Statistical Models of Domestic And SME Daily Gas Consumption - Applications To Gas Network Planning And Management

Ronan Oliver
Technological University Dublin, ronan.oliver@tudublin.ie

Follow this and additional works at: https://arrow.tudublin.ie/engdoc

Recommended Citation

This Theses, Ph.D is brought to you for free and open access by the Engineering at ARROW@TU Dublin. It has been accepted for inclusion in Doctoral by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License
Statistical Models of Domestic and SME Daily Gas Consumption - Applications to Gas Network Planning and Management

A thesis submitted to Dublin Institute of Technology in fulfilment of the requirements for the degree of Doctor of Philosophy

By

Ronan Oliver,
School of Civil and Structural Engineering,
Dublin Institute of Technology,
Bolton Street Campus, Dublin 1, Ireland.

Supervisors:
Prof. Aidan Duffy (Dublin Institute of Technology)
Ian Kilgallon (Gas Networks Ireland)

October 2016
DECLARATION

I certify that this thesis which I now submit for examination for the award of PhD, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology and has not been submitted in whole or in part for another award in any Institute.

The work reported on in this thesis conforms to the principles and requirements of the Institute’s guidelines for ethics in research.

The Institute has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis is duly acknowledged.

Signature: [Signature]

Date: Oct '16
ABSTRACT

This research is centred on three pillars of EU energy policy that aim to improve: 1) energy efficiency, in order to reduce CO$_2$ emissions and therefore limit climate change; 2) security of energy supplies, in order to protect economic output and vulnerable citizens in extreme weather; and 3) market integration, in order to increase energy supplier competition and consumer choice in each member state.

To help deliver on these policies, the EU has recently mandated that: 1) gas smart-meters are to be provided to consumers to help improve energy efficiency; 2) network operators ensure adequate gas supplies during extreme cold weather; and 3) network operators provide energy suppliers with forecasts of the volume of gas they should purchase each day in wholesale markets in order to limit the risk to suppliers when entering new markets.

Gas Networks Ireland has part-funded this research and has provided smart-metering and network gas consumption data, so that bottom-up and top-down models of gas consumption can be developed to assist with these EU requirements. Bottom-up models can be used to assess building energy efficiency and to forecast the daily volume of gas to be purchased by an energy supplier for its consumer portfolio. Top-down models can be used to forecast peak-day consumption on the network during extreme weather, and to improve the accuracy of bottom-up portfolio forecasts.

This research develops such models using both ordinary and non-linear least squares (OLS and NLS) regression modelling methods. Each of the resulting models is either based on or develops upon standard heating degree day (HDD) theory used to model
building heating system fuel consumption. It is shown that HDDs are used as an explanatory variable in linear regression models of building gas consumption and that these models can be used to infer building energy performance. This is used as a basis on which to develop a new energy efficiency benchmarking tool for domestic dwellings. This tool is for the use of energy suppliers who must assist their consumers in making energy savings. It is also shown that the HDD approach can be extended to include other variables such as wind speed and solar radiation. This is used as a basis to develop adapted HDD variables to improve estimates of daily gas consumption of individual buildings and of the Irish domestic and SME gas market. These variables are used to develop improved models for bottom-up portfolio and peak-day network forecasting.

The development of the new benchmarking tool is based on the availability of gas smart-metering and household survey data for a sample of dwellings. It is shown that these data allow each parameter of a HDD linear regression model to be estimated using non-linear regression methods rather than the traditional ‘trial and error’ methods applied to monthly or longer fuel consumption data. This improved method is used to estimate HDD models for the dwelling sample and the resulting distribution of independent parameters are presented. These parameter distributions are then characterised by multinomial logistic regression (MLR) models using descriptive household variables. These MLR models are then used to demonstrate a new energy efficiency benchmarking method by comparing the inferred energy end-use of similar buildings.

The NLS regression modelling method is also used to develop an adapted HDD variable to improve estimates of total daily domestic and SME gas market consumption. The resulting model is based on the availability of recent market consumption data and
accounts for numerous effects on gas consumption in addition to those currently estimated by the HDD variable. The improvement in modelling accuracy is quantified by applying a comparative analysis for each of the additional effects accounted for by the new adapted HDD variable. It is found that solar radiation significantly affects gas consumption and should be considered in market consumption models. The new model is used to predict year-ahead peak-day market consumption to alternative supply standards.

Finally, the research develops new models of daily gas consumption for individual consumers based on smart-metering data. These models are developed using SME smart-metering data. This is challenging because their consumption is unpredictable relative to domestic consumers, leading to forecasting difficulties for network operators and energy suppliers. Two modelling options are investigated: one that applies an adapted HDD variable (similar to that referred to above) to estimate the daily gas consumption of individual enterprises using the NLS method; and a second that applies the same market consumption estimator to each enterprises using the OLS method. It is found that OLS models are the most suitable for individual consumer forecasting in terms of the practicality of their implementation and accuracy of their forecasts.
ACKNOWLEDGEMENTS

I would like to thank Dublin Institute of Technology (DIT) and Gas Networks Ireland for sponsoring my research, and the Dublin Energy Lab for hosting this project.

My sincere thanks goes to my supervisors Prof. Aidan Duffy and Ian Kilgallon for their contributions and guidance throughout this research. My thanks to John Turner (Head of School); Dr. Bernard Enright (School of Civil and Structural Engineering) and Rodger O’Connor (Gas Networks Ireland) for their contribution to the project; Gas Networks Ireland for access to data and personnel; and the Irish Social Science Data Archive for providing the applied residential smart-metering data.

I would also like to thank Dr. Niall Murphy (Assistant Head of Applied Technology) for his continued encouragement, and Dr. Marek Rebow (Head of Engineering Research) for his assistance during the project.

I am thankful to my darling wife Ann-Marie for her love and support, and am thankful to my sons Patrick and Colin, my pride and joy and source of inspiration. I am also thankful to my parents and siblings without whom this research would not have been possible.

I would also like to thank past and present colleagues at the Dublin Energy Lab: Dr. Lacour Ayompe, Dr. Fintan McLoughlin, Daire Reilly, Brendan Cleary, Deidre Wolff, Aritra Ghosh, Iftekhar Hussain and Gianni Goretti.
# TABLE OF CONTENTS

Declaration ......................................................................................................................... i

Abstract .......................................................................................................................... ii

Acknowledgements .......................................................................................................... v

Table of Contents ............................................................................................................. vi

List of Figures .................................................................................................................... xii

List of Tables ..................................................................................................................... xvi

List of Acronyms .............................................................................................................. xvii

Nomenclature ................................................................................................................... xix

1 INTRODUCTION ......................................................................................................... 2

1.1 Overview .................................................................................................................. 2

1.2 European Context ................................................................................................... 6

1.2.1 Energy Efficiency and Smart Metering ................................................................. 6

1.2.2 Security of Supply ............................................................................................... 8

1.2.3 Market Integration ............................................................................................... 8

1.3 Aims and Objectives ............................................................................................... 9

1.4 Research Methodology ............................................................................................ 10

1.5 Thesis Layout .......................................................................................................... 13

1.6 Contributions to Knowledge ................................................................................... 17

2 DATA .......................................................................................................................... 20

