Harris, P., Lanfanco, B., Lu, B. and Comber, A. 2020. Influence of Geographical Effects in Hedonic Pricing Models for Grass-Fed Cattle in Uruguay. *Agriculture.* 10, p. 299. https://doi.org/10.3390/agriculture10070299

Supplementary Materials for: Influence of Geographical Effects in Hedonic Pricing Models for Grass-Fed Cattle in Uruguay

S.1 Hedonic Models

Linear Regression

To maintain consistency with previous studies, the theoretical model developed by Lanfranco et al. [2] was used as a reference; this model adopted the factor market method of Ladd and Martin [5] within a hedonic price framework. Hedonic price models can be estimated by an ordinary least squares (OLS) linear regression (LR). The response was the price in dollars per kilogram (US$ kg^{-1}) of live weight recorded for each lot of cattle sold at auction. For the extended model of Lanfranco and Castaño [3], predictors included in matrix **x** represented overall market conditions at the time of the sale, marketing strategy at cattle auctions, cattle attributes included in the lots, and prevalent agro-ecological characteristics of the lot's police precinct of origin. The price function *y* **(x)** allows the inclusion of nonlinear relationships; thus, quadratic terms for several of the predictors are included.

Following Lanfranco and Castaño [3], for lot $i = 1, \ldots, N$, and given a set of *K* characteristics that describe it fully, the function *y* **(x)** is expressed as follows:

$$y_i = \varphi + \sum_{k=1}^{K} \vartheta_k x_{ik} + \sum_{k=1}^{K} \rho_k x_{ik}^2 + \sum_{i=1}^{K} \sum_{k=1}^{K} \varsigma_{lk} x_{il} x_{ik} + \varepsilon_i, \qquad l \neq k \tag{1}$$

where y_i is the response (price) variable and predictors x_{ik} and x_{ik}^2 represent linear and quadratic relationships of characteristic *k,* while the product $x_{il}x_{ik}$ is the potential interaction of the predictor

variable k with the predictor variable l, for lot N. The model has K linear relationships, K quadratic relationships, and ($K \times K$) - K interactions. Model coefficients φ, ϑ_k, ρ_k and ς_{lk} are estimated by OLS. Some interactions between predictors (e.g., breed × sex) were not considered, meaning that some ς_{lk} coefficients were set to zero. For OLS, residual error ε_i is assumed independent and identically distributed, $\varepsilon_i \sim N(0, \sigma^2 I)$. Differentiating equation (1) with respect to x_k gives the price (or implicit marginal value) for a given characteristic k. Substituting the estimated coefficients for parameters $\hat{\vartheta}_k$, $\hat{\rho}_k$ and $\hat{\varsigma}_{lk}$ and ignoring the lot subscript, provides the following expression that includes quadratic relationships and factor interactions:

$$y_k = \hat{\vartheta}_k + 2\hat{\rho}_k x_k + \hat{\varsigma}_{lk} x_l, \qquad l \neq k \text{ and } k = 1, \ldots, K. \tag{2}$$

The price or marginal values in equation (2) connect the reserve equilibrium prices, δ and θ, with characteristics that determine product quality, such that $\delta(\mathbf{x}) = y(\mathbf{x}) = \theta(\mathbf{x})$. This does not reveal information concerning inherent supply and demand functions. Corrections were also applied to mitigate against heteroscedasticity of the variance-covariance matrix [3] and were only applied for LR. In summary, the following LR model was investigated:

$$PRICE = f(Beef\ and\ Market\ Cond. + Auction\ Strat. + Cattle\ Attri. + Agroeco\ Conditions) \tag{3}$$

Incorporating Spatial Effects through an Autocorrelated Error Term

For spatial extensions, it is convenient to re-name the $K = 51$ coefficients φ, ϑ_k, ρ_k, and ς_{lk}, for the model in equation (3), to give a model with a K x 1 vector of coefficients, named $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \ldots, \beta_k]$. Thus, the regression in equation (1) can be re-written:

