Bocconi

IEFE

Centre for Research on Energy and Environmental Economics and Policy

Working Paper Series - ISSN 1973-0381

# A note on asymmetries in heating degree-days and natural gas consumption dependence structure. An Archimedean copula framework on the Italian system

*Alessandro Fiorini and Antonio Sileo*

**Working Paper n. 42**

March 2011

***IEFE - The Center for Research on Energy and Environmental Economics and Policy at Bocconi University***

***via Guglielmo Röntgen, 1 - 20136 Milano***

***tel. 02.5836.3820 - fax 02.5836.3890***

**www.iefe.unibocconi.it – iefe@unibocconi.it**

# A note on asymmetries in heating degree-days and natural gas consumption dependence structure. An Archimedean copula framework on the Italian system*

*Alessandro Fiorini*† *Antonio Sileo*‡

**Abstract**: An important tool in order to carry out research for modeling and managing energy demand and supply is explaining variables with measures reflecting weather variations. For example, handlings with heating degree-days represent an easy way to account for almost 100% of natural gas consumption variations. Setting-up a linear model is a standard way to proceed but when the data under exam is other than linearly associated, that's to say jointly Normal distributed, correlation is no longer a measure of dependence and interpreting it as such is both theoretically and practically erroneous. In this context, statistical theory proposes a powerful tool, named copula function, to model flexible multivariate distributions in order to describe alternative dependence structures with respect to standard ones. The aim of the paper is to check whether heating degree-days are a consistent linear predictor for natural gas consumptions. In this context, a case study is developed on a monthly average of heating degree-days and monthly (residential and total) natural gas consumption volumes. Estimation results on alternative Archimedean copulas confirm that there is not sufficient evidence supporting a symmetric association with respect to the range of value variables can jointly assume. The statistical model to be used in this type of analysis should be robust against deviation from joint Normality and nesting linearity as a special case.

**JEL Classification**: C14, Q49

**Keywords**: *Heating degree-days, natural gas consumption, copula function, Archimedean copulas*

---

* *The authors gratefully acknowledge helpful advice and comments from Francesca Foliano.*
† Research fellow; I-Com-Institute for Competitiveness (Rome). Corresponding author: fiorini@i-com.it
‡ Research fellow; IEFE – Centre for Research on Energy and Environmental Economics and Policy - Bocconi University (Milan)

## 1. Climate determinants: heating degree-days

Temperature fluctuation is a major determinant in energy economics. In literature, most analysis takes into account the influence of temperature measures in order to estimate in-sample and forecasted patterns of energy demand and supply. Valor et *al*. (2001) investigate in the power sector the relationship between electricity loads and daily air temperatures using a population-weighted temperature index. Among the other results, it accounts for a significant nonlinear relation. This confirms the approach adopted in Engle et *al*. (1992) where the authors describe an exogenous impact of weather, expressed in a nonlinear formulation. Another interesting case study is presented in Franco and Sanstad (2008), which estimate the relationships between temperature and electricity consumption and peak demand, with respect to a group of locations around California. Results are used to predict future impact of climate changes using projected regional climate changes. The analysis of Timmerman and Lamb (2007) is in line with a vast number of studies, focused on the relationship between (winter) temperatures and residential energy consumptions. Authors develop two different temperature measures and test the maximum correlation with natural gas residential consumption; the analysis includes distinct areas of the United States with respect to a monthly and a seasonal time scale. In Gallanti et *al*. (2006) natural gas demand scenarios are designed for the Italian system, processing heating degree days and consumption volumes. In particular, for any time frequency the authors derive a crucial measure that is a “gradient of consumption” built as the ratio between volumes and heating degree days.

Heating degree-days (HDD) represents a simple way to keep weather data in a time series framework. Roughly speaking, they easily express the concept of "volume of cold", that is defined as the net exposure of a land, in a specific time period, to air temperature variations over a standard level, evaluated as an amount of degrees. From a theoretical point of view, as well as fundamental in applicative works, HDD are largely defined and studied in many disciplines, spanning from civil and mechanical engineering (Samo and Beng; 1999, Büyükalaca et *al*.; 2001), environment and energy research and management (Sarak and Satman; 2003, Gallanti et *al*.; 2006, Timmer and Lamb; 2007). Regarding a geographic area, numerical values of HDD are given by sum, over a period of interest (month or year), of positive daily differences between a conventional referencing temperature and the average external temperature. Following that, by construction, low values on HDD specify short cooling periods and average daily temperatures very close to country’s benchmark. Conversely, high values indicate long periods of severe cooling conditions, denoted by average daily temperatures which are much lower than the conventional temperature reference.

In formula, heating degree days are defined by:

$$HDD = \sum_{i=1}^{N} (\bar{t} - t_i)^+ \qquad (1.1)$$

with:

- $N$: Length of the time horizon (number of days);
- $\bar{t}$: Conventional temperature.
- $t_i$: Average temperature.
- + Sign denote that the sum is over nonnegative values.

Values for $\bar{t}$ and $t_i$ are generally ruled by law. For what concerns $\bar{t}$, according to the Eurostat calculation method, the quantity is fixed in 18 Celsius degrees (18°C). The sum is over $t_i's$ lower than or equal to 15°C. Note that country-specific legislation can establish different values, like 15.5°C. In Italy the threshold is set at 20°C. Average temperature $t_i$ is given by half the distance between maximum and minimum outdoor temperature. In Italy, it is obtained by the mean between four distinct temperatures: maximum, minimum, the temperature read at 8 a.m. and the one reading at 7 p.m. The sum is over the whole season in which the average external temperatures are lower than 12°C.

It is then obvious that the temperature is a key determinant of natural gas consumptions; and this remark is certainly enforced by looking at residential components. However, it is not clear that the basic analytical set up should be a linear relation between natural gas consumptions and heating degree-days. At least, it would be tested. In fact, whenever running a simple OLS regression, standard approach assumes $Y$ (say, natural gas consumptions) and $X$ (say, heating degree-days) following a joint Normal distribution. That allows asymptotics on parameter estimators to work; but, as will be ruled out later, modeling as Normally shaped variates which exhibit non-Normal distributional behavior is more than just a theoretical mistake. It is well known how linear dependence, that is correlation, makes sense as an association structure only in a multivariate Normal framework. Out of Normality (univariate and consequently multivariate), correlation no longer represents a dependence structure and any measure based on correlation can lead to misleading results in term of association among variables. That's the case of Pearson's correlation, maybe the most employed measure of dependence in statistics.

This paper tests the common approach of explaining natural gas consumption through a linear relation with temperature as measured by heating degree-days. Evidence from the distributional characteristics of a dataset about Italy, lead to a preference of a more cautious approach that would be robust against deviation from Normality. In other words, an extension from linear relations to the concept of functional relations is required. This implies that it is possible to search for a more suitable way to model the structure of dependence which is the joint data generating process. This role can be played by copula functions. After discussing their general features, in the following section, it will be clear how, through a very simple framework, many interesting and more appropriate remarks can be made about dependence between HDD and natural gas consumption. This will be supported by an application. Finally general comments on policy implication for the Italian gas system will be presented.