2.1 Smart Meter Trials ................................................................................................. 20
2.1.1 Domestic Sample ................................................................. 20
2.1.2 SME Sample ................................................................. 21
2.2 NDM Gas Market Data ............................................................. 22
2.3 Climate Data ........................................................................ 23
2.4 Adjusted Weighted Degree Day Data ........................................ 23
3 LITERATURE REVIEW .................................................................. 25
3.1 Irish NDM Gas Market ............................................................. 25
  3.1.1 Wholesale Gas Purchasing ............................................... 27
  3.1.2 FAR Procedures .............................................................. 28
    3.1.2.1 NDM Market Forecasts ............................................ 31
    3.1.2.2 Adjusted Weighted Degree Days .............................. 32
    3.1.2.3 Individual Consumer Models ................................. 34
    3.1.2.4 Supplier Forecasts .................................................. 36
    3.1.2.5 Supplier Allocations ............................................... 37
    3.1.2.6 Balancing Charges ............................................... 37
    3.1.2.7 Reconciliation ...................................................... 38
    3.1.2.8 Reconciliation Charges ......................................... 38
  3.1.3 Peak-Day Forecasts ....................................................... 39
    3.1.3.1 Peak-Day Forecast Procedure ............................... 39
3.2 International Practice .............................................................. 41
  3.2.1 Simple Weather Variables .............................................. 42
  3.2.2 UK-CWV .................................................................. 43
PEAK DAY FORECASTING

Methodology

Data

NDM Market Gas Consumption Model

Extreme Value Modelling

Goodness-of-Fit Test

Results and Discussion

Gas Consumption Model

Peak Consumption Estimates

Peak Day Supply Standard

Peak Week Supply Standard

Conclusion

INDIVIDUAL SME CONSUMER MODELS

Methodology

Industry Model

Ordinary Least Squares Models

Non-linear Least Squares Models

Model Estimation

Model Accuracy

Data

Results and Discussion

Model Estimation
7.2.2 Model Coefficients.................................................................157
7.2.3 Residual Autocorrelation .........................................................160
7.2.4 Model Accuracy.........................................................................161
7.3 Conclusion..................................................................................165
8 CONCLUSIONS, RECOMMENDATIONS AND AFTERWORD ...........169
  8.1 Introduction................................................................................169
  8.2 Conclusions................................................................................170
  8.3 Recommendations........................................................................172
  8.4 Afterword....................................................................................173
    8.4.1 Princeton Scorekeeping Method ............................................173
    8.4.2 Standard Assessment Procedure ..........................................176
REFERENCES.....................................................................................181
LIST OF PUBLICATIONS ................................................................195
LIST OF PRESENTATIONS .................................................................197
APPENDIX........................................................................................199
LIST OF FIGURES

Figure 1.1: Summary of the domestic, SME smart-metering and the NDM market gas consumption datasets and their application in this study. .................................................................6

Figure 1.2: A flowchart summary of the alternative HDD methods applied in this study: 1) the domestic gas end-use efficiency benchmarking tool based on HDDs and simple linear regression model parameters, 2) the peak-day NDM forecasting method based on weekday gas consumption and the first of the adapted HDD variables developed in this study, and 3) the individual consumer models based on daily SME gas consumption and the second of the adapted HDD variables developed in this study. .................................16

Figure 3.1: Daily system average prices (SAPs) and month contract prices (for the corresponding date in the previous month) in the UK National Balancing Point’s (NBP’s) spot and month futures markets [19, 20]. .........................................................27

Figure 3.2: Bottom-up NDM market forecasting procedure ..............................................29

Figure 3.3: Bottom-up NDM market allocation procedure ..................................................30

Figure 3.4: Instances of simple weather variables in published daily forecast models [23], where: $T$ are temperature values for the forecast day (D) and previous days (D-1 and D-2), and $WS$ are wind speed values. .................................................................42

Figure 3.5: Interpretation of the UK-CWV’s transformation function using gas consumption plots adapted from [24]. ....................................................................................45

Figure 4.1: Illustration of each case in the ‘Meteorological Office’ formulae in Equation 4.5 using a base temperature of 13°C and hourly outdoor temperature data from Dublin Airport during 6th - 9th October 2011 .................................................................67

Figure 4.2: HDD calculations using hourly (top) and daily (bottom) outdoor temperature data from Dublin Airport on 23rd and 24th December 2011 ........................................69
Figure 4.3: HDD regression models using ‘Meteorological Office’ (top), hourly temperatures (middle) and daily average temperature (bottom) formulae with an assumed base temperature of 15.5°C, and aggregated gas consumption data for the smart-metered domestic sample in Chapter 5.

Figure 4.4: Idealised indoor temperature profile for intermittently heated building [55].

Figure 4.5: Instantaneous energy balance of a building’s heating system.

Figure 4.6: Idealised temperature distribution across a building fabric element on consecutive days.

Figure 4.7: Temperature variation across a building fabric element.

Figure 4.8: Relationship between weekday gas consumption and daily average outdoor temperature for three sample domestic gas consumers and the NDM gas market for the gas year: October 2009 - September 2010.

Figure 5.1: Summary of the domestic energy efficiency benchmarking method.

Figure 5.2: Distribution of $R^2$ values for the HDD models (522 sample size).

Figure 5.3: Boxplot, categories and distribution of the intercept ($b_0$) parameter for HDD models with an $R^2 \geq 0.6$ (456 sample size).

Figure 5.4: Boxplot, categories and distribution of the slope ($b_1$) parameter for HDD models with an $R^2 \geq 0.6$ (456 sample size).

Figure 5.5: Boxplot, categories and distribution of the base temperature ($T_B$) parameter for HDD models with an $R^2 \geq 0.6$ (456 sample size).

Figure 6.1: Summary of the peak-day forecasting method.

Figure 6.2: Linear relationship between gas consumption for both the NDD$_{CA}$ and the HDD parameters.

Figure 6.3: Block maxima NDD$_{CA}$ histogram and the GEV model’s density distribution.
Figure 6.4: Empirical (points) and the GEV model’s (black line) NDD\textsubscript{CA} return level estimates. ........................................................................................................................................ 139

Figure 6.5: Block maxima NDD\textsubscript{CA,AVG,7D} histogram and the GEV model’s density distribution. ........................................................................................................................................ 141

Figure 6.6: Empirical (points) and the GEV model’s (black line) NDD\textsubscript{CA,AVG,7D} return level estimates. ........................................................................................................................................ 142

Figure 7.1: Summaries of the Industry, OLS, OLS\textsubscript{WD}, NLS and NLS\textsubscript{WD} Models, where arrows are used to illustrate which models will be compared to one another in terms of modelling accuracy. ........................................................................................................................................ 146

Figure 7.2: Boxplots of coefficient p-values for the OLS Models – calculated using standard errors that have been corrected for autocorrelation, and vertical lines below which the estimated coefficients are significant at 10\% and 5\% probability levels, respectively. ........................................................................................................................................ 158

Figure 7.3: Boxplots of coefficient p-values for the NLS\textsubscript{WD} Models – calculated using a one-tailed test that the coefficient is greater than zero [74], and vertical lines below which the estimated coefficients are significant at 10\% and 5\% probability levels, respectively. ........................................................................................................................................ 159

Figure 7.4: Boxplots (and means, in descending order) of $r_1$, the first-order autocorrelation coefficient for the residual error series from the individual ordinary and non-linear least squares models, and a critical region between $\pm 0.05$ (given by $\pm \sqrt{\text{no. of residuals}}$ [72]) where $r_1$ values are insignificant. ........................................................................................................................................ 160

Figure 7.5: Daily gas consumption for a sample SME between Oct.’10 and Sept.’11 (top left); and in order of accuracy, correlation plots for this consumption series (y-axis) based on the corresponding estimates given by each of the alternative models (x-axes). ........................................................................................................................................ 162
Figure 7.6: In order of accuracy, boxplots of MnAPE values for daily gas consumption estimates, given by the alternative models for each consumer in the SME sample, with the mean values of each boxplot in parenthesis.

