

## **Appendices**

**Appendix A1 - Table: Descriptive statistics of the dummy research variables (N = 22,463). WBD stands for Wild-Boars Density**

<b>Variable</b>	<b>Count</b>	<b>Relative frequency (%)</b>	<b>Cumulative frequency (%)</b>
Neighborhood 1	535	2.38%	2.38%
Neighborhood 2	4,150	18.47%	20.86%
Neighborhood 3	4,745	21.12%	41.98%
Neighborhood 4	2,652	11.81%	53.79%
Neighborhood 5	3,925	17.47%	71.26%
Neighborhood 6	2,218	9.87%	81.13%
Neighborhood 7	4,238	18.87%	100.00%
WBD, 1 <sup>st</sup> quantile, 100 m bandwidth	22,061	98.21%	98.21%
WBD, 2 <sup>nd</sup> quantile, 100 m bandwidth	352	1.57%	99.78%
WBD, 3 <sup>rd</sup> quantile, 100 m bandwidth	27	0.12%	99.90%
WBD, 4 <sup>th</sup> quantile, 100 m bandwidth	18	0.08%	99.98%
WBD, 5 <sup>th</sup> quantile, 100 m bandwidth	5	0.02%	100.00%
WBD, 1 <sup>st</sup> quantile, 300 m bandwidth	18,738	83.42%	83.42%
WBD, 2 <sup>nd</sup> quantile, 300 m bandwidth	3,048	13.57%	96.99%
WBD, 3 <sup>rd</sup> quantile, 300 m bandwidth	461	2.05%	99.04%
WBD, 4 <sup>th</sup> quantile, 300 m bandwidth	173	0.77%	99.81%
WBD, 5 <sup>th</sup> quantile, 300 m bandwidth	43	0.19%	100.00%
WBD, 1st quantile, 500 m bandwidth	15,563	69.28%	69.28%
WBD, 2nd quantile, 500 m bandwidth	4,646	20.68%	89.97%
WBD, 3rd quantile, 500 m bandwidth	1,402	6.24%	96.21%
WBD, 4th quantile, 500 m bandwidth	668	2.97%	99.18%
WBD, 5th quantile, 500 m bandwidth	184	0.82%	100.00%
WBD, 1st quantile, 700 m bandwidth	14,299	63.66%	63.66%
WBD, 2nd quantile, 700 m bandwidth	5,145	22.90%	86.56%
WBD, 3rd quantile, 700 m bandwidth	1,398	6.22%	92.78%
WBD, 4th quantile, 700 m bandwidth	1,266	5.64%	98.42%
WBD, 5th quantile, 700 m bandwidth	355	1.58%	100.00%

**Appendix A2 – Table: Neighborhood characteristics**

<b>Number in the model</b>	<b>Name</b>	<b>Municipal code</b>	<b>Area (as % of the total municipal area)</b>	<b>Population (as % of the total municipal population)</b>	<b>Socio- Economic level</b>
1	Lower city	3	3.0 %	4.3 %	Low
2	West Haifa	4	15.1 %	14.8 %	Medium- Low
3	Carmel	5	12.2 %	17.1 %	High
4	Hadar	6	4.4 %	13.8 %	Medium
5	Neve Shaanan - Izreelya	7	4.7 %	13.7 %	Medium- High
6	Ramot Neve Shaanan	8	5.9 %	7.3 %	High
7	Ramot Hacarmel	9	16.5 %	15 %	Very high

**Appendix A3 - Table: Distribution of the real estate transactions among neighborhoods**

<b>Neighborhood number</b>	<b>Number of transactions</b>
1	535
2	4,150
3	4,745
4	2,652
5	3,925
6	2,218
7	4,238
<b>Total</b>	<b>22,463</b>

**Appendix A4 - Table: Distribution of the real estate transactions among quantiles of the different wild-boars density surfaces**

<b>Quantile</b>	<b>Number of transactions (KD bandwidth = 100 m)</b>	<b>Number of transactions (KD bandwidth = 300 m)</b>	<b>Number of transactions (KD bandwidth = 500 m)</b>	<b>Number of transactions (KD bandwidth = 700 m)</b>
1 <sup>st</sup> (0.0 to 0.2)	22,061	18,738	15,563	14,299
2 <sup>nd</sup> (0.2 to 0.4)	352	3,048	4,646	5,145
3 <sup>rd</sup> (0.4 to 0.6)	27	461	1,402	1,398
4 <sup>th</sup> (0.6 to 0.8)	18	173	668	1,266
5 <sup>th</sup> (0.8 to 1.0)	5	43	184	355
<b>Total</b>	<b>22,463</b>	<b>22,463</b>	<b>22,463</b>	<b>22,463</b>

**Appendix A5 - Table: Factors affecting the price per square meter using the boars-density calculated with bandwidths of 100, 300 and 700 meters, using the full dataset**