## 2. Copula functions and limits of linear dependence

**I**. The birth of copula function dates very far in the past, but along the last two decades, their use increased in many fields of science; spanning from economics to engineering sciences (Genest and Rivest; 1993, Embrechts et *al*.; 2003, Cherubini, Luciano and Vecchiato; 2004, Dupuis; 2007). Joe (1997) and Nelsen (1999) give a complete overview on what copulas are and how they work:

***Definition* 1 *(Copula function)***. *Let* $\boldsymbol{I} = [0,1]$. *A 2-dimensional copula (2-copula) is a mapping* $C: \boldsymbol{I}^2 \to \boldsymbol{I}$ *such that:*
*a. For any* $(u, v) \in \boldsymbol{I}^2$ :

- $C(u, 0) = 0 = C(0, v)$
- $C(u, 1) = u; \; C(1, v) = v$

*b. For any* $u_1, u_2, v_1, v_2 \in \boldsymbol{I}$ *such that* $u_1 \le u_2; \; v_1 \le v_2$:

- $V_c([u_1, u_2] \times [v_1, v_2]) = C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \ge 0$

*denoting with* $V_C$ *the* $C$*-volume measured over the rectangle with vertices* $u_1, u_2, v_1, v_2$ ■

***Theorem* 1 *(Sklar's)***: *Let* $(X, Y)$ *be a bivariate real-valued random vector on a common probability space* $(\Omega, F, \wp)$, *with joint cumulative distribution function (cdf)* $H(x, y)$ *and continuous univariate margins* $F(x)$ *and* $G(y)$. *There exists a unique copula* $C$ *such that:*

$$H(x,y) = C(F(x), G(y)) \tag{2.1}$$

*given any* $x, y \in \boldsymbol{R}$ *(real axes)* ■

***Corollary 1***. *Given a 2-copula* $C(u,v)$, *the following equation holds*:

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)) \tag{2.2}$$

■

Given the continuity condition on marginals, because of Theorem 1 and Corollary 1, a joint cdf can be factorized in a unique representation in terms of copula. Conversely, if $C$ is a copula, and $F$ and $G$ are continuous univariate cdf's, $H$ it is a bivariate joint cdf. Therefore, given the copula structure of a certain joint cdf, large alternative distributions can be modeled just keeping $F$ and $G$ fixed and by varying the copula. Viceversa, identity (2.1) shows the possibility of isolating, at least theoretically, the underlying copula in any joint (continuous) cdf. Finally, noting that the probability integral transform defines a couple of Uniform (0,1) random variables, say *U* and *V*, can be observed that:

$$C(u,v) = \Pr(U \leq u, V \leq v) \tag{2.3}$$

which shows how a copula $C$ induces a probability measure on the metric space $\boldsymbol{I}^{\boldsymbol{2}}$, or simply, $C$ is nothing but a joint cdf with standard Uniform margins. Furthermore, given that (2.3) can be obtained from any guessed functional form for $F$ and $G$, it follows that copulas are not affected by "marginal behaviors", playing as a factor that completely and uniquely governs the dependence structure among variables.

***Definition* 2 *(Archimedean copula)***. *Let* $\varphi: I \to [0,\infty]$ *be a continuous strictly decreasing function such that* $\varphi(1) = 0$ and $\varphi^{[-1]}$ *the pseudo-inverse of* $\varphi$. *Further, let the function* $C: \boldsymbol{I}^2 \to \boldsymbol{I}$ *be given by the relation*:[1]

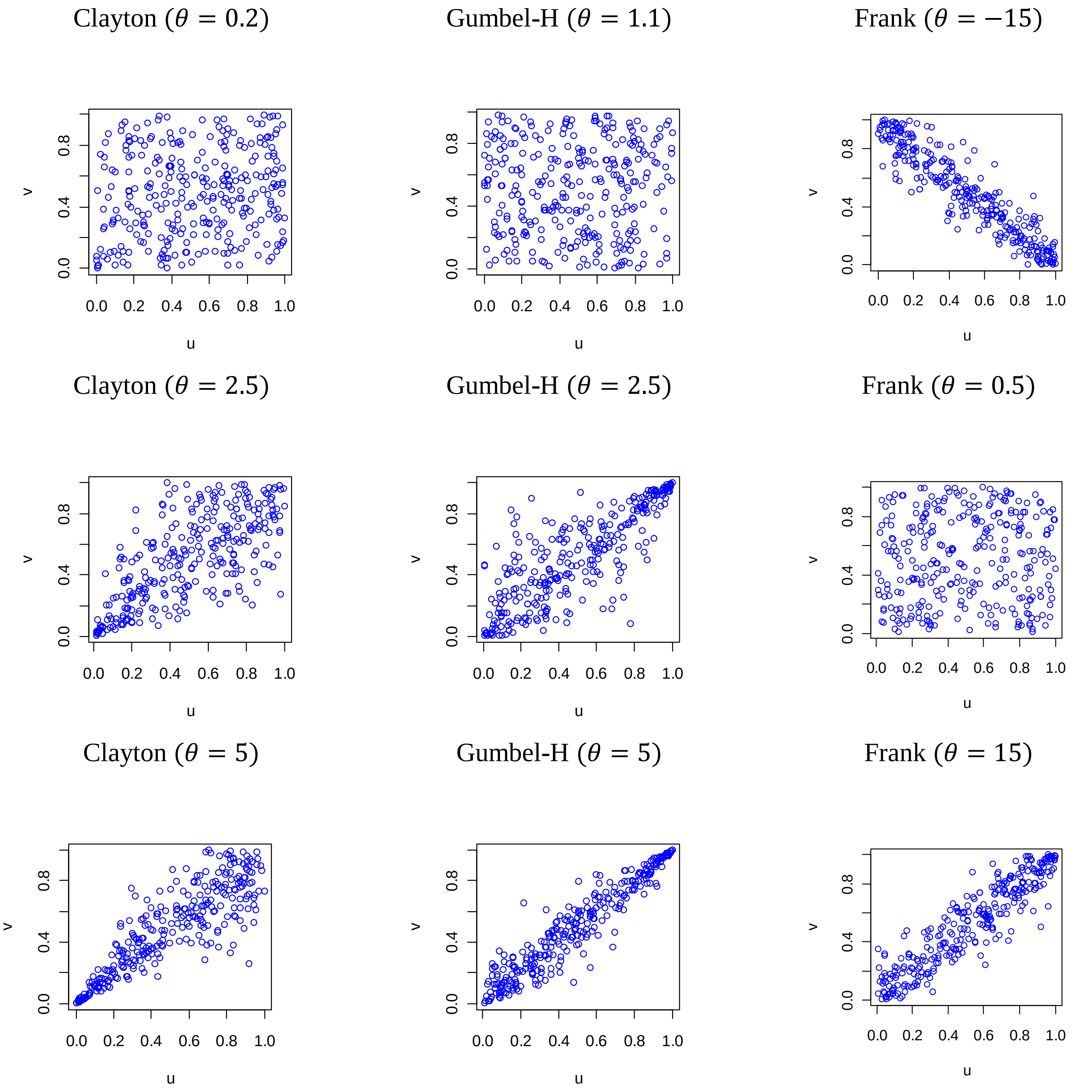
$$C(u,v) = \varphi^{[-1]}(\varphi(u), \varphi(v)) \tag{2.6}$$

*for any* $u, v \in \boldsymbol{I}$. *If* $\varphi$ *is a convex function,* $C$ *is a copula called Archimedean, while* $\varphi$ *is defined generator function of* $C$ ■

Notice that the parameters' value uniquely determines the shape of the generating function, as well as the functional form of the copula and characteristic of dependence it describes (Genest and Favre; 2007). A deeper focus on Archimedean copulas can be visually appreciated in Figure 2.1. Note that, as the value of the parameters increase towards positive values, copulas define a tighter positive association between data, reaching, asymptotically, the exact functional relation expressed by the identity $C_\theta = M(u,v)$. They show an analogue of behaviors for negative values, except for Gumbel-H which, by construction, describes only positive degrees of association.