Figure 7.7: In order of accuracy, boxplots of nAPE values for daily aggregated (or portfolio) estimates given by the alternative model types, with MnAPE values in parenthesis.

Figure A.1: Linear relationship between gas consumption for the HDD parameter in Table 6.4.

Figure A.2: Linear relationship between gas consumption for the HDD($\gamma_1$ & $\alpha_1$=0, wind-speed and outdoor temperature) parameter in Table 6.4.

Figure A.3: Linear relationship between gas consumption for the HDD($\gamma_2$ & $\alpha_1$=0, solar radiation and outdoor temp.) parameter in Table 6.4.

Figure A.4: Linear relationship between gas consumption for the HDD($\gamma_1$ & $\gamma_2$=0, effective outdoor temperature) parameter in Table 6.4.

Figure A.5: Linear relationship between gas consumption for the HDD($\gamma_1$=0, wind-speed and effective outdoor temp.) parameter in Table 6.4.

Figure A.6: Linear relationship between gas consumption for the HDD($\gamma_2$=0, solar radiation and effective outdoor temp.) parameter in Table 6.4.

Figure A.7: Linear relationship between gas consumption for the HDD WA parameter in Table 6.4.

Figure A.8: Linear relationship between gas consumption for the NDD WA parameter in Table 6.4.

Figure A.9: Linear relationship between gas consumption for the NDD CA parameter in Table 6.4.
LIST OF TABLES

