proximity on agricultural land values. Land Economics 85(2):252–264

Hess S, Daly A (2014) Handbook of choice modelling. Edward Elgar Publishing, Inc., Northampton, MA.
Hogervorst J, Plusquin M, Vangronsveld J, Nawrot T, Cuypers A, Van Hecke E, Roels HA, Carleer R, Staessen JA (2007) House dust as possible route of environmental exposure to cadmium and lead in the adult general population. Environmental Research 103:30–37
Howland M (2000) The impact of contamination on the Canton/Southeast Baltimore land market. Journal of the American Planning Association 25(3):411–420
Hu H, Jin Q, Kavan P (2014) A study of heavy metal pollution in China: Current status, pollution-control policies and countermeasures. Sustainability 6:5820–5838
Huang H, Miller GY, Sherrick BJ, Gómez MI (2006) Factors influencing Illinois farmland values. American Journal of Agricultural Economics 88(2):458–470
Jackson TO (2002) Environmental contamination and industrial real estate prices. Journal of Real Estate Research 23(1/2):179–199
Johnson BB, Chess C (2003) How reassuring are risk comparisons to pollution standards and emission limits? Risk Analysis 23(5):999–1007
Koenker R, Bassett G (1978) Quantile regressions. Econometrica 46(1):33–50
Koenker R, Hallock KF (2001) Quantile regression. Journal of Economic Perspectives 15(4):143–156
Kostov P (2010) Do buyers' characteristics and personal relationships affect agricultural land prices? Land Economics 86(1):48–65
Leggett CG, Bockstael NE (2000) Evidence of the effects of water quality on residential land prices. Journal of Environmental Economics and Management 39(2):121–144
Livanis G, Moss CB, Breneman VE, Nehring RF (2006) Urban sprawl and farmland prices. American Journal of Agricultural Economics 88(4):915–929
Lizin S, Schreurs E, Van Passel S (2015) Farmers' perceived cost of land use restrictions: A simulated purchasing decision using discrete choice experiments. Land Use Policy 46:115–124
Ma S, Swinton SM (2012) Hedonic valuation of farmland using sale prices versus appraised values. Land Economics 88(1):1–15
Maclean JC, Webber DA, Marti J (2014) An application of unconditional quantile regression to cigarette taxes. Journal of Policy Analysis and Management 33(1):188–210
Maddison D (2009) A spatio-temporal model of farmland values. Journal of Agricultural Economics 60(1):171–189
McMillen DP (2013) Quantile Regression for Spatial Data. Springer Briefs in Regional Science
Messer KD, Schulze WD, Hackett KF, Cameron TA, McClelland GH (2006) Can stigma explain large property value losses? The psychology and economics of superfund. Environmental and Resource Economics 33(3):299–324
Nawrot T, Van Hecke E, Thijs L, Richart T, Kuznetsova T, Jin Y, Vangronsveld J, Roels HA, Staessen JA (2008) Cadmium-related mortality and long-term secular trends in the cadmium body burden of an environmentally exposed population. Environmental Health Perspectives 116(12):1620–1628
Patton M, McErlean S (2003) Spatial effects within the agricultural land market in Northern Ireland. Journal of Agricultural Economics 54(1):35–54
Plantinga AJ, Lubowski RN, Stavins RN (2002) The effects of potential land development on agricultural land prices. Journal of Urban Economics 52(3):561–558
Poor PJ, Boyle KJ, Taylor LQ, Bouchard R (2001) Objective versus subjective measures of water clarity in hedonic property value models. Land Economics 77(4):482–493
Schreurs E, Voets T, Thewys T (2011) GIS-based assessment of the biomass potential from phytoremediation of contaminated agricultural land in the Campine region in Belgium. Biomass and Bioenergy 35(10):4469–4480
Snyder SA, Kilgore MA, Hudson R, Donnay J (2007) Determinants of forest land prices in northern Minnesota: A hedonic pricing approach. Forest Science 53(1):25–36
Thanos S, Bristow AL, Wardman MR (2014) Residential sorting and environmental externalities: The case of nonlinearities and stigma in aviation noise. Journal of Regional Science. DOI: 10.1111/jors.12162
Uematsu H, Khanal AR, Mishra AK (2013) The impact of natural amenity on farmland values: A quantile regression. Land Use Policy 33(July):151–160
Van Meirvenne M, Goovaerts P (2001) Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. Geoderma 102(1/2):75–100
Van Meirvenne M, Meklit T (2010) Geostatistical simulation for the assessment of regional soil pollution. Geographical Analysis 42(2):121–135

Appendix A: Differences Between CQR and UQR – Illustrative Simulation Example

This appendix presents a simple simulation example (borrowed with some modifications) from McMillen (2013) to illustrate the differences between traditional CQR and new UQR, and to point to potential problems that may arise when interpreting the estimates returned by CQR.

Consider a single explanatory variable, X, which is limited to take integer values from 1 to 10. Each integer value occurs 200 times in the simulated data set, leading to a total of 2,000 observations. The regression equation is specified as $Y = 10 - 0.5X + \varepsilon$. To ensure an R^2 of approximately 0.8 for the OLS (mean) regression, we specify $\varepsilon \sim N[0, \sigma_\varepsilon(X)]$, where $\sigma_\varepsilon^2(X) = \frac{0.2\sigma_X^2}{X}$. This variance function for ε implies that the quantiles of the conditional distribution of Y given X converge; that is, the dispersion (variance) of the conditional distribution of Y decreases with higher values of X.

The results from OLS, CQR, and UQR are summarized in Table A.1. Selected CQR regression lines, representing the 10th, 50th, and 90th percentiles of the conditional distribution of Y given X, are shown in Fig. A.1. The OLS regression line represents the conditional mean of Y.

Insert **Table A.1** about here

Insert **Fig. A.1** about here

The points a and b in Fig. A.1 show why standard CQR results may be misinterpreted. Point a is associated with a high value of X on the 90th-percentile regression line, whereas point b represents a low value of X on the 10th-percentile regression line. The value of the dependent variable Y at point a is *lower* than its value at point b, where the former is positioned at the bottom end of the unconditional distribution of Y. Therefore, it is *not* the case that an increase in X leads to a greater decline in Y whenever Y is *unconditionally* high. The steeper slope at the 90th percentile merely indicates that an increasing X leads to greater declines in Y along the 90th percentile of Y than at its 10th percentile, *conditional* on the values of X.