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \tag{4}$$

where in matrix terms, the coefficients $\boldsymbol{\beta}$ are estimated via OLS using:

$$\widehat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\mathbf{y} \tag{5}$$

where, $\mathbf{X}$ is a $\left(N \times (K+1)\right)$ predictor data matrix and $\mathbf{y}$ is a $(N \times 1)$ response data vector. Spatial autocorrelation in the error term can be accounted for, if present, by fitting a linear mixed model (LMM), where the coefficients are unbiasedly estimated using restricted maximum likelihood (REML) [59,31]:

$$\widehat{\boldsymbol{\beta}}_{REML} = (\mathbf{X}^{\mathrm{T}}[\boldsymbol{\Sigma}_{\Delta}]^{-\mathbf{1}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}[\boldsymbol{\Sigma}_{\Delta}]^{-\mathbf{1}}\mathbf{y} \tag{6}$$

and where $[\Sigma_{\Delta}]^{-1}$ represents unbiased variogram information of the (spatially-autocorrelated) residual process, $\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{REML}$. In this study, an exponential variogram model was chosen for this purpose, which can be defined as:

$$\gamma(h) = c_0 + c_1\left(1 - \exp\left(-\frac{h}{a}\right)\right) \quad (7)$$

where the parameters c_0 and c_1 represent the partial sills of the variogram and a is the correlation range; h is distance, assuming isotropy. Details for specifying variograms for an LMM can be found in Schabenberger and Gotway [16]. Failure to account for a spatially-autocorrelated error term in regression modelling are unreliable coefficient and uncertainty estimates from a naïve (OLS) LR fit. As with the LR model, an LMM was investigated following the same functional form given in equation (3). As this study is dealing with areal data, an alternative spatial regression could have been formulated using a simultaneous or conditional autoregressive model (i.e. a SAR or CAR model, respectively). This study's reporting of an R^2 value for the LMM, should be viewed cautiously, as the R^2 value (or coefficient of determination) is specifically designed for an OLS LR fit and therefore does not account for error variation due to residual spatial autocorrelation.

Incorporating Spatial Effects through Scale-dependent Relationships

For the second spatial model, a geographically weighted regression (GWR) model was used [17]. GWR is a spatially varying coefficient (SVC) model that investigates how relationships between the response and predictors may vary across space. It is underpinned by the idea that global or whole map statistical models such as an LR/LMM may make unreasonable stationary assumptions amongst the regression's coefficients under investigation. GWR provides a measure of process spatial heterogeneity in data relationships, where local coefficients (i.e. the SVCs) and associated measures of uncertainty (e.g. *t*-values) can be mapped to explore this [36,44].

In its standard form, GWR calculates a series of local regressions at target locations, using nearby weighted data falling under a kernel at the center of each location. Only a single kernel bandwidth is used, which is limiting in that it unrealistically assumes the same level of spatial smoothness for each set of SVCs. Thus, when some relationships operate at a large-scale while others operate at a small-scale, standard GWR will not capture these differences and only find a 'best-on-average' scale of relationship nonstationarity (as using only a single kernel bandwidth). As a first step to mitigate against this limitation, mixed GWR can be implemented in which some relationships are taken as stationary (*globally-fixed*) whilst others are taken as nonstationary (*locally-varying)* [18]. However, a mixed GWR model does not fully address the limitation, as the subset of relationships that are locally-varying are all still taken to operate at the same spatial scale. Instead, multiscale GWR (MGWR) can be used [32,33,27,47,34,48], in which each relationship is characterized through its own bandwidth. Thus, the scale of relationship nonstationarity is allowed to vary for each response to predictor relationship.

Following that given in Murakami et al. [48], a basic linear SVC model can be defined as:

$$y_i = \sum_{k=1}^{K} x_{ik}\beta_k(s_i) + \varepsilon_i, \qquad E[\varepsilon_i] = 0, \qquad Var[\varepsilon_i] = \sigma^2, \tag{8}$$

where $\beta_k(s_i)$ denotes the k-th SVC for site i. This local approach estimates coefficients at the i-th site, $\{\beta_1(s_i),\ldots\ \beta_k(s_i),\ldots\ \beta_K(s_i)\}$, where in the case of standard GWR, a weighted least squares

estimation is applied to nearby sub-samples that are weighted through some distance-decay function at site *i*. Supposing $\boldsymbol{\beta}(s_i) = [\beta_1(s_i),... \beta_k(s_i),... \beta_K(s_i)]^{\mathrm{T}}$, the standard GWR estimator gives:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{GWR}}(s_i) = [\mathbf{X}^{\mathrm{T}}\mathbf{W}(s_i)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(s_i)\mathbf{y} \tag{9}$$