Table 3-1: Meter reading frequency in the Irish NDM market [2]. ..........................25
Table 3-2: Composition of annual gas consumption in Ireland for the gas-year: October 2010 to September 2011 [3]..................................................................................................................26
Table 3-3: European peak gas supply standards ........................................................................56
Table 5-1: Applied survey data and their relationship to the HDD regression model. 105
Table 5-2: Parameter starting values and limits ....................................................................107
Table 5.3: Multinomial logistic regression model for the intercept parameter \( b_0 \) .......114
Table 5.4: Multinomial logistic regression model for the slope parameter \( b_1 \) ...........115
Table 5.5: Multinomial logistic regression model for the base temperature parameter \( (T_b) \)..........................................................................................................................116
Table 5.6: Energy efficiency assessments ............................................................................119
Table 6.1: Gas consumption effects estimated by the NDD\textsubscript{CA} parameter .........125
Table 6.2: NDM market model parameter starting values ..................................................128
Table 6.3: NLS solution summary for the gas consumption model and the NDD\textsubscript{CA} parameter..........................................................................................................................133
Table 6.4: NLS solution and in-sample accuracy summaries of incrementally updated HDD based parameters. ........................................................................................................136
Table 6.5: Block maxima NDD\textsubscript{CA} GEV model solution summary .................138
Table 6.6: Block maxima NDD\textsubscript{CA,AVG,7D} GEV model solution summary. .........141
Table 7.1: Parameter starting values and limits for the non-linear least squares models. ...........................................................................................................................................154
# LIST OF ACRONYMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Anderson-Darling</td>
</tr>
<tr>
<td>AR</td>
<td>Autoregressive</td>
</tr>
<tr>
<td>ARIMA</td>
<td>Autoregressive Integrated Moving Average</td>
</tr>
<tr>
<td>AWDD</td>
<td>Adjusted Weighted Degree Day</td>
</tr>
<tr>
<td>BER</td>
<td>Building Energy Rating</td>
</tr>
<tr>
<td>CER</td>
<td>Commission for Energy Regulation (Ireland)</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CWV</td>
<td>Composite Weather Variable</td>
</tr>
<tr>
<td>DEAP</td>
<td>Dwellings Energy Assessment Procedure – for calculating BERs in Ireland</td>
</tr>
<tr>
<td>DM</td>
<td>Daily Metered</td>
</tr>
<tr>
<td>DoW</td>
<td>Day of Week</td>
</tr>
<tr>
<td>DV</td>
<td>Dummy Variable</td>
</tr>
<tr>
<td>HDD</td>
<td>Heating Degree Day</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>EU-EEOS</td>
<td>European Union - Energy Efficiency Obligation Scheme</td>
</tr>
<tr>
<td>EU-GNC</td>
<td>European Union - Gas Network Code</td>
</tr>
<tr>
<td>FAR</td>
<td>Forecasting, Allocation and Reconciliation</td>
</tr>
<tr>
<td>GEV</td>
<td>Generalised Extreme Value</td>
</tr>
<tr>
<td>GNC</td>
<td>Gas Network Code</td>
</tr>
<tr>
<td>GNI</td>
<td>Gas Networks Ireland</td>
</tr>
<tr>
<td>LDM</td>
<td>Large Daily Metered</td>
</tr>
<tr>
<td>LOESS</td>
<td>Local Polynomial Regression</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean Absolute Percentage Error</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MnAPE</td>
<td>Mean normalised Absolute Percentage Error</td>
</tr>
<tr>
<td>MLR</td>
<td>Multinomial Logistic Regression</td>
</tr>
<tr>
<td>NAC</td>
<td>Normalised Annual Consumption</td>
</tr>
<tr>
<td>NBP</td>
<td>National Balancing Point – the UK’s wholesale gas market</td>
</tr>
<tr>
<td>NDD</td>
<td>Network Degree Day</td>
</tr>
<tr>
<td>NDM</td>
<td>Non-Daily Metered</td>
</tr>
<tr>
<td>NLS</td>
<td>Non-linear Least Squares</td>
</tr>
<tr>
<td>NWP</td>
<td>Numerical Weather Prediction</td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary Least Squares</td>
</tr>
<tr>
<td>PRISM</td>
<td>PRinceton Scorekeeping Method</td>
</tr>
<tr>
<td>SAP</td>
<td>System Average Price – in the UK’s wholesale gas market; or Standard Assessment Procedure – for calculating BERs in the UK</td>
</tr>
<tr>
<td>SME</td>
<td>Small-to-Medium Enterprise</td>
</tr>
<tr>
<td>TSO</td>
<td>Transmission System Operator</td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>US-EPA</td>
<td>United States - Environmental Protection Agency</td>
</tr>
<tr>
<td>WC</td>
<td>Wind Chill</td>
</tr>
</tbody>
</table>
## NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.33NV</td>
<td>Building air-infiltration heat loss coefficient ( (W/°C) )</td>
</tr>
<tr>
<td>A</td>
<td>Area ( (m^2) )</td>
</tr>
<tr>
<td>A^2</td>
<td>Anderson-Darling test statistic</td>
</tr>
<tr>
<td>AU^2</td>
<td>Modified-Anderson-Darling upper tail test statistic</td>
</tr>
<tr>
<td>AWDD</td>
<td>Adjusted weighted degree day ( (°C \cdot \text{day}) )</td>
</tr>
<tr>
<td>AWDD_D</td>
<td>AWDD forecast ( (°C \cdot \text{day}) ) for day (D)</td>
</tr>
<tr>
<td>ΣAWDD_M</td>
<td>Total AWDDs ( (°C \cdot \text{day}) ) for metered period (M)</td>
</tr>
<tr>
<td>c_{eff}</td>
<td>Building fabric effective heat capacity ( (kJ/°C) )</td>
</tr>
<tr>
<td>C</td>
<td>Gas consumption – varied units: kWh, MWh or GWh</td>
</tr>
<tr>
<td>C_D</td>
<td>Gas consumption for day (D)</td>
</tr>
<tr>
<td>Ĉ_D</td>
<td>Gas consumption estimate/forecast for day (D)</td>
</tr>
<tr>
<td>C_M</td>
<td>Gas consumption for metered period (M)</td>
</tr>
<tr>
<td>ΔC_M</td>
<td>Difference between forecast and actual gas consumption for metered period (M)</td>
</tr>
<tr>
<td>Ĉ_{NDM,D}</td>
<td>Forecast NDM market gas consumption for day (D)</td>
</tr>
<tr>
<td>C_{NDM,D-n}</td>
<td>NDM market gas consumption for day (D-n)</td>
</tr>
<tr>
<td>C_{NDM,WD}</td>
<td>NDM market gas consumption for weekdays (WD)</td>
</tr>
<tr>
<td>Ĉ_{Sl,D}</td>
<td>Gas consumption forecast/allocation for supplier (i) for day (D)</td>
</tr>
<tr>
<td>CW</td>
<td>Composite weather (temperature) value ( (°C) )</td>
</tr>
<tr>
<td>dT_{BF}/dt</td>
<td>Rate of change of the building fabric temperature</td>
</tr>
<tr>
<td>D</td>
<td>Instantaneous heat demand ( (kW) )</td>
</tr>
<tr>
<td>Days_M</td>
<td>Number of days in metered period (M)</td>
</tr>
<tr>
<td>DoW</td>
<td>Day of week adjustment factor</td>
</tr>
<tr>
<td>DoW_{Dom.}</td>
<td>DoW adjustment factor for domestic (Dom.) consumers</td>
</tr>
<tr>
<td>DoW_{SME}</td>
<td>DoW adjustment factor for SME consumers</td>
</tr>
<tr>
<td>DoW_{WD}</td>
<td>DoW adjustment factor for weekdays (WD)</td>
</tr>
<tr>
<td>DoW_{WE/Hol.}</td>
<td>DoW adjustment factor for weekends/holidays (WE/Hol.)</td>
</tr>
<tr>
<td>DV</td>
<td>Dummy variable</td>
</tr>
<tr>
<td>E</td>
<td>Energy consumption</td>
</tr>
<tr>
<td>F</td>
<td>Fuel consumption</td>
</tr>
<tr>
<td>F_{TS,D}</td>
<td>Fuel consumption due to thermal storage for day (D)</td>
</tr>
</tbody>
</table>
GR  Global radiation (J/cm²)
HDD  Heating degree day (°C·day)
ΣHDDM  Total HDDs (°C·day) for metered period (M)
HDDSA  Solar adjusted HDD (°C·day)
HDDS&W  Solar and wind adjusted HDD (°C·day)
HDDWA  Weather adjusted HDD (°C·day)
HDDWDA  Weather and day-type adjusted HDD (°C·day)
H₀  Null hypothesis
HLC  Overall heat loss coefficient (kW/°C)
HLVD  Heat loss variable (kW/°C) for day (D)
k  GEV model shape parameter
MAPE  Mean absolute percentage error (%)
MnAPE  Mean normalised absolute percentage error (%)
N  Number of air changes per hour (1/h)
ND  Mean number of air changes per hour (1/h) for day (D).
NDD  Network degree day (°C·day)
NDDCA  Climate-adjusted network degree day (°C·day)
NDDCA,0.02  1-in-50 year NDDCA (°C·day)
NDDCA,AVG,7D  7-day average NDDCA (°C·day)
NDDWA  Weather-adjusted network degree day (°C·day)
NLS  Non-linear least squares regression model
NLSWD  Within-Day NLS regression model
OLS  Ordinary least squares regression model
OLSWD  Within-Day OLS regression model
P  Return period (years)
Q  Heat (kW)
QAI  Air infiltration heat loss (kW)
QBF  Building fabric heat loss (kW)
QHS  Heating system output (kW)
QIG  Internal heat gain (kW)
QSG  Solar heat gain (kW)
QTS  Thermal storage heat gain/loss (kW)
Q̄TS  Average thermal storage heat gain/loss (kW)
r  Correlation coefficient
R^2  Coefficient of determination
s  Slope parameter
SF_D  Scaling factor for day (D)
S-HDD_D  Seasonal HDD (°C·day) for day (D)
S-NDDWA,D  Seasonal NDDWA (°C·day) for day (D)
S-NDDWA,d  Seasonal NDDWA (°C·day) for day of the year (d)
SS-NDDWA,D  Smoothed Seasonal-NDDWA (°C·day) for day (D)
T  Temperature (°C)
T_B  Base temperature parameter (°C)
T_B,DoW  Base temperature parameter for a given day of the week (°C)
T_B,low  Lower base temperature parameter (°C)
T_B,upr  Upper base temperature parameter (°C)
\bar{T}_D  Average temperature (°C) for day (D)
T_{EFFECT,D}  Effective temperature (°C) for day (D)
T_{EFFECT,O,D}  Effective outdoor temperature (°C) for day (D)
ST_{EFFECT,D}  Seasonal effective temperature (°C) for day (D)
T_G  Equivalent temperature (°C) due to heat gains in a building
T_HI  High temperature parameter (°C)
T_I  Indoor temperature (°C)
\bar{T}_I  Average indoor temperature (°C)
T_{IG}  Equivalent temperature effect of internal heat gain (°C)
T_{LWR}  Lower temperature parameter (°C)
T_O  Outdoor temperature (°C)
\bar{T}_{O,D}  Average outdoor temperature (°C) for day (D)
T_O,h  Outdoor temperature (°C) at hour (h)
T_{O,MAX}  Maximum outdoor temperature (°C)
T_{O,MIN}  Minimum outdoor temperature (°C)
T_{SG}  Equivalent temperature effect of solar heat gain (°C)
T_{SP}  Set-point temperature (°C)
T_{TS}  Equivalent temperature effect of thermal storage (°C)
T_{UPR}  Upper temperature parameter (°C)
\Delta T  Temperature differential (°C)
\Delta T_{BF}  Change in building fabric temperature over a day (°C/day)
U  U-value (W/m²·°C)
ΣUA  Building fabric heat loss coefficient (W/°C)
V  Volume of heated space (m³)
WD-n  ‘Weather’ value for day (D-n)
ŴD  ‘Weather’ forecast for day (D)
ŴD-n  ‘Weather’ forecast for day (D-n)
WC_D  Wind-chill on day (D)
WDD  Weighted degree day (°C·day)
WS_D  Wind speed (knots) for day (D)
WS_B  Base wind speed (knots)
ΨS  Average wind speed (knots)