Another way to illustrate the differences between the estimated effects of X on Y using either CQR or UQR is to look at the *quantile plots* presented in Fig. A.2. Specifically, they show the CQR and UQR estimates (along with their 95% confidence bands) of the effects of X at nine different quantiles (from the 10th to 90th percentile). It can immediately be seen that the estimated effect obtained using UQR is highly non-monotonic, exhibiting a pronounced U-shaped pattern, whereas the effect obtained using standard CQR declines monotonically. It is interesting to see that the pattern displayed by the quantile coefficients obtained using UQR is close to a "mirror image" of the inverse U-shaped pattern shown in FFL (2009, their Fig. 1, p. 965).

Insert **Fig. A.2** about here

The quantile coefficients obtained using CQR are the estimated slopes of the regression lines in Fig. A.1, which show how the quantiles of the *conditional* distribution of Y given X are affected by a 1-unit change in X. Thus, CQR and UQR estimate two different things, where CQR focuses on the unobserved heterogeneity of Y reflected by the distribution of Y for a given value of X, $F(Y|X) = F(\varepsilon)$, while UQR centers on the observed distribution of Y for the entire population, $F(Y)$. Yet it is curious to see that in many empirical applications of CQR, results are often erroneously interpreted as if they were UQR results.

Appendix B: Within- and Between-Municipality Covariances among Farmland Prices

This appendix provides simple descriptive measures of the covariance (or correlation) patterns among observed prices as a function of the distances between the farmland plots.

Following Barrios et al. (2012), we calculated the *within*- and *between*-municipality covariances for pairs of prices of farmland plots whose (straight-line) distance in km is within some bandwidth h of a given distance d; that is,

$$\widehat{C_W(d)} = \frac{\sum_{i<j} \mathbf{1}_{m_i=m_j} \cdot \mathbf{1}_{|d_{ij}-d| \leq h} \cdot \tilde{p}_i \cdot \tilde{p}_j}{\sum_{i<j} \mathbf{1}_{m_i=m_j} \cdot \mathbf{1}_{|d_{ij}-d| \leq h}} \tag{B.1}$$

and

$$\widehat{C_B(d)} = \frac{\sum_{i<j} \mathbf{1}_{m_i \neq m_j} \cdot \mathbf{1}_{|d_{ij}-d| \leq h} \cdot \tilde{p}_i \cdot \tilde{p}_j}{\sum_{i<j} \mathbf{1}_{m_i \neq m_j} \cdot \mathbf{1}_{|d_{ij}-d| \leq h}} \tag{B.2}$$

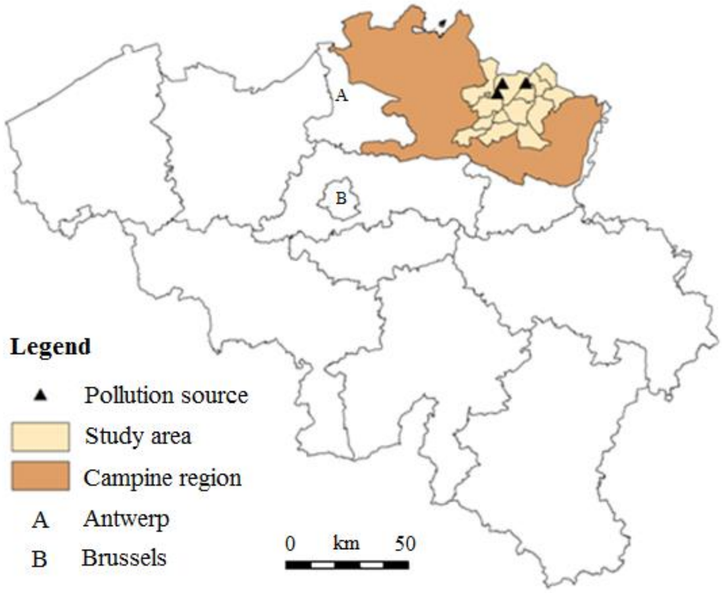
where $\tilde{p}_i, \tilde{p}_j$ are the "demeaned" sales prices (measured as deviations from the overall sample mean). The symbols m_i and m_j denote the municipalities in which the plots i and j are located. Specifically, $m_i = m_j$ ($m_i \neq m_j$) means that farmland plots i and j are (not) located in the same municipality. Note that all farmland plots have been ordered, and the condition $i < j$ just takes account of the symmetry of the variance-covariance matrix.

The panels in Fig. B.1 show the covariance functions for bandwidth h going from 1 to 4 or the within- and between-municipality covariances. For example, with bandwidth $h = 1$, we look at the prices of all plots j located at a distance between 0 and 2 km from plot i for $d = 1$, between 1 and 3 km for $d = 2$, between 2 and 4 km for $d = 3$, and so on.