where $\mathbf{W}(s_i)$ is an $N \times N$ diagonal matrix whose j-th element $g(s_i, s_j)$ represents the weight assigned to the j-th sample, and where, $g(s_i, s_j)$ is calculated by a kernel weighting function [56], such as a bi-square kernel:

$$g(s_i, s_j) = \begin{cases} 1 - \left(\dfrac{d(s_i, s_j)^2}{b^2}\right), d(s_i, s_j) < b \\ 0, \quad otherwise \end{cases} \tag{10}$$

where b denotes the bandwidth parameter and $d(s_i, s_j)$ is the distance between locations s_i and s_j. The SVCs from a standard GWR will tend to the constant coefficients of the OLS LR, if b is set sufficiently large enough; otherwise, the SVCs will be local. Bandwidths can be specified as a *fixed distance*, or an adaptive distance, where for the latter, bandwidths vary according to a *fixed local density* of sub-samples.

Standard GWR as described above, ignores differences of spatial scale across the SVCs, as the same (single) bandwidth is specified for all relationships. To counter this, each set of SVCs can be found using its own bandwidth, and thus extend GWR with multiple (or flexible) bandwidths, one for each relationship (i.e. MGWR). Here the bi-square kernel for MGWR is defined as:

$$g_k(s_i, s_j) = \begin{cases} 1 - \left(\frac{d(s_i, s_j)^2}{{b_k}^2} \right), d(s_i, s_j) < b_k \\ 0, \quad otherwise \end{cases} \tag{11}$$

where b_k is the (*fixed distance*) bandwidth for the k-th coefficient. The estimated SVCs describe a global-scale process provided b_k is set sufficiently large, and a local-scale process provided b_k is set sufficiently small. Standard GWR is estimated in a two-stage procedure where first b is optimally found through some measure of model fit (say, by maximizing the Akaike Information Criterion (AIC) [19]); and second, the SVCs are estimated by substituting the optimal value of b into equation (9). The SVCs of MGWR are estimated in a similar manner except that a back-fitting approach is used in the first stage, which sequentially iterates the calibration of b_k (or $b(s_i)_k^{ad}$) with the assumption that all bandwidth parameters are known. For this study, MGWR is calibrated using the bi-square function in equation (11) with fixed bandwidths optimally found via an AIC-based back-fitting approach.

As with the LR and the LMM, an MGWR was investigated following the same functional form given in equation (3). Here, the use of nonlinear quadratic terms in MGWR is unusual given nonstationary effects often mimic nonlinear effects but were retained for coherence and consistency between model forms. Standard or mixed GWR forms were not investigated. LR and LMM do not suffer from the problem of multiple hypothesis testing when determining coefficient significance (i.e. false discovery rates), whereas GWR models do. Thus, *p*-values from MGWR are adjusted for multiple comparisons as advocated in Fotheringham et al. [32]; Yu et al. [34].

S.2 Supplementary Results

Table S1. Summaries (part 1) for the LR model.