Greek symbols:

\( \alpha_1 \)  Thermal storage parameter
\( \gamma_1 \)  Global radiation coefficient
\( \gamma_2 \)  Wind speed coefficient
\( \Delta \)  Difference
\( \varepsilon \)  Model error
\( \varepsilon_D \)  Model error for a day (D)
\( \varepsilon_M \)  Model error for metered period (M)
\( \varepsilon_{WD} \)  Model error for a weekday (WD)
\( \eta \)  Building heating system efficiency (%) 
\( \mu \)  GEV model location parameter
\( \nu \)  Random model error
\( \rho \)  Autoregressive parameter
\( \sigma \)  GEV model scale parameter
\( \chi^2 \)  Likelihood ratio
\( \omega_1 \)  Seasonal coefficient
CHAPTER 1

INTRODUCTION
1 INTRODUCTION

1.1 Overview

In Ireland, the gas network forms a key element of the country's energy supply infrastructure, delivering gas to electric power stations, large industry, small-to-medium enterprises (SMEs) and domestic dwellings, for example. The delivery and maintenance of this infrastructure is provided by Gas Networks Ireland (GNI), who must ensure the safety and security of gas supply to consumers, and manage their consumption on behalf of energy suppliers.

Currently, GNI manages the supply of gas to approximately 650,000 domestic dwellings and SMEs, the latter comprising any non-domestic consumer with an annual gas requirement below 5,550,000 kWh [1]. These are known as ‘non-daily metered’ (NDM) consumers since their consumption is recorded manually by GNI, four to twelve times per year, depending on their annual gas requirement [2]. Together they account for almost 60% of Ireland’s annual gas consumption excluding electricity generation [3].

Gas consumed by this NDM market is purchased in advance by energy suppliers from wholesale markets and then transported to the consumer by GNI. Unlike electricity that is generated and consumed instantaneously and necessitates half-hour trading periods in wholesale markets, gas can be delivered and stored on the network before it is consumed and is traded in daily volumes. GNI therefore manages NDM market consumption on a daily basis. It is responsible for the planning and operation of the network including the following tasks:
• metering this consumption at network and individual consumer levels;

• forecasting the daily gas requirement of each energy supplier’s NDM consumer portfolio;

• forecasting and securing the supply of peak-day NDM market consumption; and

• delivering on Ireland’s recent commitment to roll out smart gas metering.

To facilitate daily NDM market forecasts, GNI has developed individual models of daily gas consumption for each consumer in the market using their meter readings gathered at monthly to bi-monthly intervals. The advantage of these models is that the resulting daily estimates of an individual consumer’s gas consumption can be easily assigned to each supplier’s consumer portfolio, which is constantly changing as customers switch suppliers, new connections are made and old ones terminated. These are then aggregated and the resulting estimates are used by GNI to inform each supplier of the purchases they must make from the wholesale gas market each day. This interaction allows GNI to have greater certainty with respect to maintaining gas network pressure.

GNI must also ensure that gas supplies are available to the NDM market during extremely cold weather, as over 95% of buildings in this market are domestic dwellings [3], with occupants such as the elderly who can be vulnerable to such weather conditions. This maximum supply capacity is based on a probabilistic peak-day forecast that is estimated before each heating season using a daily market gas consumption model. The estimate is used by GNI to inform network investment and year-ahead
operating and maintenance plans, and also to base network capacity charges to suppliers.

Improvements to these individual consumer and peak-day modelling processes can benefit GNI, energy suppliers and the consumers they serve. Improved individual consumer models allow energy suppliers to trade in wholesale gas markets more accurately and GNI to manage the supply and transport of this gas more effectively. Errors in this individual consumer modelling process results in energy suppliers either purchasing too little or too much gas for a consumer on a given day. GNI must apply charges to either recoup the cost of purchasing the additional volume of gas consumed or credit back to energy suppliers the value of unnecessary gas purchased. Because these charges may be recouped or credited back at rates higher or lower than that originally paid, any deficits can result in higher costs to the consumer. Improved peak-day models allow for more reliable assessments of the adequacy of the network to potential extreme cold weather periods, and hence allow GNI to develop network development and maintenance plans which will result in better preparedness in the most cost-effective manner. Any financial savings resulting from these improved models can result in savings to customers and society.

GNI is committed to the continuous improvement of its gas management processes, and has co-funded this research so that improvements to these individual consumer and peak-day gas consumption models can be explored. For this research, GNI has provided training and support in addition to smart metering and daily market consumption data. These smart metering data are available as a result of recent trials that led to the decision to roll out smart meters to the Irish NDM market from 2018 [4], and were collected for over 1500 domestic and 50 SME consumers.
The daily gas consumption data available from these smart meters have the potential to be used to develop more accurate individual consumer models than those currently based on monthly (or longer) meter readings. SME rather than domestic smart metering data is to be used for this purpose. This is because the daily gas consumption data of SMEs is relatively more difficult to model given different industries’ diverse gas requirements and significant variation in this consumption on different days of the week. Consequently, GNI has found that SMEs are the most significant source of error in the current modelling process. The smart metered SME sample therefore provides a better basis in which to develop improved individual consumer models for the domestic and SME (or NDM) gas market.

GNI has also supported the development of a gas end-use efficiency benchmarking tool for the consideration of energy suppliers who are now required by the EU to assist consumer energy savings [5]. The smart metered domestic sample is used for this purpose because they are the most critical component of NDM consumption and because data on building and household variables were collected (dwelling type, construction year, number of bedrooms and occupants) which are fundamental for identifying energy efficiency measures; these variables were not collected with the SME data. These household data can be used to quantify the efficiency of gas consumption in these smart metered dwellings. The resulting benchmarking tool can be used by suppliers to help screen gas consumers and target suitable energy saving opportunities at the most appropriate households.

Figure 1.1 provides a high-level summary of how the available gas consumption data are applied in this study. A more detailed flowchart describing the research methodology is provided later in Figure 1.2.
1.2 European Context

These opportunities to develop improved gas consumption models and a gas end-use efficiency benchmarking tool for the Irish NDM gas market are also relevant to other European network operators and energy suppliers. This is due to developments of common interest at an EU level that aim to improve building energy efficiency, security of gas supply and energy market integration.

1.2.1 Energy Efficiency and Smart Metering

In the EU, domestic buildings are responsible for 26% of annual energy consumption and 37% of this energy is consumed as gas [6]. Domestic gas consumers can therefore make a significant contribution to the EU’s 2020 targets of: 1) a 20% reduction in greenhouse gas emissions from 1990 levels; 2) a 20% increase in energy from
renewable resources; and 3) a 20% improvement in energy efficiency [7]; and thus help to meet the objective of decarbonising energy end-use in Europe.

To help realise such improvements and a reduction in fossil fuel imports, the EU has mandated that smart meters are made available to gas consumers in each member state, except those states where an adverse cost benefit has been established [8]. This has resulted in the planned installation of these meters in many countries across the EU including Ireland and, for example, the United Kingdom (UK) where 22 million are planned for installation by 2019 and France where, 11 million could be in place before 2020 [9]. In such countries, consumers will have access to high resolution time-of-use consumption data. Sampling intervals for smart meters are typically hourly (or less) compared to monthly (or more) for traditional manually-read meters. Access to such high-frequency data will enable consumers to manage their gas consumption more effectively and identify readily achievable energy savings.