Insert **Fig. B.1** about here

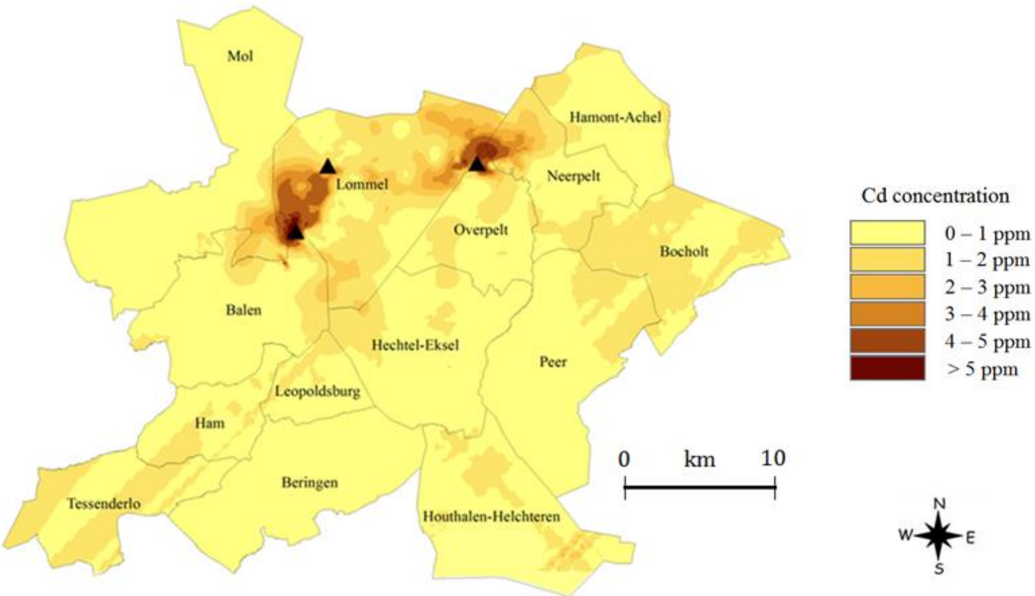
The main conclusions from the graphs presented in Fig. B.1 are that the covariances decrease with distance, and that the between-municipality covariances are of a magnitude much smaller than the within-municipality covariances. Therefore, the results provide some empirical support for (i) the importance of using simple *inverse* distances, and (ii) the land-market *segmentation* along municipal boundaries. In a way, the latter can be interpreted as a manifestation of "home biases" along municipal boundaries, due to higher transaction or information-search costs associated with the prices prevailing "on the other side of the fence". Obviously, this is just a conjecture, which has not been further tested here.

Based on the findings from this covariance analysis, our empirical application in the main text uses the only the *within*-municipal farmland prices in constructing the spatiotemporal lag of the dependent variable.



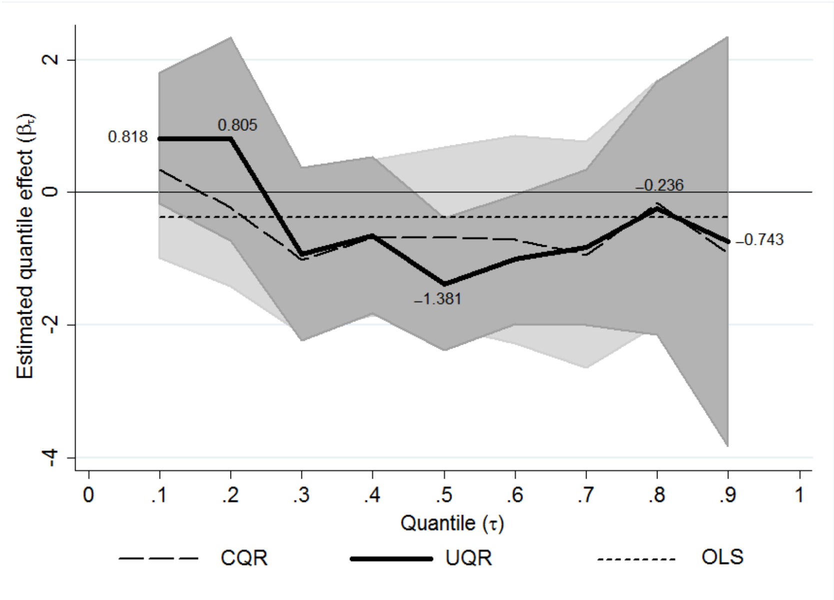
The study area covers 14 municipalities: Balen, Beringen, Bocholt, Ham, Hamont-Achel, Hechtel-Eksel, Houthalen-Helchteren, Leopoldsburg, Lommel, Mol, Neerpelt, Overpelt, Peer, and Tessenderlo.

Fig. 1 Location map of Campine region and study area in Belgium



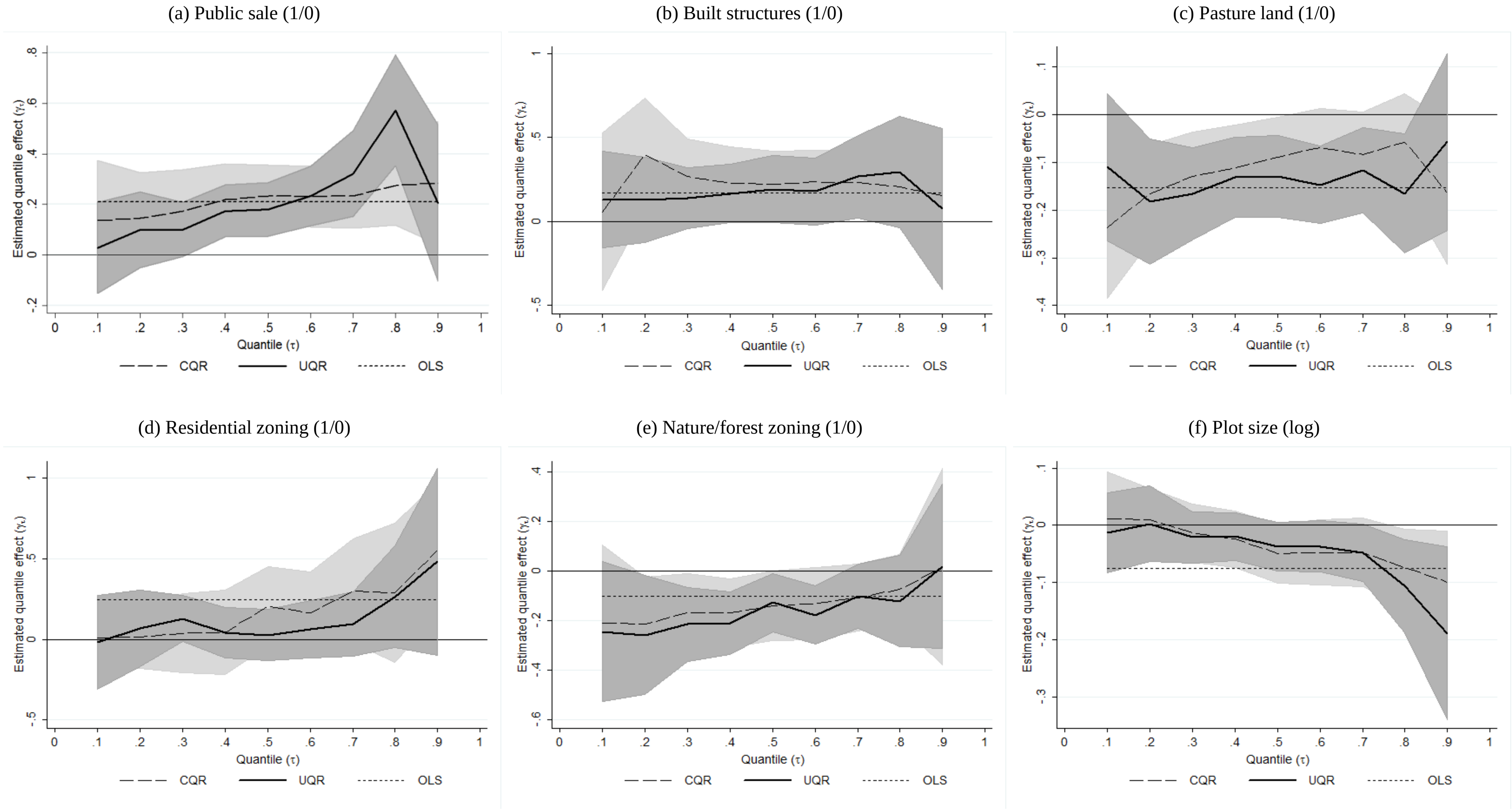
The map shows the three sites associated with the historical Cd contamination due to zinc-smelting/-alloying activities. The sites are located in the municipalities of Balen, Lommel, and Overpelt. During our sample period 2004–2011, the sites in Balen and Overpelt were operated by Nyrstar (http://www.nyrstarbalenoverpelt.be), which reduced airborne Cd pollution to zero (by using clean, non-thermal zinc-refining technologies), whereas the site in Lommel was completely re-developed (after levelling the old structures) by Sibelco, whose business is the extraction, production, and distribution minerals for industrial use (http://www.sibelco.be).