Description	Variable	Coefficient	Standard Error	*t*-statistic	*p*-value & significance	
Constant (intercept)	C	0.65939	0.11990	5.49960	0.00000	***
Steer price (US $ / kg)	PSTEER	0.67655	0.01488	45.47410	0.00000	***
Exch. rate (UY$/US$)	EXRT	-0.00810	0.00101	-8.04384	0.00000	***
Order of entry (#)	ORDER	-0.00040	0.00007	-5.92430	0.00000	***
ORDER quadratic (#)	ORDER2	0.00000	0.00000	7.05475	0.00000	***
Lot Size (#)	LOTSZ	0.00054	0.00019	2.88571	0.00394	**
LOTSZ quadratic (#)	LOTSZ2	0.00000	0.00000	-2.27386	0.02305	*
Recommended lot (Y/N)	RECOM	0.02954	0.00514	5.74496	0.00000	***
Males (Yes / No)	MALE	0.16724	0.00466	35.89716	0.00000	***
Live weight (kg)	KLW	-0.00166	0.00016	-10.71059	0.00000	***
KLW quadratic (kg^2)	KLW2	0.00000	0.00000	6.62661	0.00000	***
Class (scored 3 to 10)	CLASS	0.01121	0.00132	8.51169	0.00000	***
Condition (scored 3 to 10)	COND	0.00038	0.00440	0.08725	0.93048	
Age uniformity (Y/N)	AGEU	0.00610	0.00469	1.30013	0.19366	
Shape uniformity (Y/N)	UNIF	-0.00062	0.00430	-0.14393	0.88556	
Improved nutrition (Y/N)	INUT	-0.00449	0.00503	-0.89222	0.37235	
Tick area (Y/N)	TKAR	-0.00082	0.00378	-0.21789	0.82753	
Mio-Mio (Y/N)	BCAR	0.00395	0.00544	0.72557	0.46816	
Aberdeen Angus (Y/N)	BD2	0.01645	0.00448	3.66980	0.00025	***
Other British (Y/N)	BD3	-0.00078	0.00568	-0.13782	0.89039	
Continental breed (Y/N)	BD4	0.00304	0.00455	0.66879	0.50369	
Dairy breed (Y/N)	BD5	-0.01071	0.00461	-2.32213	0.02030	*
Zebu breed (Y/N)	BD6	0.00469	0.00637	0.73529	0.46223	
Hereford × Angus (Y/N)	CZ1	-0.02115	0.00505	-4.18866	0.00003	***
Brit. × Continental (Y/N)	CZ2	0.00026	0.00913	0.02815	0.97754	
Dairy × Zebu (Y/N)	CZ3	0.00059	0.00794	0.07459	0.94054	
Lot size × Live weight (kg)	LXW	0.00000	0.00000	-0.84324	0.39917	
Condition × Live weight (kg)	CXW	0.00002	0.00002	1.09228	0.27481	

Significance level (α) of *Student's* t_{12466} (two-tailed): *** 0.1%; ** 1%; * 5%, and ^ 10%, where the probability of committing a Type I error and statistical significance for four levels of α (0.1%, 1%, 5%, and 10%) are shown in the last two columns.

Table S2. Summaries (part 2) for the LR model.

Description	Variable	Coefficient	Standard Error	*t*-statistic	*p*-value & significance	
Soil productivity (#)	CONEAT	0.00018	0.00040	0.45099	0.65203	
CONEAT quadratic (#)	CONEAT2	0.00000	0.00000	0.42903	0.66793	
Water holding capacity (mm)	WHC	-0.00106	0.00037	-2.86619	0.00419	**
WHC quadratic (mm^2)	WHC2	0.00000	0.00000	2.06729	0.03880	*
Summer (Yes / No)	T1	-0.01377	0.06787	-0.20283	0.83929	
Fall (Yes / No)	T2	-0.12913	0.05592	-2.30929	0.02100	*
Winter (Yes / No)	T3	0.09679	0.07510	1.28878	0.19758	
NDVI (#)	NDVI	-0.01071	0.00341	-3.14370	0.00169	**
NDVI quadratic (#)	NDVI2	0.00010	0.00003	3.59105	0.00034	***
Surface water runoff (mm)	SWR	-0.00090	0.00041	-2.19622	0.02816	*
SWR quadratic (mm^2)	SWR2	0.00000	0.00000	-5.76850	0.00000	***
Available water (%)	PAW	0.00357	0.00120	2.96594	0.00304	**
PAW quadratic ($\%^2$)	PAW2	-0.00002	0.00001	-2.20301	0.02768	*
SWR × PAW	SXP	0.00002	0.00000	3.68608	0.00023	***
NDVI × Summer	NXT1	0.00405	0.00099	4.08890	0.00004	***
NDVI × Fall	NXT2	0.00255	0.00089	2.87043	0.00413	**
NDVI × Winter	NXT3	-0.00114	0.00102	-1.10968	0.26723	
SWR × Summer	SXT1	-0.00010	0.00017	-0.56159	0.57444	
SWR × Fall	SXT2	0.00051	0.00015	3.34261	0.00084	***
SWR × Winter	SXT3	-0.00013	0.00018	-0.72649	0.46760	
PAW × Summer	PXT1	-0.00318	0.00057	-5.55860	0.00000	***
PAW × Fall	PXT2	-0.00043	0.00047	-0.90305	0.36658	
PAW × Winter	PXT3	-0.00065	0.00049	-1.32628	0.18486	

Significance level (α) of *Student's* t_{12466} (two-tailed): *** 0.1%; ** 1%; * 5%, and ^ 10%, where the probability of committing a Type I error and statistical significance for four levels of α (0.1%, 1%, 5%, and 10%) are shown in the last two columns.