The EU has also recommended that energy distributors and/or suppliers provide assistance to consumers to help reduce their energy consumption. In this regard, each EU member state can implement an ‘Energy Efficiency Obligation Scheme’ (EU-EEOS) to ensure that suppliers achieve energy savings each year from 2014 to 2020 that are at least equivalent to 1.5% of their consumers’ average annual energy consumption between 2010 and 2012 [5]. Such schemes can benefit from smart metered gas consumption data and the gas end-use efficiency benchmarking tool developed in this study.
1.2.2 Security of Supply

The EU is dependent on imports for over 87% of its annual gas consumption [6], and these may be subject to supply restrictions as a result of extreme cold weather or disruptions as a consequence of geopolitical risks. In this regard, the EU has recently developed a supply standard that stipulates different scenarios during which supplies must be ensured for ‘protected’ (mainly domestic) consumers [10]. However, this includes a new peak supply standard that is different to that already applied to the Irish NDM market.

In the new EU standard, peak supply capacity is quantified by extreme or 1-in-20 year temperatures over a 7-day peak period [10]; whereas in the Irish standard, it is quantified by a 1-in-50 year ‘composite weather variable’ (CWV) for a weekday [11] – CWVs have been developed by the gas industry to account for numerous weather effects on gas consumption in addition to temperature such as wind-chill, for example (see Sections 3.1.3.1 and 3.2.2). Consequently, this study quantifies the difference in peak supply capacity required by alternative supply standards. This assessment is based on Irish NDM market consumption data and an adapted HDD variable developed later in this study.

1.2.3 Market Integration

Improved energy market integration is central to EU energy policy [12]. To facilitate new market entrants, network operators such as GNI have been established in each member state to assume control of gas networks from the incumbent suppliers [13]. And to increase competition, industrial and domestic consumers have been free to choose alternative energy suppliers since 2004 and 2007, respectively [13]. However, in order
to encourage greater market integration across the EU, the role of network operators has recently been harmonised by the establishment of a common gas network code (EU-GNC) [14].

Part of this code is the requirement that network operators provide energy suppliers with daily gas consumption forecasts for their portfolio of NDM consumers [14], similar to those already provided by GNI. However with smart metering, more accurate individual consumer models can be developed than those based on monthly or longer meter readings. Such models are developed later in this study for the benefit of European network operators currently in the process of adopting smart metering infrastructure.

1.3 Aims and Objectives

The aim of this research is to use newly available data sources to develop models of individual consumer and NDM market gas consumption, which can be used by the gas industry to inform consumer level energy efficiency initiatives, network planning operations and daily gas procurement processes. Specific objectives include developing methods for estimating:

- the efficiency of gas end-use consumption in individual dwellings;

- daily gas consumption for the NDM market;

- peak-day gas consumption for the NDM market;

- daily gas consumption of individual SMEs.
1.4 Research Methodology

This research develops statistical models based on heating degree day (HDD) theory taking the general form: $C_D = f(HDD)$, where $C_D$ is gas consumption for a given day (D). The HDD ($^\circ$C·day) variable is an established estimator of building heat consumption that is commonly used in gas consumption models. Traditionally, HDDs have been applied in linear regression models using consumption data from monthly (or longer) gas bill meter readings. In these models HDDs are used to account for the cumulative indoor-outdoor temperature differential necessitating building heat consumption each month.

This HDD temperature differential is calculated as the difference between a base temperature parameter (which is related to indoor temperatures and is used to estimate the outdoor temperature above which heating is not required) and outdoor temperature data for the local weather station (see Equation 4.7). However, traditional HDD models are limited by monthly (or longer) meter readings which make it difficult to estimate a building’s base temperature or the effects of other factors on building heating such as wind speed and solar gains. For simplicity, published HDD data with an assumed base temperature from the local meteorology service are applied in traditional models. Such HDDs are applied instead of estimating the actual base temperature of the individual building by manual ‘trial and error’ methods. The effects of weather variables other than temperature are ignored, as plots of monthly consumption against monthly HDDs exhibit much less scatter than corresponding daily plots where other weather effects are more apparent.
This research overcomes these limitations by employing non-linear least squares (NLS) regression methods. NLS is an iterative computational method used to fit non-linear models to measured data. It can be used with daily gas consumption data to estimate a building’s actual base temperature parameter within, rather than separately to the HDD regression model by ‘trial and error’, as in the case of the traditional method. This improved base temperature estimation method is used in the development of the new gas end-use efficiency benchmarking tool referred to previously.

This benchmarking method is illustrated in Figure 1.2. It begins by using NLS to estimate a HDD regression model for each consumer in the domestic smart metering dataset. The resulting distributions of independent parameters which relate to alternative domestic gas end-uses are then presented. These distributions are then related to household characteristics using multinomial logistic regression (MLR) models based on descriptive household survey data that were generally known by consumer sample. These MLR models can be used to estimate the probability that an individual consumer’s HDD model parameter estimates are higher or lower than expected when compared to similar households. The MLR models are then used to compare the inferred energy efficiency of a sample of buildings with similar characteristics thus allowing energy saving interventions which are likely to be appropriate for the consumer to be identified.

The non-linear regression method is also used to develop improvements to the HDD variable for modelling daily gas consumption. HDDs are based on a building energy model of heat consumption which accounts for internal and external temperatures only. This model is used to adjust the HDD variable to account for additional effects such as solar radiation and wind speed. Two types of adapted HDD variables were developed in
this regard. The first of these variables is used to forecast year-ahead peak-day gas consumption for the NDM market.

This forecasting method is illustrated in Figure 1.2. It begins by estimating a daily gas consumption model for the NDM market including the coefficients of each weather variable within the adapted HDD variable. Adapted HDDs are then calculated using long term (>30 years) climate data so that various return levels (e.g. 1-in-50 year values) of the variable can be estimated by an extreme value model. These extreme values are then applied in the NDM market model to quantify the difference in year-ahead peak-day forecasts using alternative gas supply standards.

The second type of adapted HDD variable developed in this study is used to model the individual daily gas consumption data of smart-metered SMEs. This modelling method is also illustrated in Figure 1.2. It employs a NLS method to estimate building base temperatures for each day of the week, as well as the effect of weather on each enterprise’s daily gas consumption. The accuracy of the resulting models is compared to alternative OLS models and an industry model which apply the same market consumption estimator to each enterprise. Such market consumption estimators assume that the annual profile of daily gas consumption or weather response for each consumer follows that of the market. This assumption has been traditionally applied by network operators to estimate daily gas consumption of consumers limited to monthly (or longer) manually read meter readings.

However, with smart-metering data this assumption need no longer be made. The second of the adapted HDD variables is used to assess the benefit of independently estimating the daily gas consumption of an individual consumer’s building heating
system to varied weather conditions. Because of the computational intensity of the individualised NLS approach, computation times are compared to the alternative OLS models and an industry model.