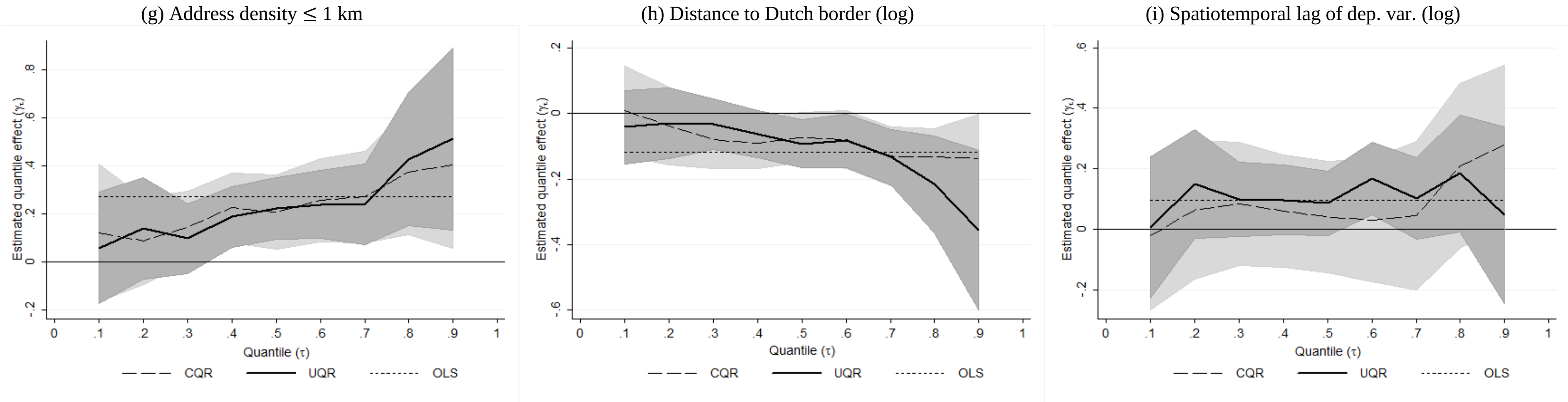
Fig. 2 Reference map of study area and spatial distribution of Cd pollution



The light and dark grey areas in the graph represent the 95% confidence intervals for the CQR and UQR estimates of the quantile effects of Cd pollution, respectively.

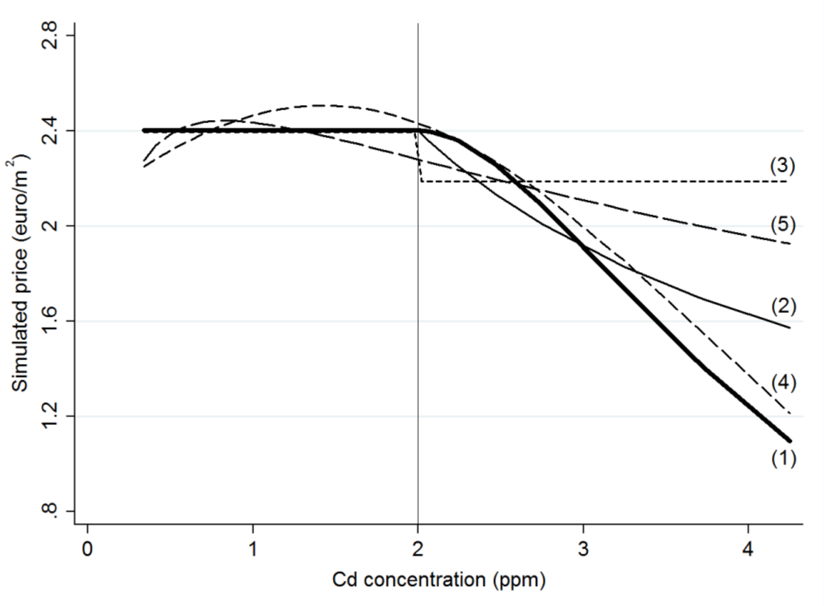
Fig. 3 Quantile plots for estimated coefficients on Cd-pollution term


(a) Public sale (1/0)
(b) Built structures (1/0)
(c) Pasture land (1/0)
(d) Residential zoning (1/0)
(e) Nature/forest zoning (1/0)
(f) Plot size (log)
Estimated quantile effect (γτ)
Quantile (τ)
CQR
UQR
OLS



The light and dark grey areas in the graphs represent the 95% confidence intervals for the CQR and UQR estimates of the quantile effects of other farmland characteristics, respectively.