Table S3. Summaries (part 1) for the LMM.

Description	Variable	Coefficient	Standard Error	*t*-statistic	*p*-value & significance	
Constant (intercept)	C	0.70575	0.12386	5.69778	0.00000	***
Steer price (US $ / kg)	PSTEER	0.68200	0.01078	63.28919	0.00000	***
Exch. rate (UY$/US$)	EXRT	-0.00767	0.00102	-7.48880	0.00000	***
Order of entry (#)	ORDER	-0.00040	0.00007	-5.83070	0.00000	***
ORDER quadratic (#)	ORDER2	0.00000	0.00000	6.41632	0.00000	***
Lot Size (#)	LOTSZ	0.00057	0.00017	3.26219	0.00112	**
LOTSZ quadratic (#)	LOTSZ2	0.00000	0.00000	-2.10862	0.03507	*
Recommended lot (Y/N)	RECOM	0.02819	0.00513	5.49078	0.00000	***
Males (Yes / No)	MALE	0.16732	0.00475	35.19650	0.00000	***
Live weight (kg)	KLW	-0.00165	0.00013	-12.94867	0.00000	***
KLW quadratic (kg^2)	KLW2	0.00000	0.00000	7.90443	0.00000	***
Class (scored 3 to 10)	CLASS	0.01066	0.00147	7.23449	0.00000	***
Condition (scored 3 to 10)	COND	-0.00090	0.00407	-0.22048	0.82552	
Age uniformity (Y/N)	AGEU	0.00514	0.00419	1.22771	0.21966	
Shape uniformity (Y/N)	UNIF	0.00025	0.00432	0.05707	0.95449	
Improved nutrition (Y/N)	INUT	-0.00179	0.00471	-0.38006	0.70393	
Tick area (Y/N)	TKAR	0.00075	0.00449	0.16703	0.86736	
Mio-Mio (Y/N)	BCAR	-0.00101	0.00598	-0.16856	0.86616	
Aberdeen Angus (Y/N)	BD2	0.01665	0.00434	3.83338	0.00013	***
Other British (Y/N)	BD3	-0.00140	0.00607	-0.22995	0.81815	
Continental breed (Y/N)	BD4	0.00235	0.00466	0.50311	0.61493	
Dairy breed (Y/N)	BD5	-0.01086	0.00454	-2.39410	0.01673	*
Zebu breed (Y/N)	BD6	0.00410	0.00664	0.61813	0.53654	
Hereford × Angus (Y/N)	CZ1	-0.01845	0.00563	-3.27863	0.00106	**
Brit. × Continental (Y/N)	CZ2	0.00007	0.00969	0.00718	0.99427	
Dairy × Zebu (Y/N)	CZ3	-0.00015	0.00865	-0.01724	0.98625	
Lot size × Live weight (kg)	LXW	0.00000	0.00000	-1.14491	0.25235	
Condition × Live weight (kg)	CXW	0.00002	0.00001	1.25120	0.21097	

Significance level (α) of *Student's* t_{12466} (two-tailed): *** 0.1%; ** 1%; * 5%, and ^ 10%, where the probability of committing a Type I error and statistical significance for four levels of α (0.1%, 1%, 5%, and 10%) are shown in the last two columns.

Table S4. Summaries (part 2) for the LMM.