[1] For $\varphi(0) = \infty$, implying $\varphi^{[-1]} = \varphi^{-1}$, then $\varphi$ is defined strict generator and $C$ is a strict Archimedean copula (Nelsen; 1999).

**Figure 2.1: Scatterplots of 300 bivariate observations generated from Archimedean copulas on $I^2$**




The shape of the generator functions also accounts for the independence copula. Finally, observe that far from independence, Gumbel-H copula stands as a framework for heavier dependence on upper tail values (upper-right quadrant), while lower tails in the case of Clayton (lower-left quadrant). Frank copula is symmetric going from upper to lower tail values.

***Theorem* 2**. *Let $C$ be an Archimedean copula with generator function $\varphi$ and let*:

$$K_C(t;\theta) = V_C(t;\theta)\left[\{(u,v) \in \boldsymbol{I}^2 : C_\theta(u,v) \le t\}\right]$$

*then for all $t \in \boldsymbol{I}$*:

$$K_C(t;\theta) = t - \frac{\varphi(t)}{\varphi'(t)} \tag{2.7}$$

■

Equation (2.7) is crucial because it shows how, with Archimedean copulas, any $n$-dimensional problem should be reduced to a univariate one, through a simple 1-dimensional cdf describing the random variable $(U,V)$. Efficacy will be clear in the evaluation of the dependence measure called Kendall's $\tau$, that will be discussed later in the paper.

***Corollary 2***. *If $C$ is an Archimedean copula with generator $\varphi$, $(U,V)$ a random vector with joint distribution $C$; then the function in (2.7) is the distribution function of the variable $C(U,V)$.* ■

**III**. Pearson's linear correlation coefficient, says $r$ is successfully employed as an indicator of dependence, mainly for its simple evaluation and scale-independence property. But its relevant drawbacks are well-known. Being a cross-product moment function, $r$ completely depends on marginal distributions. In the case of multivariate Normal distributions (elliptical as a general case), marginals, of any dimension, are always of the same form, up to a linear transformation. Thus, Pearson's correlation describes nothing more than what kind of linear relation links the two variables. The more variables deviate from the condition of linear linkage, the more $r$ loses the wider meaning of measure of dependence. In addition, it is a function of moments, and for some standard distributions $r$ it is forced to be undetermined or misleading. That's the case, whenever the absolute moment's existence condition is no longer verified.[2]

A solution can be found in nonparametric measures of concordance, such as, among the others, Kendall's $\tau$, Spearman's $\rho$ or Gini co-graduation index. In this context, for a matter of convenience, attention will be paid to the first one. Since it can be evaluated through the probability of concordance and discordance Nelsen (1999) states the following:

***Theorem 3***. *Let $X$ and $Y$ be continuous random variables whose copula is $C$. Then the population version of Kendall's $\tau$ for $X$ and $Y$ and is given by*:

$$\tau\,(X,Y) = 4\iint_{\boldsymbol{I}^2} C(u,v)\,\mathrm{d}C(u,v) - 1 \tag{2.8}$$

■

Thus notice, that when the underlying copula is Archimedean with generator $\varphi$ Genest and Rivest (1993) prove that:

$$\tau\,(X,Y) = 1 + \frac{\varphi(t)}{\varphi'(t)}\mathrm{d}t \tag{2.9}$$

Finally, handling with generators, Genest and Favre (2007) derive:

- Clayton: $\tau\,(X,Y) = \theta/(\theta+2)\,;$ $\quad\theta \geq -1$
- Gumbel: $\tau\,(X,Y) = (\theta-1)/\theta;$ $\quad\theta \geq 1$
- Frank: $\tau\,(X,Y) = 1 - 4/\theta + 4\,D_1(\theta)/\theta;$ $\quad\theta \in \boldsymbol{R}$

[2] See, among the others, Vitali (1999).

with: $D_1(\theta) = \int_0^\theta \frac{x}{\theta}(e^x - 1)^{-1}\,\mathrm{d}x$ (first order Debye function)

These relations are crucial in order to derive an estimation technique, which shares the rationale of a moment based procedure, independent from marginal distributions (Genest and Rivest; 1993).

In an inferential problem on a copula set-up, the starting point is to focus on a sample analog which shares the same invariance property. Pairs of ranks associated with any sample couple, $(R_1, S_1)\dots(R_n, S_n)$, meet that requisite (Genest and Favre; 2006). Note that by normalizing ranks on a factor $1/(n+1)$, one gets a set of points on $\boldsymbol{I}$ which form the domain of the empirical distribution counterparts of $F$ and $G$, say $F_n$ and $G_n$ That formally defines the empirical copula:

$$C_n(u, v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{\frac{R_i}{n+1} \le u, \frac{S_i}{n+1} \le v\right\} \qquad (2.10)$$

with $\mathbf{1}\{\cdot\}$ the indicator function. $C_n$ is the probability distribution that puts a mass $1/n$ on each point of the rank plot. Following the fact that inference for any copula-based measure of association mimics the same approach via the empirical copula $C_n$. In other terms, given that $\tau = f(C)$ it follows that $\tau_n = f(C_n)$. Genest and Favre (2007) show that under suitable regularity conditions $C_n \to C$ as $n$ increases, assuring that $\tau_n$ is an asymptotically unbiased estimator of $\tau$. Further, being $\tau_n$ a $U$-statistic, it is asymptotically Normal. These are the basic properties to derive, by inverting formulas as derived by (2.9), a robust and easy moment-like estimation procedure for Archimedean copulas' parameters as suggested in Genest and Rivest (1993). Whenever alternative Archimedean copula models have been estimated, goodness-of-fit testing should be run by the method based on the Kendall process discussed in Genest et *al*. (2009).

### 3. Dependence features for natural gas consumption and heating degree days

The data set is composed by two balanced time series for monthly average heating degree days and monthly natural gas residential and total consumption volumes, respectively from Eurostat and ENI – Snam Rete Gas. The sample size is 89 observations, covering the time frame from January/2003 to May/2010. For a deeper investigation of co-movements between temperature and natural gas consumption, it was split in two distinct samples representing colder and warmer months (say, "cold" and "warm" months). The first one was formed collecting observations from October to March, the second from April to September.