The research is completed by an assessment of the benefit of real-time smart metering data for these individual SME models. This data is only available if smart-metered gas consumption is uploaded on a daily basis by the network operator at an additional cost, instead of downloading such data less regularly on a monthly basis, for example. The benefit of real-time data is that the gas consumption value for the previous day can be used to improve the accuracy of next or within-day gas consumption forecasts. Because of the additional cost of real-time smart metering data, the forecasting accuracy of individual SME models with and without such data is compared.

1.5 Thesis Layout

Chapter 2 Data: is a description of the data used in this study, including data from the Irish domestic and SME smart meter trials, daily NDM market gas consumption data, and climate data from Dublin Airport used to calculate the alternative HDD variables applied in this study.

Chapter 3 Literature Review: initially this is a review of the Irish gas industry with particular reference made to the NDM market and the current gas modelling methods and operational codes used to manage daily gas consumption for this market. A review of international gas consumption modelling and peak-day forecasting literature is then presented, followed by a description of current methods used to benchmark the energy efficiency of buildings using metered energy consumption data.
Chapter 4 Heating Degree Days: is a detailed description of the HDD variable, the building energy model on which it is based, and the development of the adapted HDD variables used to model the daily gas consumption of either individual SMEs or the Irish domestic and SME gas market.

These alternative HDD variables are applied in this study as illustrated in Figure 1.2 – i.e. models of domestic gas consumption using the HDD variable are applied first; a model of weekday NDM market gas consumption using the first of the adapted HDD variables is applied next; and models of daily SME gas consumption using the second of the adapted HDD variables is applied last.

Chapter 5 Benchmarking: presents the statistical benchmarking method illustrated in Figure 1.2. This can be used by energy suppliers to infer the efficiency of cooking, hot water and space heating gas consumption in buildings in their domestic portfolio, so that energy saving interventions can be targeted to suitable consumers. The method is demonstrated using a small sample of consumers.

Chapter 6 Peak-Day Forecasting: presents a methodology to forecast year-ahead peak-day gas consumption for the NDM market as illustrated in Figure 1.2. This is used to quantify the difference in year-ahead peak day NDM market forecasts to alternative supply standards. It is also shown that solar radiation significantly affects gas consumption and should be considered in gas consumption models.

Chapter 7 Individual SME Consumer Models: presents new individual consumer models of daily SME gas consumption based on the availability of smart metered data. The second of the adapted HDD variables is applied in this assessment, as illustrated in
Figure 1.2. Additional models based on the market consumption estimator applied by GNI are also assessed in the model accuracy comparisons referred to at the bottom of this flowchart. It is found that OLS models are the most suitable in terms of the practicality of their implementation and accuracy of their forecasts.

**Chapter 8 Conclusions, Recommendations and Afterword**: completes the thesis and provides conclusions for the research and further areas that can be investigated.
**Simple Linear Regression Models**
- **HDDs**
  - NLS Estimation of Model Parameters
- Linear and Base Temperature Parameter Distributions
  - Multinomial Logistic Regression Models
  - Domestic Energy Efficiency Assessments

**Peak-Day NDM Market Forecasts**
- **HDDs**
  - NLS Estimation of Model Parameters
- Extreme Value Models
  - Peak-Day NDM Market Forecasts

**Individual SME Models**
- Multiple Linear Regression Model
  - **adapted HDDs**
  - NLS Estimation of Model Parameters
  - Multiple Linear Regression Model
    - **adapted HDDs**
    - Real-time smart metering data
  - NLS Estimation of Model Parameters
  - Individual SME Model Accuracy Comparisons

---

**Figure 1.2:** A flowchart summary of the alternative HDD methods applied in this study: 1) the domestic gas end-use efficiency benchmarking tool based on HDDs and simple linear regression model parameters, 2) the peak-day NDM forecasting method based on weekday gas consumption and the first of the adapted HDD variables developed in this study, and 3) the individual consumer models based on daily SME gas consumption and the second of the adapted HDD variables developed in this study.
1.6 Contributions to Knowledge

The contributions to knowledge of this research can be summarised as follows:

- The first time application of NLS methods to estimate building base temperatures within HDD regression models of (smart metered) daily gas consumption. Traditionally, with longer interval fuel consumption data, base temperatures are either assumed in published HDD data or are estimated by ‘trial and error’ methods as a secondary step to an OLS model. The benefit of the NLS method is that it estimates this non-linear base temperature parameter simultaneously with the linear parameters of the HDD regression model in a single step with high precision.

- The development of a benchmarking tool to infer the efficiency of gas consumption in smart-metered dwellings. Current benchmarking tools are based on energy intensity parameters normalised by building floor area e.g. kWh/m²/year. However, these tools presuppose that floor area data are readily available even though this research later finds that many householders are unable to provide this measurement when surveyed. The benefit of the new benchmarking method is that it quantifies the efficiency of common gas end-uses using simple household survey data known to consumers so that appropriate energy saving interventions can be identified.

- The first-time application of NLS methods to estimate effects other than base and outdoor temperature within a HDD regression model of daily gas market consumption. Current state-of-the-art models apply composite weather variables
(CWVs) that account for numerous weather effects such as temperature and wind-chill. However, these CWVs do not account for the important effect of solar radiation. The NLS method allows building heat gain effects due to solar radiation to be correctly estimated as an equivalent temperature effect within an adapted HDD variable. This and other important effects are estimated in the resulting adapted HDD variable simultaneously to the linear parameters of the applied regression model.

- The development of a second and similar adapted HDD variable to assess the practicality of NLS models and such state-of-the-art gas consumption estimators for individual consumers in NDM markets. The benefit of this assessment is that finds that such methods are impractical due to excessive computation time and because NLS convergence issues were found for some consumers with irregular daily gas consumption. Consequently, simpler OLS models that do not suffer with these issues were also developed using the market consumption estimator applied by GNI.
CHAPTER 2

DATA
This study is based on data from the Irish domestic and SME smart meter trials, daily gas consumption data for the Irish non-daily metered (NDM) gas market and long-term climate data from Dublin Airport.

2.1 Smart Meter Trials

The planned roll-out of smart metering to the Irish NDM gas market is the result of a positive cost benefit analysis [15], as recommended by the EU Directive which calls for the availability of these meters [8]. As part of this cost benefit analysis, smart meters were trialled at over 1,500 domestic and 50 SME consumers during different time-periods between 2009 and 2011.

2.1.1 Domestic Sample

The domestic smart meter trial participants were selected to be representative of the domestic consumers in the NDM gas market. This selection was confirmed by a pre-trial telephone survey, which collected the following data for each dwelling: the number of adult and children occupants; the type, size and age of the building; the alternative gas uses – for example: cooking, hot-water and space heating; and type of heating controls utilised.

In order to determine the benefit of smart meters and additional energy efficiency stimuli such as detailed energy statements and in-house (gas consumption) display devices, these participants were allocated into various groups before the end of a smart metered benchmark period. This included a control group of over 500 consumers who...
received no stimuli and were requested to continue using their gas as normal during the trial period. The effect of the various stimuli such as in-home (energy) display devices was assessed using statistical tests which compared the difference in gas consumption between the benchmark and trial periods for each test group to that for the control group. Based on these and net present value tests, the trial established a positive cost benefit for the provision of smart meters, in-house display devices and detailed energy statements to domestic gas consumers in the NDM gas market [16].