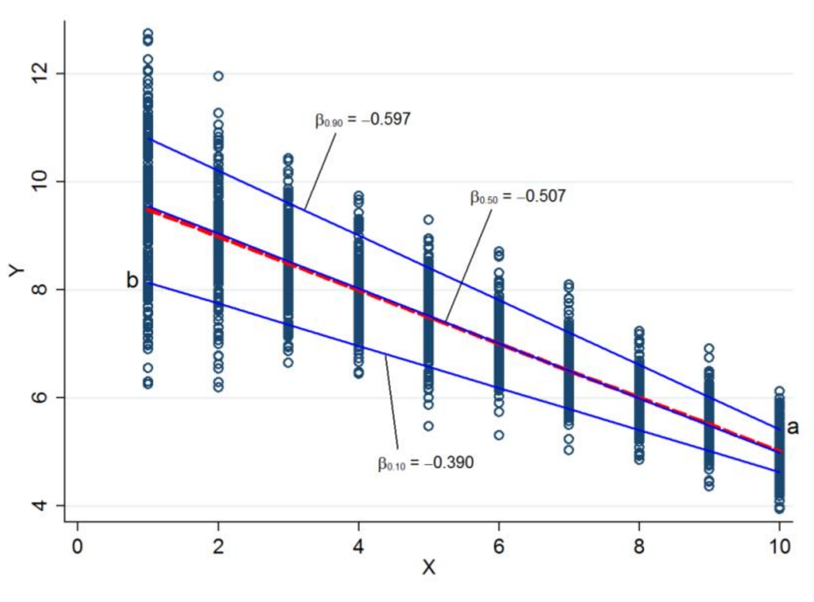
Fig. 4 Quantile plots for estimated coefficients on other farmland characteristics



The simulated prices for different Cd levels (while holding everything else constant) have been calculated at the means of the continuous variables (with the exception of Cd), and with all binary variables set to zero. The mean values in the estimation sample (511 observations) are: 1.42 for plot size; 0.414 for address density; 10.5 for distance to Dutch border; and 2.80 for the spatiotemporal price lag. The estimated constant term is 0.927. The numbers shown in the interior part of the graph, between parentheses, correspond to the column numbers of Table 4. The (thick) solid curve (1) represents the price-reponse curve predicted from our baseline specification, as defined in Eq. (8). For all specifications considered the (constant or average) predicted prices for Cd $\leq$ 2 ppm are almost identically equal to 2.40 euro/m^2. This price level matches closely with the observed median price in the estimation sample (2.37 euro/m^2).

Fig. 5 Simulated prices for alternative specifications of Cd-pollution term

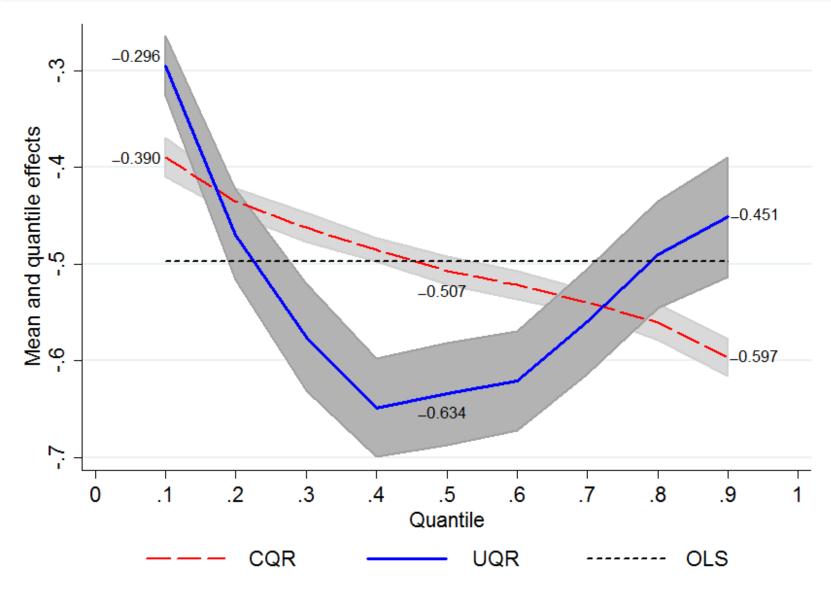
Appendix A



The blue lines represent the 10th, 50th, and 90th percentile-regression lines returned by CQR, whose slopes are equal to −0.390, −0.507, and −0.597, respectively. The red line is the mean-regression line obtained from using OLS, which practically coincides with the 50th percentile-regression line. The variance of the conditional distribution of Y given X decreases from about 1.8 for X = 1 to 0.2 for X = 10.

Fig. A.1 Predicted CQR and OLS regression lines

Appendix A

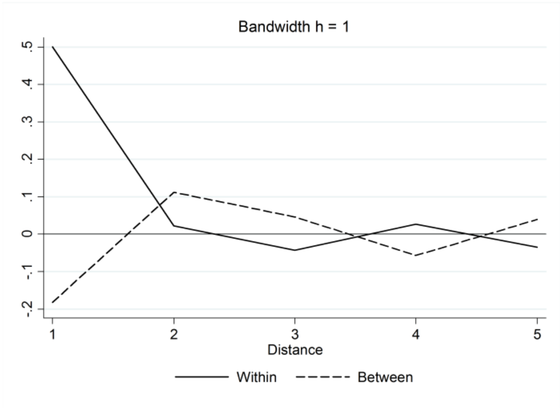


The vertical axis measures the mean and quantile effects at different quantiles; i.e., the estimated coefficients on X at different quantiles of the conditional and unconditional distributions of Y for CQR and UQR, respectively. The shaded areas represent the 95% confidence intervals for the CQR estimates (light grey) and UQR estimates (dark grey).