**Table 3.1: Dataset**

| *Variable* | *Symbol* |
|---|---|
| Heating degree-days (Celsius degree) | HDD |
| Residential natural gas consumption (G-$m^3$) | RES |
| Total natural gas consumption (G-$m^3$) | TOT |
| Heating degree-days – cold months (Celsius degree) | HDD-C |
| Residential natural gas consumption – cold months (G-$m^3$) | RES-C |
| Total natural gas consumption – cold months (G-$m^3$) | TOT-C |
| Heating degree-days – warm months (Celsius degree) | HDD-W |
| Residential natural gas consumption – warm months (G-$m^3$) | RES-W |
| Total natural gas consumption – warm months (G-$m^3$) | TOT-W |

**Figure 3.1: Dataset**

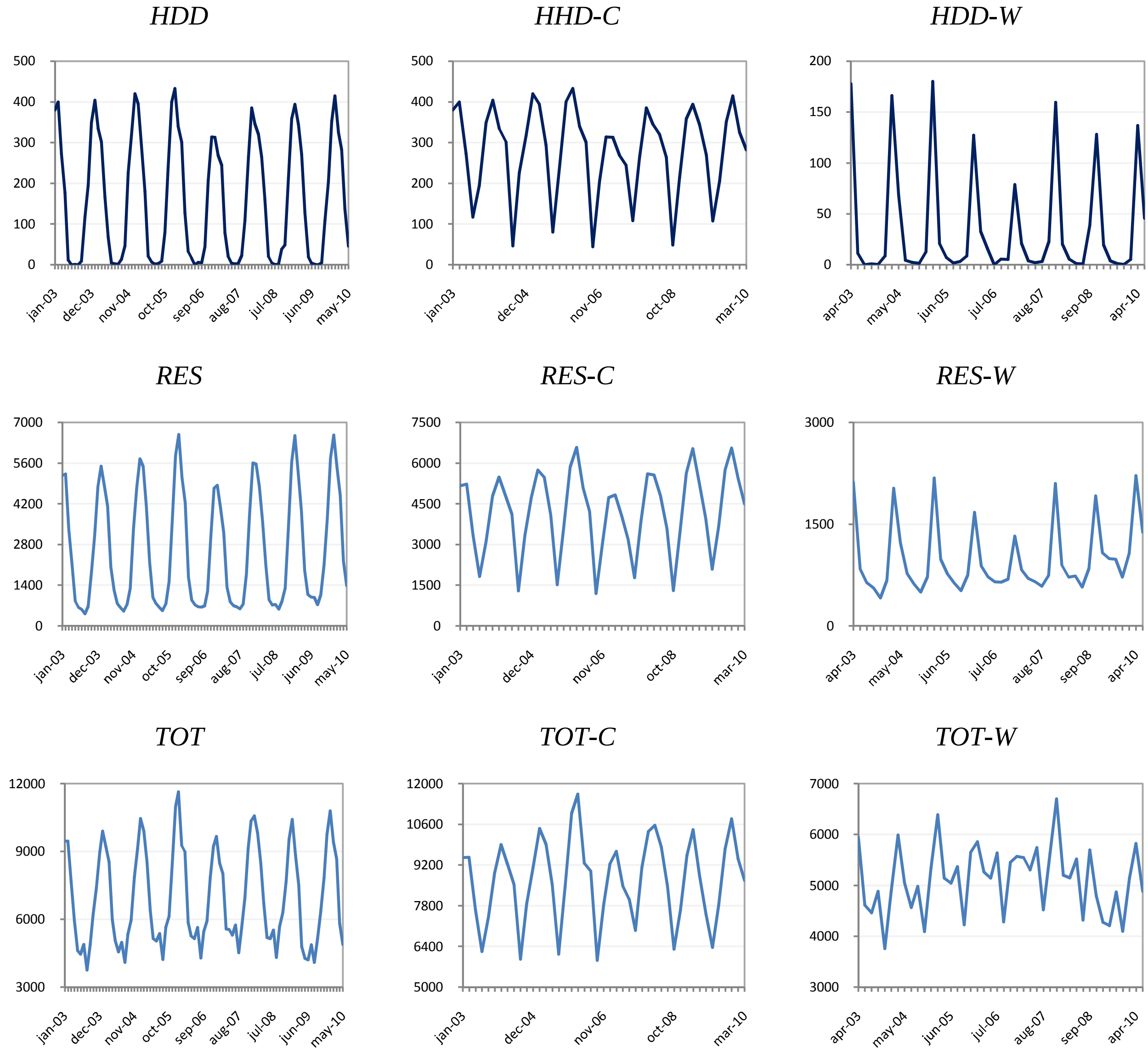



**Table 3.2: Summary statistics**

| | *Mean* | *Standard deviation* | *Kurtosis (excess)* | *Asymmetry* | *n* |
|---|---|---|---|---|---|
| HDD | 159.4930 | 150.0114 | -1.482149 | 0.3412381 | 89 |
| RES | 2620.387 | 1968.182 | -1.295883 | 0.4957636 | 89 |
| TOT | 6926.468 | 2118.985 | -1.224119 | 0.3946046 | 89 |
| HDD-C | 280.7531 | 107.6963 | -0.3721586 | -0.7540116 | 45 |
| RES-C | 4219.453 | 1484.572 | -0.6682637 | -0.4980789 | 45 |
| TOT-C | 8699.823 | 1434.825 | -0.6671762 | -02809184 | 45 |
| HDD-W | 35.47707 | 55.31497 | 1.024112 | 1.597304 | 44 |
| RES-W | 984.9786 | 509.6495 | 0.4587629 | 1.337684 | 44 |
| TOT-W | 5112.809 | 652.3233 | -0.5052863 | 0.05475134 | 44 |