Predictors	Model 2 B (t) <sup>(a)</sup>	Model 3 B (t) <sup>(a)</sup>	Model 4 B (t) <sup>(a)</sup>
Constant	-1,666,819 (-101.65)***	-1,669,958 (-102.38)***	-1,679,293 (-103.88)***
Number of rooms	214 (8.71)***	231 (9.43)***	240 (9.89)***
Residence age	-31 (-22.29)***	-33 (-23.86)***	-35 (-25.25)***
Transaction year	836 (102.45)***	838 (103.23)***	843 (104.67)***
Neighborhood 1	-6,240 (-42.92)***	-5,955 (-40.77)***	-5,663 (-37.81)***
Neighborhood 2	-4,248 (-58.61)***	-4,050 (-55.31)***	-3,748 (-49.04)***
Neighborhood 3	500 (7.55)***	454 (6.84)***	545 (8.08)***
Neighborhood 4	-5,485 (-64.91)***	-5,204 (-60.18)***	-4,915 (-52.34)***
Neighborhood 5	-4,277 (-59.78)***	-4,037 (-55.17)***	-3,784 (-47.25)***
Neighborhood 6	-1,363 (-16.48)***	-1,232 (-14.90)***	-815 (-9.63)***
Wild-Boars Density, 1st quantile	-1,149 (-1.64)	-2,909 (-12.47)***	-2,208 (-22.06)***
Wild-Boars Density, 2nd quantile	760 (1.06)	-2,297 (-9.69)***	-2,336 (-24.03)***
Wild-Boars Density, 3rd quantile	1,847 (2.02)**	-887 (-3.28)***	-1,124 (-9.59)***
Wild-Boars Density, 5th quantile	571 (0.34)	-554 (-1.06)	351 (1.88)*
No of obs.	22,463	22,463	22,463
R <sup>2</sup>	0.5415	0.5463	0.5551
Adjusted R <sup>2</sup>	0.5413	0.5460	0.5548

**Notes:**

Model 2: Wild boars density calculated with bandwidth = 100 m.

Model 3: Wild boars density calculated with bandwidth = 300 m.

Model 4: Wild boars density calculated with bandwidth = 700 m.

(a) B and t statistic in the parentheses

(\*) Indicates a two-tailed 0.1 significance level

(\*\*) Indicates a two-tailed 0.05 significance level

(\*\*\*) Indicates a two-tailed 0.01 significance level

## **A6 - Note about the Bonferroni correction:**

When only one hypothesis is tested, the statistical significance is defined by comparing the p-value to a target significance level (the probability of type 1 error)  $\alpha$ , where  $\alpha$  is usually set to 0.05. It means, that once all model assumptions are true, the probability to reject a null hypothesis that there is no association between X and Y, when this null hypothesis is true, is at most  $\alpha$ . In other words, while it is always possible to make a false discovery while performing statistical test, the probability for such lack of fortune is bounded by  $\alpha$ .

However, when more than one hypothesis is tested, things become more complicated. A naïve comparing of every single p-value to  $\alpha$  can lead to much bigger probability of a false discovery than  $\alpha$ . For example, if a multiple regression model is fitted, many t-tests are performed simultaneously, each for a separate variable. If each and every p-value is compared to a target significance level, say  $\alpha=0.05$ , then the probability of at least one null hypothesis false rejection, i.e., the "discovery" of at least one un-existing association between variables is not bounded by  $\alpha$ , but rather by much bigger probability (which equals 1 on the extreme). In fact, the more hypotheses are tested, the bigger chance to declare at least one (and sometimes many more than one) such un-existing association (see, for example in Bretz et al, 2011).

To address this issue, different adjustment procedures were developed, each has advantages and disadvantages. Possibly the most widely known procedure is the Bonferroni correction (i.e., adjustment). The main advantage of this procedure is its simplicity: all is needed to compare every single p-value to  $\alpha/m$  instead of  $\alpha$ , where  $m$  is a number of tests performed. Acting this way will guarantee the probability of at least one false discovery be at most  $\alpha$ . For example, if one performs five statistical hypotheses tests and the purpose is to bound the probability of at least one false discovery by 0.05, then, according to Bonferroni adjustment, every single p-value should be compared to  $0.05/5=0.01$ . The disadvantage of this procedure is its well-known conservatism in most situations, meaning that too many true associations may not be discovered (a usual trade-off between Type 1 and 2 error probabilities for a fixed sample size). However, once a discovery is made by applying the Bonferroni

correction, one can be sure that the same discovery would have been made by any other adjustment procedure .

In this context, the regression model we fitted (see above) includes  $m=13$  explanatory variables. To apply Bonferroni correction with a target level of  $\alpha=0.05$ , we need to compare each corresponding p-value to  $\alpha'=\alpha/m=0.003846$ . As can be seen from table with the regression output, all-individual p-value are indeed smaller than  $\alpha'$ , hence all explanatory variables may be considered as significant.