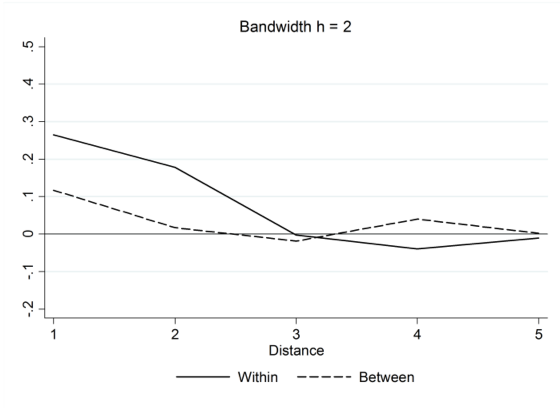
Fig. A.2 Quantile plots for OLS, CQR, and UQR

Appendix B

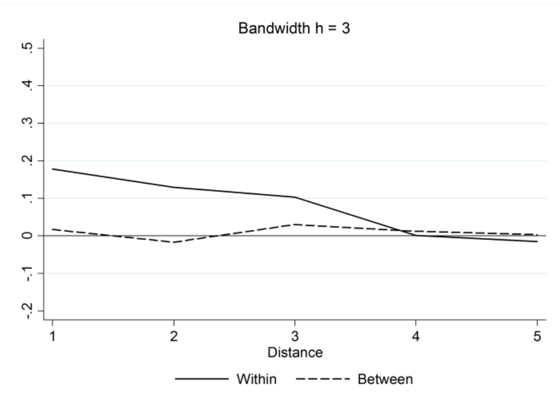
(a)



(b)



(c)



(d)

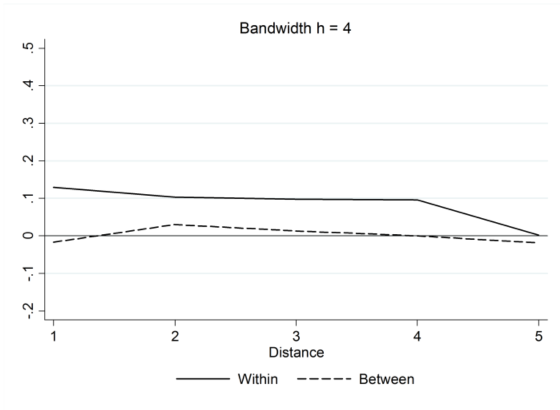


Fig. B.1 Within- and between-municipality price covariances

Table 1 Basic descriptive statistics

Variable	Description	Measurement unit	#Obs.	Mean	S.D.	Min.	Q25	Median	Q75	Max.
(a) Dependent variable										
rpricefland	Real price of farmland per m^2	euro	599	2.82	2.35	0.17	1.82	2.34	2.98	23.57
(b) Transaction characteristics										
public	Public sale vs. private sale	=1 if public sale	599	0.184	0.388	0				1
plotsize	Plot size	hectare	599	1.40	4.55	0.02	0.34	0.68	1.39	103.15
residential	Residential vs. agricultural zoning	=1 if residential	599	0.078	0.269	0				1
natureforest	Nature/forest vs. agricultural zoning	=1 if nature/forest	599	0.109	0.311	0				1
pasture	Pasture land vs. arable and other land	=1 if pasture	599	0.424	0.495	0				1
structures	Built structures vs. vacant	=1 if structure	599	0.052	0.222	0				1
addressdensity	Address density ≤ 1 km	thousands	599	0.414	0.340	0	0.175	0.326	0.579	2.051
distcenter	Distance to nearest local urban center	km	599	3.6	1.8	0.3	2.2	3.3	4.7	14.8
distborder	Distance to Dutch border	km	599	10.5	6.1	0	6.3	9.8	14.6	31.1
(c) Municipality characteristics										
cultland	Cultivated land	proportion	599	0.307	0.173	0.064	0.179	0.229	0.518	0.564
rpricedevland	Real price of developable land per m^2	euro	599	97.75	22.02	51.04	80.67	98.50	113.18	151.13
(d) Environmental variables										
cd	Cd-pollution level	ppm (mg kg^{-1})	599	1.007	0.688	0.255	0.750	0.750	1.250	5.436
	- Cd ≤ 2 ppm		555	0.852	0.358	0.255	0.750	0.750	1.250	1.873
	- Cd > 2 ppm		44	2.958	0.857	2.070	2.250	2.750	3.250	5.436
distsource	Distance to nearest pollution source	km	599	10.0	5.0	1.6	5.3	10.2	13.6	25.6

The sample contains 599 observations, distributed over the 14 municipalities belonging to the study area. The number of spatiotemporal lags of farmland prices are based on 597 for all prices (regardless of municipality), and 567 observations for "within-municipality" prices. Prices of farmland per square meter are expressed in constant 2011 euros (year and month CPI deflated). Prices per square meter of developable land are measured as five-year moving averages at the municipality level (CPI deflated annually). The variables *distcenter* (distance to nearest urban (local town) center), *distsource* (distance to nearest pollution source), *cultland* (proportion of cultivated land), and *rpricedevland* (prices of developable land) are only used for robustness checks (see table 5 below). The spatiotemporal lag of the dependent variable is constructed on the basis of the "within-municipality" prices realized for nearby land plots in the recent past.