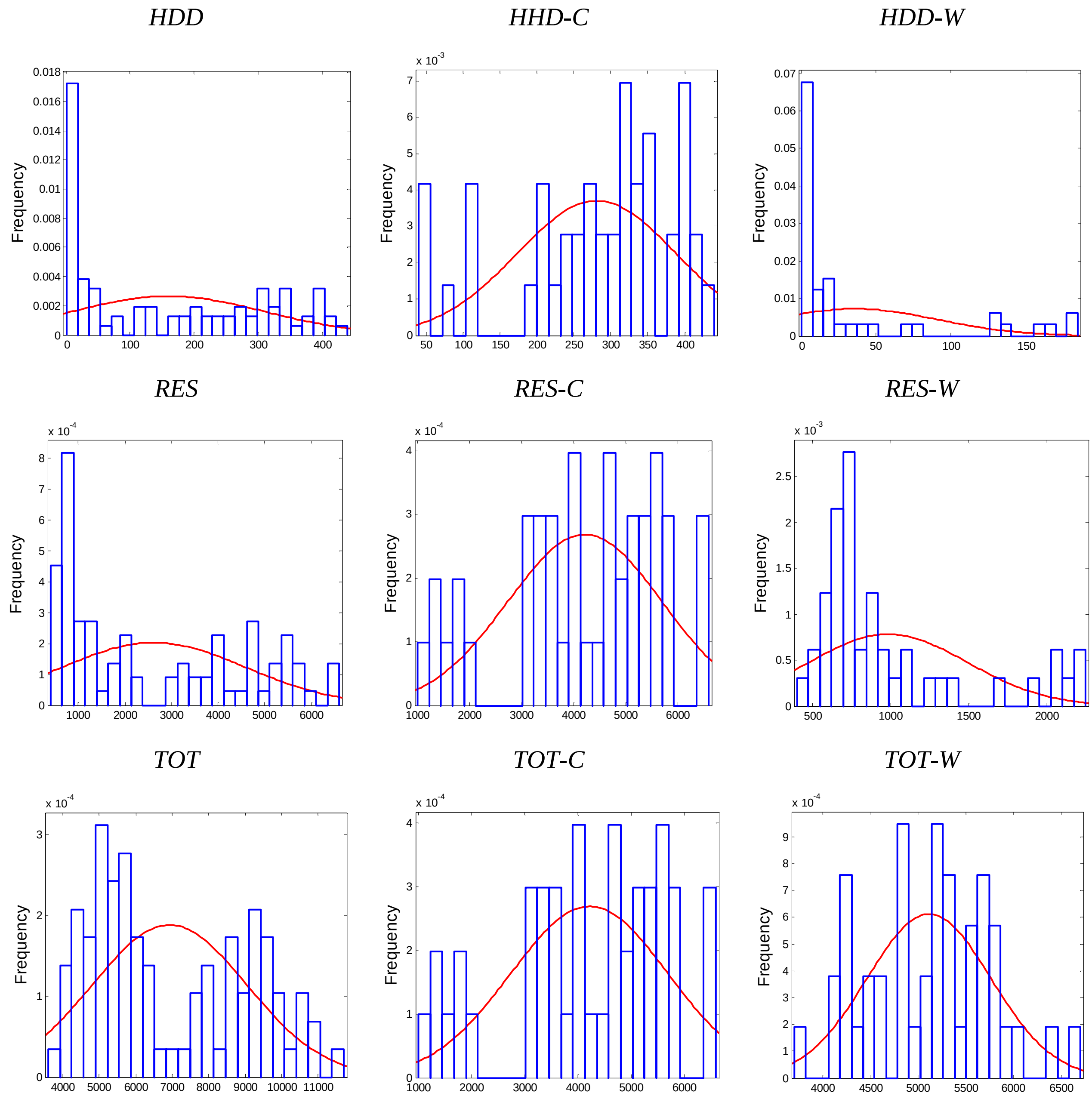
Summary statistics and graphics for each dataset are reported in Table 3.2 and Figure 3.1. It is certainly not surprising to check how data are affected by a prominent seasonal component. That's absolutely obvious thinking that temperature strictly depends on a natural sequence of seasons. Similarly, natural gas consumption, depending on temperature variations. Table 3.3 reports Pearson's correlation index estimated with respect to each couple "heating degree days/gas consumptions". The values confirm what intuition and literature suggests about a consistent and exceptionally strong correlation between such variables, above all in case of HDD vs RES. In addition, season-specific samples reflect the same structure of dependence, except for HDD vs TOT, for which correlation coefficient drops down to 0.6234. It can be argued that the effect of power generation and industrial consumption component, not properly temperature-related, significantly weakens the association between heating degree-days and total consumption whenever their relative weight tends to increase, particularly during warm months.

**Table 3.3: Pearson correlation index**

| | **Pearson's rho** | ***p*-value** |
|---|---|---|
| *Overall correlation* | | |
| HDD / Residential | 0.9874 | 1.7234e-071 |
| HDD / Total | 0.9643 | 5.4261e-052 |
| *Seasonal correlation* | | |
| HDD-C vs RES-C | 0.9729 | 5.5106e-029 |
| HDD-C vs TOT-C | 0.9463 | 1.0415e-022 |
| HDD-W vs RES-W | 0.9581 | 2.0045e-024 |
| HDD-W vs TOT-W | 0.6234 | 6.1859e-006 |

As we noticed earlier, when two (or more) variables are not jointly normal, Pearson's correlation fails to represent a measure of dependence. We can start by at least checking whether the data under examination can be assumed as drawn by a univariate Gaussian data generating process, recalling that a bivariate Normal has univariate Normal margins. Graphics in Figure 3.2 offer a first visual analysis of the distributional features of our data. Note that, in any case, Normal curves do not come out as an adequate fitting for the empirical distributions, both for lack of symmetries and for meso-kurtosis. This is confirmed by indexes reported in Table 3.2. Observe that a nonzero skewness coefficient denotes the presence of asymmetry, while the (excess of) kurtosis tails are thinner/fatter than the Gaussian distribution. For what concerns symmetries, distributions appear to be essentially right skewed in case of cold months and otherwise left skewed. Full time series reveal the prevalence of left-skewness for the unconditional distribution. This suggests how, in warm months, each variable showed lower variability over its range with respect to cold months, displaying a heavy mass of data points concentrated around lower values. In this case, that aspect can be essentially explained with the methodology described in (1.1) for HDD. For the measure of the "severity of coldness", the measure's support is mainly shifted towards cold temperatures. TOT-W can be considered as an exception, being not very far from a balanced shape around the mean. Also for what concern kurtosis, indexes suggest the presence of wider peaked distributions than the Gaussian. Exceptions emerge again from warm months: HDD-W and RES-W kurtosis index reveal higher probability (than a Normally distributed variable) to assume extreme values.

**Figure 3.2: Empirical distributions VS Normal theoretical density**




Symmetry and kurtosis indexes strictly depend on the sample's 2nd, 3rd and 4th order unconditional moments. Anyway, deviation from Normality exhibited by the data is strongly confirmed through statistical testing. First of all, in Figure 3.3, standard Normal quantiles are plotted against empirical quantiles for each variable (Normal QQ-plots). Notice that in any case, scatterplots tend to assume "anomalous" patterns in correspondence of upper and lower tails. It is extremely clear for HDD, RES and TOT. In HDD-W and RES-W that the distance is very large approaching the upper tail. Finally, in Table 3.4 two traditional Normality tests are performed. Regarding Jarque and Bera test, the null hypothesis of a Normal DGP is strongly rejected for HDD, RES, HDD-W and RES-W. In the case of TOT the null is weakly rejected (5%). For RES-C and TOT-C there is insufficient evidence to suggest the rejection of a Gaussian DGP. Distinct and clearer results are obtained by running a Kolmogorov-Smirnov test over a Gaussian benchmark. In every case, the null is not statistically supported.