Description	Variable	Coefficient	Standard Error	*t*-statistic	*p*-value & significance	
Soil productivity (#)	CONEAT	0.00026	0.00053	0.49327	0.62186	
CONEAT quadratic (#)	CONEAT2	0.00000	0.00000	0.24961	0.80291	
Water holding capacity (mm)	WHC	-0.00128	0.00051	-2.49462	0.01267	*
WHC quadratic (mm^2)	WHC2	0.00000	0.00000	1.97112	0.04881	*
Summer (Yes / No)	T1	-0.01178	0.06479	-0.18185	0.85571	
Fall (Yes / No)	T2	-0.13940	0.05625	-2.47835	0.01326	*
Winter (Yes / No)	T3	0.05952	0.06936	0.85818	0.39086	
NDVI (#)	NDVI	-0.01181	0.00345	-3.42363	0.00063	***
NDVI quadratic (#)	NDVI2	0.00011	0.00003	3.83767	0.00013	***
Surface water runoff (mm)	SWR	-0.00080	0.00047	-1.70119	0.08902	^
SWR quadratic (mm^2)	SWR2	0.00000	0.00000	-5.24602	0.00000	***
Available water (%)	PAW	0.00296	0.00128	2.30118	0.02145	*
PAW quadratic ($\%^2$)	PAW2	-0.00001	0.00001	-1.53039	0.12603	
SWR × PAW	SXP	0.00002	0.00001	3.08254	0.00207	**
NDVI × Summer	NXT1	0.00390	0.00088	4.44173	0.00001	***
NDVI × Fall	NXT2	0.00276	0.00080	3.42638	0.00062	***
NDVI × Winter	NXT3	-0.00071	0.00090	-0.79261	0.42807	
SWR × Summer	SXT1	-0.00013	0.00021	-0.59719	0.55043	
SWR × Fall	SXT2	0.00050	0.00015	3.24618	0.00118	**
SWR × Winter	SXT3	-0.00016	0.00019	-0.83109	0.40600	
PAW × Summer	PXT1	-0.00303	0.00065	-4.67993	0.00000	***
PAW × Fall	PXT2	-0.00045	0.00051	-0.87510	0.38159	
PAW × Winter	PXT3	-0.00049	0.00050	-0.97798	0.32817	

Significance level (α) of *Student's* t_{12466} (two-tailed): *** 0.1%; ** 1%; * 5%, and ^ 10%, where the probability of committing a Type I error and statistical significance for four levels of α (0.1%, 1%, 5%, and 10%) are shown in the last two columns.

Table S5. Estimated stationary coefficients for MGWR (medians from MGWR outputs).

Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
C	-	BCAR	-0.00812	T3	-
PSTEER	-	BD2	-	NDVI	-
EXRT*	-0.00817	BD3	-0.00200	NDVI2*	0.00010
ORDER*	-0.00040	BD4	-	SWR	-0.00083
ORDER2*	0.00000	BD5*	-0.00951	SWR2*	0.00000
LOTSZ*	0.00045	BD6	-	PAW*	0.00311
LOTSZ2	-0.00000	CZ1*	-0.01270	PAW2	-0.00001
RECOM*	0.02548	CZ2	-	SXP*	0.00002
MALE	-	CZ3	-0.00270	NXT1*	0.00271
KLW	-	LXW	-0.00000	NXT2*	0.00181
KLW2*	0.00000	CXW	-0.00002	NXT3	-0.00115
CLASS*	0.00889	CONEAT	0.00014	SXT1	-0.00011
COND	-0.00310	CONEAT2	0.00000	SXT2	-
AGEU	0.00177	WHC*	-0.00078	SXT3	-
UNIF	-	WHC2	0.00000	PXT1*	-0.00298
INUT	-	T1	-	PXT2	-0.00042
TKAR	0.00022	T2	-0.07877	PXT3	-0.00003

Entries highlighted in blue are nonstationary (see main text). * Significant at the 5% level or lower.

S.3 Data Preparation

In section 2.3 of the main text, a preferential sampling issue was resolved by taking a stratified random sample from the full database used in previous studies [2,3], to yield a study dataset of N = 2,845 observations. Here, a more sophisticated weighted approach via spatial kernel density estimation could have been adopted similar to that proposed in Diggle et al. [35]. This study's pragmatic approach to preferential sampling is still subject to bias, but not as great as that would be found using the full dataset. Further, a sensitivity analysis on the effects of sampling variation could have been achieved via a parametric bootstrap but would have been computationally expensive. It would also have been possible to assess the *out-of-sample* prediction of accuracy of the study models using the larger set-aside dataset (the LMM then becomes a regression kriging model). As previous studies used the full database, then some differences would be expected between previous LR results and this study's LR results, as reported above.