Table 2 Estimation results obtained using OLS, CQR, and UQR

	OLS	CQR			UQR		
		Q10	Q50	Q90	Q10	Q50	Q90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Public sale	0.209***	0.136	0.234***	0.285**	0.028	0.180***	0.205
	(0.055)	(0.121)	(0.062)	(0.124)	(0.092)	(0.054)	(0.157)
Log plot size	−0.076***	0.012	−0.049*	−0.099**	−0.013	−0.037*	−0.189**
	(0.020)	(0.041)	(0.027)	(0.046)	(0.036)	(0.022)	(0.077)
Built structures	0.168*	0.058	0.220**	0.156	0.130	0.193*	0.075
	(0.087)	(0.238)	(0.101)	(0.118)	(0.147)	(0.102)	(0.245)
Pasture land	−0.153***	−0.236***	−0.089**	−0.163**	−0.109	−0.130***	−0.058
	(0.033)	(0.076)	(0.043)	(0.076)	(0.079)	(0.043)	(0.095)
Residential zoning	0.244**	0.009	0.205	0.554***	−0.018	0.026	0.480*
	(0.097)	(0.109)	(0.125)	(0.214)	(0.148)	(0.081)	(0.290)
Nature/forest zoning	−0.100*	−0.209	−0.140**	0.019	−0.244*	−0.128**	0.019
	(0.049)	(0.160)	(0.071)	(0.202)	(0.144)	(0.060)	(0.169)
Address density ≤ 1 km	0.272***	0.123	0.207***	0.405**	0.060	0.224***	0.511***
	(0.089)	(0.146)	(0.079)	(0.177)	(0.118)	(0.065)	(0.193)
Log distance to Dutch border	−0.117**	0.008	−0.073*	−0.137**	−0.042	−0.093**	−0.356***
	(0.038)	(0.070)	(0.039)	(0.069)	(0.057)	(0.037)	(0.123)
Log spatiotemporal lag of dependent variable	0.098	−0.021	0.040	0.279**	0.006	0.086	0.047
	(0.083)	(0.126)	(0.093)	(0.134)	(0.120)	(0.055)	(0.149)
$I(Cd > 2) \times [\log(Cd/2)]^2$	−0.378	0.343	−0.675	−0.928	0.818	−1.381***	−0.743
	(0.247)	(0.677)	(0.693)	(1.023)	(0.502)	(0.507)	(1.574)
Municipality dummies jointly = 0, *p* value	0.000	0.000	0.018	0.475	0.000	0.020	0.410
Year dummies jointly = 0, *p* value	0.000	0.855	0.053	0.057	0.968	0.018	0.626
Wald-type *F* test for equality of coefficients		1.52		0.06	10.27***		0.13
(*p* value)		(0.218)		(0.812)	(0.001)		(0.717)
Number of observations	511	511	511	511	511	511	511

The dependent variable in the case of UQR is the RIF of the (logged) real farmland prices. All models include municipality and year dummies. Omitted (reference) categories are private sale, vacant, arable and other land, and agricultural zoning. Standard errors are clustered around municipalities for OLS, and bootstrapped (500 replications) for CQR and UQR, and are reported in parentheses. The estimated constant terms and coefficients on the municipality and year dummies are not reported in the table to save space. Four influential observations, with Cd levels larger than 4.25 ppm (members of the 1% upper tail of the Cd distribution in the original sample), were removed from the estimation sample. The spatiotemporal lags are constructed on the basis of the within-municipality farmland prices only (see Appendix B). OLS and CQR were implemented using Stata's in-built commands "regress" and "bsqreg", respectively, while UQR was estimated using Stata's "rifreg" procedure, which is available at http://faculty.arts.ubc.ca/nfortin/datahead.html. The Wald *F* test is for the null hypothesis of pair-wise equality of the coefficients at the 50th and 10th or 90th percentiles.

*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

Table 3 Probit classification of farmland into price segments on the basis of observed characteristics

	Lower 20% price tail (1)	Middle 60% price range (2)	Upper 20% price tail (3)
Dependent variable: $\Pr(y = 1 \mid x)$ that farmland plot is within relevant price range or tail			
Public sale	−0.349**	−0.623***	1.055***
	(0.178)	(0.229)	(0.179)
Log plot size	−0.039	0.163**	−0.227***
	(0.084)	(0.067)	(0.062)
Built structures	−0.258	−0.220	0.558
	(0.349)	(0.341)	(0.424)
Pasture land	0.501***	−0.078	−0.358**
	(0.122)	(0.112)	(0.145)
Residential zoning	−0.136	−0.265	0.507**
	(0.308)	(0.290)	(0.248)
Nature/forest zoning	0.516***	−0.238*	−0.349**
	(0.176)	(0.134)	(0.172)
Address density ≤ 1 km	−0.503**	−0.413*	0.802***
	(0.263)	(0.217)	(0.210)
Log distance to Dutch border	0.135	0.240**	−0.401***
	(0.171)	(0.108)	(0.144)
Cd-pollution level (*Cd*)	−0.283*	0.126	0.086
	(0.162)	(0.139)	(0.158)
Number of observations used	497	511	505
Percentage of observations correctly classified	81.3%	67.3%	84.4%

Municipality and year dummies were included in all models. If a probit model is empirically under-identified (in our case, due to collinearity problems caused by the inclusion of municipality dummies in combination with several other binary variables), some of the original 511 observations in the estimation sample were dropped by Stata (see, e.g., Hess and Daly 2014, p. 520). Robust standard errors adjusted for intra-municipality correlations are given in parentheses.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

Table 4 UQR results for alternative Cd-pollution terms, for the median of the price distribution

	(1)	(2)	(3)	(4)	(5)
(a) Estimated coefficients					
$I(Cd > 2) \times [\log(Cd/2)]^2$	−1.381*** (0.507)				
$I(Cd > 2) \times \log(Cd/2)$		−0.560** (0.264)			
$I(Cd > 2)$			−0.090 (0.095)		
Cd				0.260* (0.141)	
Cd^2				−0.091** (0.039)	
$\log(Cd)$					−0.034 (0.058)
$[\log(Cd)]^2$					−0.089* (0.052)
(b) Implied elasticities evaluated at selected contamination levels					
Cd = 0.225 ppm	0	0		0.054 (0.036)	0.210 (0.154)
Cd = 1 ppm	0	0		0.078 (0.161)	−0.034 (0.058)
Cd = 2 ppm	0	0		−0.209 (0.422)	−0.157* (0.092)
Cd = 2.07 ppm	−0.095*** (0.035)	−0.560*** (0.264)		−0.242 (0.445)	−0.163* (0.095)
Cd = 3 ppm	−1.120*** (0.411)	−0.560*** (0.264)		−0.859 (0.823)	−0.229* (0.128)
Cd = 4 ppm	−1.914*** (0.703)	−0.560*** (0.264)		−1.874 (1.377)	−0.282* (0.155)
Cd = 4.25 ppm	−2.082*** (0.764)	−0.560*** (0.264)		−2.185 (1.539)	−0.292* (0.161)