**Figure 3.3: Normal QQ-Plots**

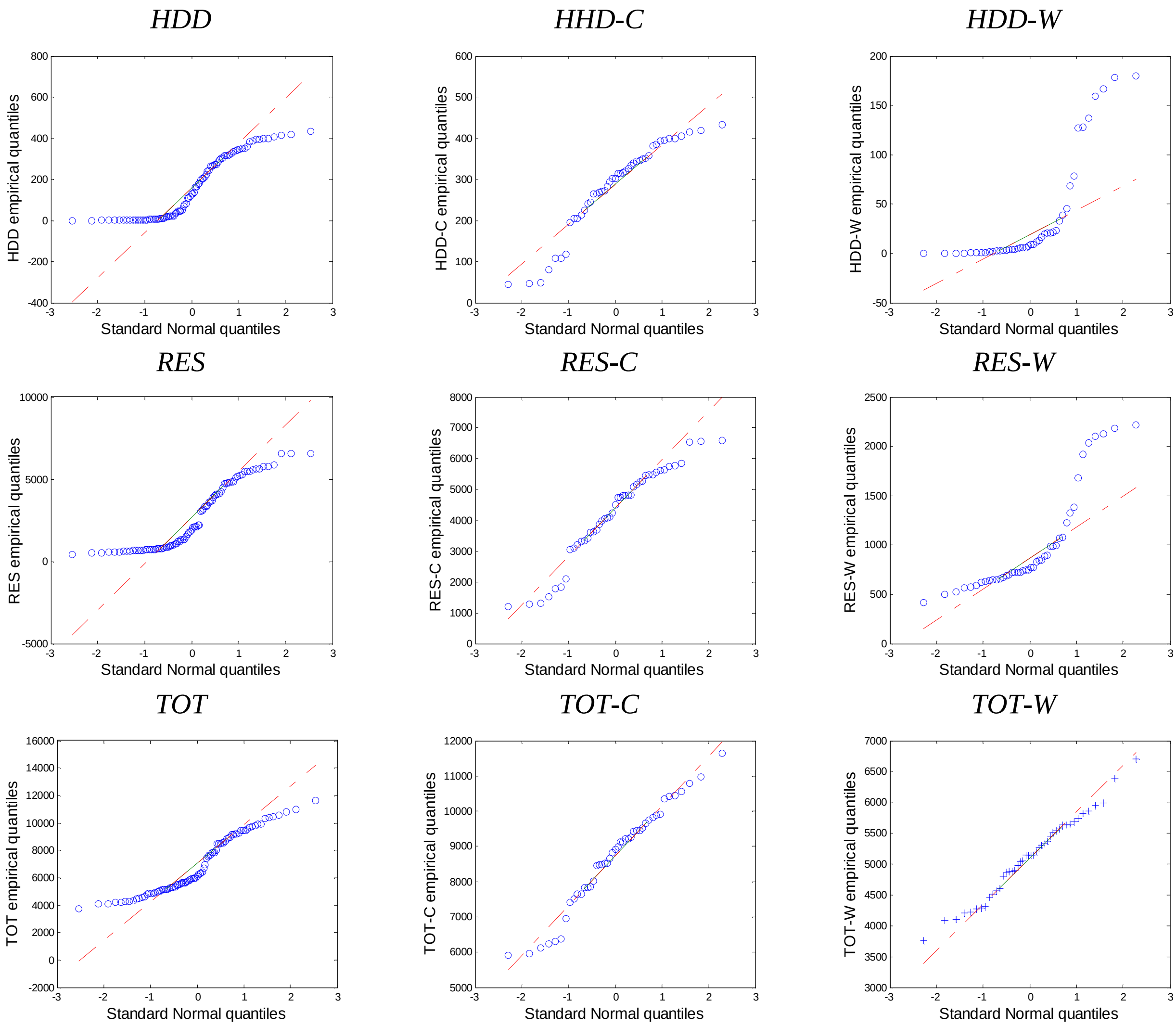
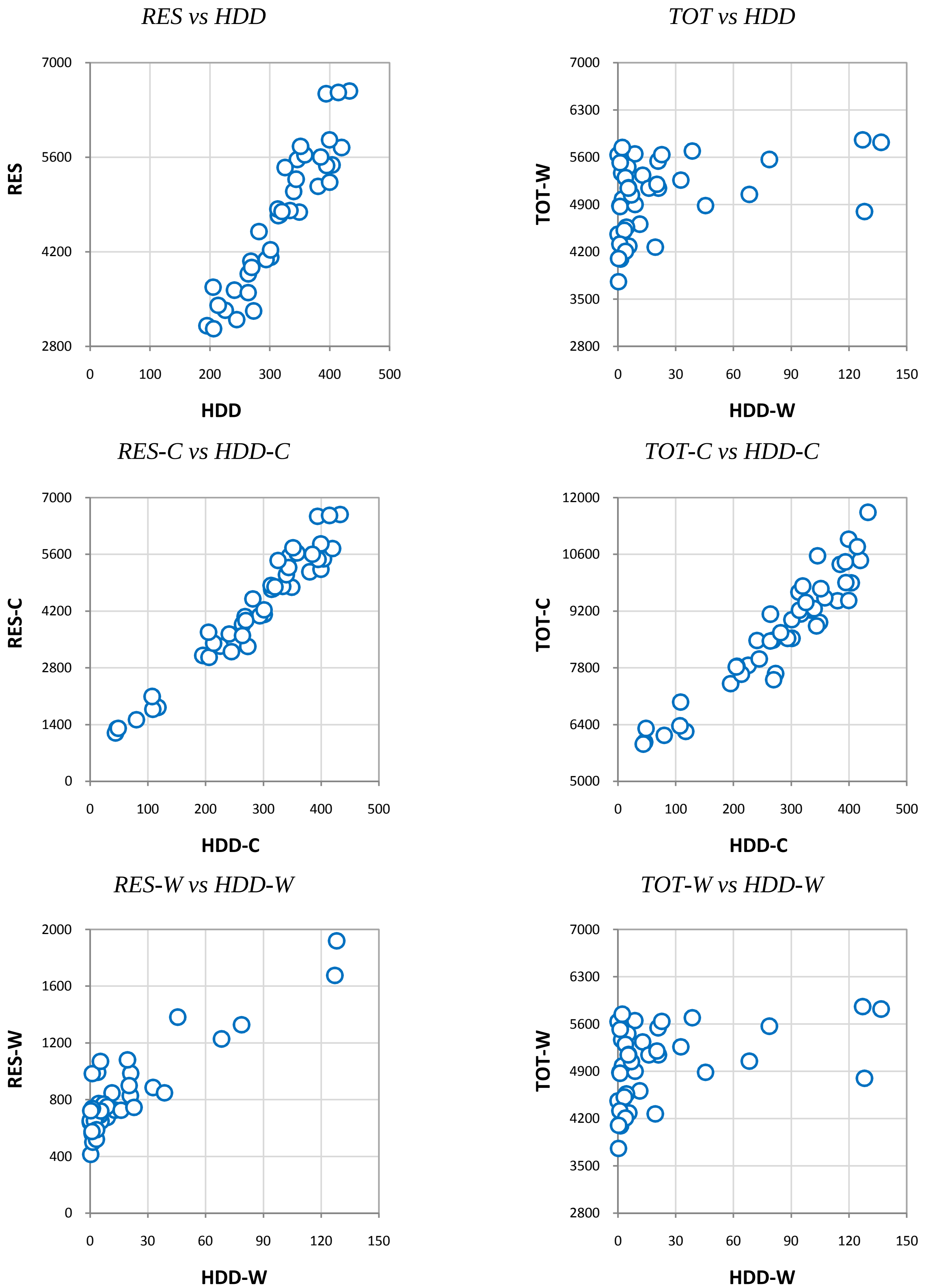