All the coefficients and elasticities reported in the table are for the median ($\tau = 0.50$). The number of observations is invariably equal to 511 across the columns in the table. All models include municipality and year dummies. Bootstrapped standard errors (using 500 replications) are provided in parentheses. For the calculation of elasticities, we selected some specific values of Cd contamination, where 0.255 ppm is the minimum in the sample; 2.07 ppm is the smallest Cd level in the sample above the regulatory standard of 2 ppm; and 4.25 ppm is the maximum value in the estimation sample. The elasticities are calculated according to Eq. (10). It should be noted that the specification in column 4 returns *semi*-elasticities. For the sake of comparison, the semi-elasticities have been divided by a factor $100/\mathrm{Cd}$, after which they can be interpreted as the percentage changes in the median price due to a 1% increase in Cd pollution at each evaluation point. Despite the fact that the coefficient on the linear term is positive (0.260), the elasticity value is zero at Cd = 1.43 ppm, and becomes negative thereafter. In unreported work, we found that other (linear) specifications including Cd or log(Cd) yielded insignificant results with unexpected signs. Since the estimated coefficients associated with all other covariates are barely affected, the table shows only the coefficients on the alternative Cd-pollution terms.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

Table 5 Robustness of UQR results across different model specifications (for median of price distribution)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Public sale	0.180***	0.186***	0.180***	0.190***	0.168***	0.181***	0.172***
	(0.054)	(0.055)	(0.055)	(0.051)	(0.052)	(0.051)	(0.052)
Log plot size	−0.037*	−0.042*	−0.037*	−0.022	−0.037*	−0.022	−0.027
	(0.022)	(0.023)	(0.022)	(0.021)	(0.022)	(0.020)	(0.019)
Built structures	0.193*	0.213**	0.195*	0.198**	0.210**	0.206**	0.195**
	(0.102)	(0.101)	(0.102)	(0.087)	(0.102)	(0.086)	(0.090)
Pasture land	−0.130***	−0.129***	−0.132***	−0.124***	−0.127***	−0.121***	−0.123***
	(0.043)	(0.044)	(0.044)	(0.045)	(0.042)	(0.040)	(0.042)
Residential zoning	0.026	0.076	0.027	0.049	0.015	0.041	0.061
	(0.081)	(0.080)	(0.082)	(0.077)	(0.075)	(0.080)	(0.080)
Nature/forest zoning	−0.128**	−0.132**	−0.130**	−0.116*	−0.150**	−0.128*	−0.117*
	(0.060)	(0.059)	(0.060)	(0.063)	(0.061)	(0.067)	(0.064)
Address density $\leq$ 1 km	0.224***		0.218***	0.133**	0.202***	0.124**	0.175***
	(0.065)		(0.066)	(0.062)	(0.067)	(0.063)	(0.066)
Log dist. to nearest local urban center		−0.086**					
		(0.042)					
Log distance to Dutch border	−0.093**	−0.089**	−0.088**	−0.100***	−0.109***	−0.096***	−0.048*
	(0.037)	(0.037)	(0.038)	(0.023)	(0.037)	(0.023)	(0.025)
Log spatiotemporal lag of dependent variable	0.086	0.082	0.086	0.098*	0.091	0.107*	0.099*
	(0.055)	(0.056)	(0.055)	(0.055)	(0.060)	(0.055)	(0.054)
$I(Cd > 2) \times [\log(Cd/2)]^2$	−1.381***	−1.328***	−1.421***	−1.172***	−1.394***	−1.098**	−0.987*
	(0.507)	(0.515)	(0.513)	(0.509)	(0.507)	(0.538)	(0.538)
Log distance to nearest pollution source			−0.038				
			(0.075)				
Proportion of cultivated land (municipality level)							0.451***
							(0.135)
Log prices of developable land (municipality level)							0.412***
							(0.097)
Municipality dummies	Yes	Yes	Yes	No	Yes	No	No
Year dummies	Yes	Yes	Yes	Yes	No	No	No
Number of observations	511	511	511	511	511	511	511

All coefficients reported in the table are for the median ($\tau = 0.50$). To ease comparison, column 1 is a replication of the results reported in column 6 of Table 2. Omitted (reference) categories are private sale, vacant, arable and other land, and agricultural zoning. The correlation between address density $\leq$ 1 km and the distance to the nearest urban (local town) center is -0.462, and the correlation between the Cd-pollution level and the distance to the nearest pollution source is −0.392.
*, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively.

Appendix A

Table A.1 Simulation results for OLS, CQR, and UQR – Estimated coefficients on *X*

OLS	CQR			UQR		
	Q10	Q50	Q90	Q10	Q50	Q90
−0.497 (0.007)	−0.390 (0.010)	−0.507 (0.008)	−0.597 (0.010)	−0.296 (0.016)	−0.634 (0.027)	−0.451 (0.032)
Wald-type *F* test (p value)	84.3 (0.000)	104.6 (0.000)		95.2 (0.000)	15.3 (0.000)	

Q10, Q50, and Q90 are the 10th, 50th, and 90th percentiles of the conditional (CQR) and unconditional (UQR) distributions of *Y*, respectively. Robust standard errors for OLS and bootstrapped standard errors for CQR and UQR (based on 500 replications) are given in parentheses. The resulting R^2 for OLS is 0.795. To test the null hypothesis of equality of the unconditional quantile effects, we first performed a bootstrap estimation (based on 500 replications), which was necessary to obtain the full covariance matrix of the estimated UQR coefficients (note that there is no in-built Stata command for testing the equality of the quantile coefficients available for UQR, as is the case with CQR).