**Table 3.4: Normality tests**

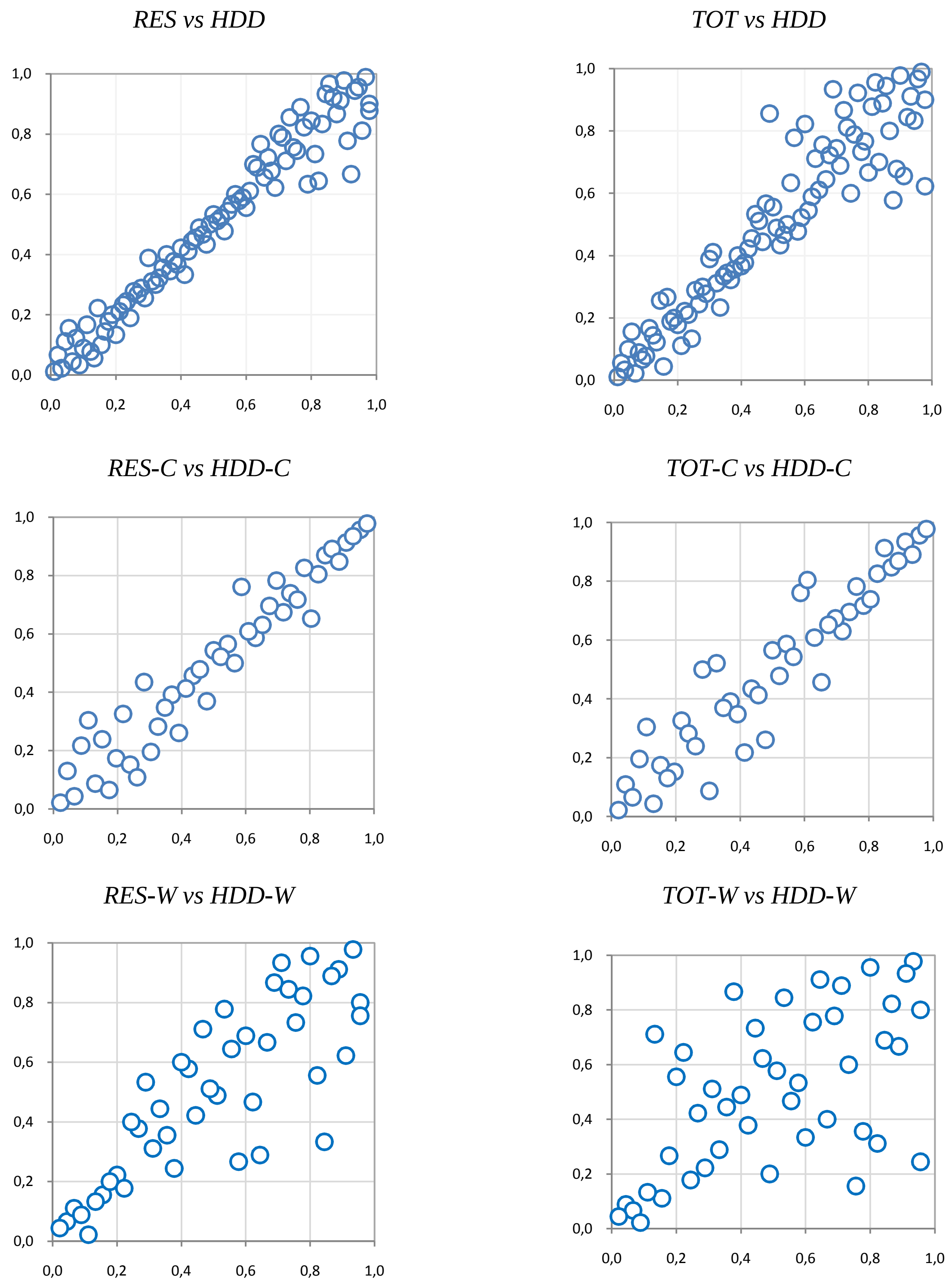
| | *Jarque-Bera Test ($H_0$: Normal)* | | *Kolmogorov-Smirnov Test ($H_0$: Normal)* | |
|---|---|---|---|---|
| | Test-statistic | *p*-value | Test-statistic | *p*-value (two-sided) |
| HDD | 9.5562 | 0.008412 | 0.8745 | 2.2e-16 |
| RES | 9.6302 | 0.008106 | 1.0000 | 1.665e-15 |
| TOT | 7.5838 | 0.02255 | 1.0000 | 1.665e-15 |
| HDD-C | 4.6798 | 0.09634 | 1.0000 | 6.661e-16 |
| RES-C | 2.5806 | 0.2752 | 1.0000 | 6.661e-16 |
| TOT-C | 1.221 | 0.5431 | 1.0000 | 6.661e-16 |
| HDD-W | 22.7456 | 1.150e-05 | 0.7680 | 2.2e-16 |
| RES-W | 14.7674 | 0.0006213 | 1.0000 | 6.661e-16 |
| TOT-W | 0.2994 | 0.861 | 1.0000 | 6.661e-16 |

**Figure 3.4: Scatterplots on real values**




Analysis provided so far suggests that a general explanatory model of natural gas consumptions against heating degree-days should, at least, nest linear relations as a special case. However, in general there is strong evidence confirming how conditional distribution of natural gas consumptions given heating degree-days, cannot be ruled out as bivariate Normal distributions describing a correlation-based structure of dependence between the couple. A flexible alternative, as explained, can be reached with a copula framework. In Figure 3.4 and 3.5 scatterplots on real axes and an empirical copula domain are compared.

**Figure 3.5: Scatterplots on standardized ranks values**




As can be seen in the latter, each sample shows distinct association profiles. In the case of full ones, data points seem to be strictly associated around average values in each domain, becoming sparser on upper-right quadrant values. That is even more evident in the case of TOT vs HDD. Concerning season-based samples, interesting comments can be made. Observe that in cold months, there seems to be a functional relation linking natural gas consumptions and heating degree-days characterized by a clustering on the upper-right side values, both for RES as for TOT; whereas in warm seasons the clustering concerns the lower left ones. In TOT-W vs HDD-W the cloud is definitively going to approximate an independent relation between variables. Estimation results and diagnostics reported in Table 3.5 support and strengthen the visual assessment: Frank copula is never the best model to fit unconditional bivariate distributions between natural gas consumptions and heating degree-days; that's to say the joint behavior between variables is nowhere as symmetric along the range of values but strictly depends on whether heating degree-days assume low or high values. Clayton was selected in case of both full and warm month's samples. This advises for an explanatory power of HDD for RES and TOT decreasing as values of variables increase. The opposite is true in the case of cold months, when natural gas consumptions and heating degree-days appear to be related by a functional relation that becomes tighter and tighter as values increase (Gumbel-H copula). This somehow confirms the aforementioned apparent prevalence of the warm month's variables joint behavior with respect to the whole year.

In conclusion, there is no reason to assume a linear relation linking heating degree-days and natural gas consumptions. Reduction of natural gas consumptions caused by a drop in the severity of coldness results to an asymmetricity with an increase in consumptions caused by an increase in the severity of coldness. It follows that a 1% of temperature variation does not always necessarily induce an *x*% in natural gas consumptions. It strictly depends on the season in which variations occur together with the magnitude of temperature rigidity. In particular, cold months seem to be characterized by a tighter relation between high volumes of consumptions and high severity. Alternatively, in warm periods, the more heating degree-days increase, the more natural gas consumption variations follow other determinants than just the temperature.

**Table 3.5: Copula models estimation**

| | Kendall-τ | *Clayton* | | *Gumbel-H* | | *Frank* | |
|---|---|---|---|---|---|---|---|
| | | Est. | GoF | Est. | GoF | Estimate | GoF |
| RES vs HDD | 0.8751 | 14.0128 | 3.4112e-4 | 8.0064 | 8.8510e-4 | 30.2862 | 4.4588e-4 |
| TOT vs HDD | 0.7893 | 7.4922 | 5.3773e-4 | 4.7461 | 0.0020 | 17.1615 | 0.0011 |
| RES-C vs HDD-C | 0.8505 | 11.3779 | 0.0014 | 6.6890 | 5.9242e-4 | 24.9950 | 6.3412e-4 |
| TOT-C vs HDD-C | 0.8040 | 8.2041 | 0.0016 | 5.1020 | 3.3331e-4 | 18.6037 | 5.7784e-4 |
| RES-W vs HDD-W | 0.6586 | 3.8582 | 6.4003-e4 | 2.9291 | 0.0022 | 9.7381 | 00.16 |
| TOT-W vs HDD-W | 0.4197 | 1.4465 | 0.0013 | 1.7232 | 0.0024 | 4.4389 | 0.0023 |

## 4. Closing remarks

A suitable procedure to explain energy variables through temperature variations needs to be based on a deeper examination of the structure of association describing the multivariate phenomenon. Given that, in any case the latter seems to not be a core issue, it follows in the literature we rarely encounter an alternative framework and then a linear representation. This is certainly intuitive and easy, but relevant mistakes are well known whenever some fundamental conditions do not hold. A crucial one is that linear relations arise in the context of multivariate Normal DGP's (elliptical in general). Otherwise, associations between variables should be analyzed under the wider framework of functional relations, because forcing the model to linear relations can produce misleading results. A simple way to avoid this mistake is to use copula modeling, whose advantages have been already discussed from a theoretical point of view. Even though they reached a strong popularity, copula functions are still a work in progress for which pros and cons have sparked a vivid debate.[3]

This paper focused on the nature of the relationship linking heating degree-days to natural gas consumption for the Italian energy system. Analysis pointed out that at least multinormality is not statically significant to describe the joint behavior of the couple. A set of a 1-parameter Archimedean copula have been tested: Clayton, Gumbel-H and Frank. The choice had been guided by a double reason: they provide a very simple but rigorous estimation procedure and represent each kind of all relevant association features. Estimation outcomes suggest how the nature of the relationship linking natural gas consumptions to heating degree-days essentially accounts for many sources of asymmetries. The main one is that intensity of positive association between natural gas volumes and heating degree-days strictly depends on whether temperature variations occur in cold or warm months. Cold months are likely to show a tighter association between variations on high values with respect to variations occurring on low values. The opposite happens in the case of a warm month. Of course, this evidence is stronger in the case of a residential withdrawal. It can be concluded that a 1% of temperature variation does not always necessarily induce an *x*% in natural gas consumptions. It strictly depends on the season in which variations occur together with a magnitude of temperature rigidity. In particular, cold months seem to be characterized by a tighter relation between high volumes of consumptions and high severity. Alternatively, in warm periods, the more heating degree-days increase, the more natural gas consumption variations follow other determinants rather than the temperature.

This can be considered a useful result in order to drive decision-making processes on a vast range of energy management branches; from pricing on the spot and forward markets to supply investment planning, in addition with supporting (indirectly) how improving interventions on the efficiency of domestic dwellings should be effective.

From a scientific point of view, the topic, not discussed in literature, can be subject to a future development. The highlighted "overall" copula dependence can be certainly extended in a time varying context, through a dynamic copula model (Patton; 2006). In this sense, the aim is investigating basic determinants of both short and long range relationships, trying to add key features for a more general theory and technique about temperature filtering on energy variables.

[3] See for example the large discussion launched by Mikosch (2006), which actually counts 10 replies and a rejoinder.

**References**:


BÜYÜKALACA, O.; BULUT, H.; YILMAZ, T. (2001): "Analysis of variable-base heating and cooling degree-days for Turkey", *Applied Energy*, n. 69 (4), 269-283

CHERUBINI, U.; LUCIANO, E.; VECCHIATO, W. (2004): *Copula Methods in Finance*; Wiley, Chichester.

DUPUIS, D.J. (2007): "Using copulas in hydrology: Benefits, cautions, and issues"; *Journal of Hydrologic Engineering*, vol. 12(4): 381–393.

EMBRECHTS, P.; LINDSKOG, F.; MCNEIL A. (2003): "Modelling Dependence with Copulas and Application to Risk Management"; in Rachev, S.T. (ed.): *Handbook of Heavy Tailed Distributions in Finance*; Elsevier/North-Holland, Amsterdam.

ENGLE, R.F.; MUSTAFA, C.; RICE, J. (1992): Modelling peak electricity demand. *Journal of Forecasting*, n. 11, 241–251.

FRANCO, G.; SANSTAD, A.H. (2008): "Climate change and electricity demand in California"; *Climatic Change*, n. 87 (1), 139-151.

GALLANTI, M.; BORGARELLO, M.; GELMINI, A. (2006): Le criticità del sistema gas per l'inverno 2006-2007. *L'Energia Elettrica*, n. luglio/agosto.

GENEST, C.; FAVRE, A.C. (2007): Everything you always wanted to know about copula modeling but where afraid to ask; *Journal of Hydrologic Engineering*, vol. 12(4): 347-368.

GENEST, C.; RIVEST, L.P. (1993): "Statistical Inference Procedure for Bivariate Archimedean Copulas"; Journal of the American Statistical Association, vol. 88 (423): 1034-1043

GENEST, C.; RÉMILLARD, B.; BEAUDOIN, D. (2009): "Goodness-of-fit tests for copulas: A review and a power study; *Insurance: Mathematics and Economics*, vol. 44 (2): 199-213.

HÖEFFDING (1940): "Massstabinvariante Korrelationstheorie" *Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik*, Berlin University, n. 5, 181-233.

JOE, H. (1997): *Multivariate Models and Dependence Concepts*; Chapman & Hall, London.

KOPOCINSKI, R. (2007): Simulating dependent random variables using copulas. Applications to Finance and Insurance; Master Thesis, Wroclaw University of Technology, 05-12/2007.

KOTZ, S.; BALAKRISHNAN, N.; JOHNSON, N.L. (2000): *Continuous Multivariate Distributions – Volume I: Models and Applications, 2nd edition*; Wiley, New York.

MIKOSCH, T. (2006): "Copulas: Tales and facts"; *Extreme*, n. 1(9), 3-20.

NELSEN, R.B. (1999): An Introduction to Copulas; Lecture Notes in Statistics n. 44, Springer, New York.

PATTON, A.J. (2006): "Modelling Asymmetric Exchange Rate Dependence"; *International Economic Review*, n. 47 (2), 527-556.

SAMO, S.R.; MENG, B.E. (1999): **"**Calculation of heating degree-days for Pakistan"; *Building Services Engineering Research and Technology*, Vol. 20, No. 1, 41-44 (1999)

SARAK, H.; SATMAN, A. (2003): "The degree-day method to estimate the residential heating natural gas consumption in Turkey"; *Energy*, n. 28, 929-939

SKLAR, A. (1959): "Fonction de répartition a *n* dimensions et leurs marges; *Publication of the Institut de Statistique*, Université de Paris, n.8: 229-231.

TIMMER, R.P.; LAMB, P.J. (2007): “Relations between Temperature and Residential Natural Gas Consumption in the Central and Eastern United States”; *Journal of Applied Meteorology and Climatology*, n. 46 (11), 1993-2013

VALOR, E.; MENEU, V.; CASELLES, V. (2001): “Daily Air Temperature and Electricity Load in Spain”; *Journal of Applied Metereology*, n. 40, 1413-1421

VITALI, O. (1999): *Statistica per le scienze applicate, Volume I*; Cacucci, Bari.