The survey and the corresponding smart metered gas consumption data from this trial is available publically in anonymised format [17]. Although the exact locations of these households are not available, it is known that they are located in either in Dublin (64%), or in urban centres no more than approximately 250km from Dublin [18]. The control group in this dataset is used to develop the benchmarking tool later in this thesis, as these consumers were not subject to energy saving stimuli during the smart meter trials. These data were recorded between December 2009 and May 2011.

2.1.2 SME Sample

Given the diversity of SME sectors in the NDM gas market it was not practical to assess the benefit of smart metering for each sector using a statistical experiment similar to that applied in the domestic smart meter trials. For example, such experiments would have resulted in a very large samples or proportion of the (relatively small) population of over 23,500 SMEs in the NDM gas market. Therefore, the qualitative approach described below was adopted for the SME smart meter trial.

For this trial, over 50 SMEs were selected to represent the largest non-domestic gas consuming sectors in Ireland, including: restaurants, public houses, takeaways;
government buildings; hotels and/or leisure facilities; healthcare buildings; educational buildings; and industrial facilities. These were surveyed by telephone after the trial and were questioned if energy savings had been made during the trial due to the availability of smart metered data. Respondents found that they could attribute between 5% and 10% energy savings in this regard [18].

Gas consumption data from the SME trial participants between October 2010 and September 2011 were used to develop individual consumer models for the NDM market. These data has been provided by GNI and are not available publically.

2.2 NDM Gas Market Data

Daily gas consumption data for the Irish NDM gas market is calculated rather than metered. It is given by the difference between the total gas supplied to the network and that consumed by GNI (to operate the network), electricity generators and daily metered consumers (large industrial users) as well as that lost from the system. Such data for the three years between October 2009 and September 2012 were used to develop an adapted HDD parameter to forecast year-ahead peak-day gas consumption for the NDM market in Chapter 6. Again, this data has been provided by GNI and is not available publically. Importantly, this data includes two extreme cold weather periods that were observed in Ireland during January and December 2010. Such extreme weather periods are rare and provide a unique opportunity to test the accuracy of market models for extreme cold weather.
2.3 Climate Data

Daily climate data from Dublin Airport is used to calculate the various HDDs variables applied in this study. This climate dataset consists of daily temperature, wind speed and global radiation values since 1976 – as long-term data is required for peak-day gas consumption forecasting. The length of this data series has been limited by global radiation, which is only available since this date. Occasional missing global radiation values in the dataset have been replaced by their equivalent 30 year average values. These missing values account for less than 1% of the dataset.

2.4 Adjusted Weighted Degree Day Data

The adjusted weighted degree day (AWDD) variable is an estimator of market consumption that is described in detail in Section 3.1.2.2. It is used by GNI to forecast the daily gas consumption of individual domestic and SME consumers in the NDM market. AWDD data has been provided by GNI for the same time-period as the SME smart-metering trials. These are used to replicate the current individual consumer model applied by GNI to monthly-metered SMEs in the NDM market. This model is described in detail in the literature review and is used to assess the accuracy of the new individual consumer models developed using the SME smart-metering dataset in Chapter 7. Some of these new models also apply the AWDD data provided by GNI. Again these data are not available publically.
CHAPTER 3

LITERATURE REVIEW
3 LITERATURE REVIEW

This review is split into three main sections. The first section describes the scale of the Irish NDM gas market relative to large industrial consumers and electricity power generators, the meter reading frequencies currently applied to this market, the alternative wholesale gas purchasing strategies employed by energy suppliers operating in the market, and the procedures used by GNI to manage and forecast daily and peak-day consumption for the market. The second section summarises international literature related to these gas management and forecasting processes. And the final section reviews current domestic energy efficiency benchmarking methods based on metered energy consumption data.

3.1 Irish NDM Gas Market

The Irish NDM gas market includes approximately 620,000 domestic and 23,500 SME consumers [3], whose meters are read at the frequencies given in Table 3-1.

Table 3-1: Meter reading frequency in the Irish NDM market [2].

<table>
<thead>
<tr>
<th>Meter Reading Frequency</th>
<th>Annual Gas Consumption (kWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-monthly: 4 actual plus 2 estimated readings per year</td>
<td>&lt; 72,999</td>
</tr>
<tr>
<td>Monthly: 12 actual readings per year</td>
<td>≥ 73,000</td>
</tr>
</tbody>
</table>

These NDM consumers compare to 239 daily metered (DM) and 44 large daily metered (LDM) consumers [3], which have an annual gas requirement over 5,550,000kWh and 57,500,000kWh, respectively [1]. Although the annual gas consumption of individual
NDM consumers is small relative to DM and LDM consumers, the aggregate consumption of NDM market is significant, as it accounts for 21.8% of Ireland’s annual gas consumption as shown in Table 3-2.

Table 3-2: Composition of annual gas consumption in Ireland for the gas-year: October 2010 to September 2011 [3]

<table>
<thead>
<tr>
<th>Consumer Category</th>
<th>Annual Gas Consumption (kWh)</th>
<th>Share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electricity Power Generators</td>
<td>35,432</td>
<td>62.6</td>
</tr>
<tr>
<td>LDM Consumers</td>
<td>4,911</td>
<td>8.7</td>
</tr>
<tr>
<td>DM Consumers</td>
<td>3,020</td>
<td>5.3</td>
</tr>
<tr>
<td>NDM Consumers:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>● SME Consumers</td>
<td>(4,023)</td>
<td>(7.1)</td>
</tr>
<tr>
<td>● Domestic Consumers</td>
<td>(8,340)</td>
<td>(14.7)</td>
</tr>
<tr>
<td>GNI</td>
<td>889</td>
<td>1.6</td>
</tr>
<tr>
<td>Total</td>
<td>56,615</td>
<td>100</td>
</tr>
</tbody>
</table>

In order to manage the significant market share of domestic and SME consumers on a daily basis, GNI has developed forecasting, allocation and reconciliation (FAR) procedures for the NDM market that are described in detail in Section 3.1.2. The forecast models in these procedures are used to inform the amount of gas that each energy supplier purchases for a given day.
3.1.1 Wholesale Gas Purchasing

It is understood that Irish energy suppliers purchase or hedge a proportion of their NDM portfolios’ daily gas requirement in month to two-year ahead futures markets, and wait until GNI’s (or their own) next- or within-day forecasts before trading in next-day or spot markets to purchase the remaining balance. In general, these next- or within-day forecasts should be more accurate than seasonal forecasts, as they benefit from the availability of recent NDM market consumption data and near-time weather forecasts from the local meteorological service; whereas, monthly or longer gas contracts can only account for seasonal consumption or weather and are at risk from abnormal winters or consumer switching, for example.

![Chart showing daily system average prices (SAPs) and month contract prices](image)

**Figure 3.1:** Daily system average prices (SAPs) and month contract prices (for the corresponding date in the previous month) in the UK National Balancing Point’s (NBP’s) spot and month futures markets [19, 20].
In Figure 3.1, it is seen that any savings to be made from monthly contracts compared to spot markets, for example, is unclear as lower wholesale gas prices have alternated between these markets in recent years. However, this research is not concerned by these alternative purchasing strategies, as GNI and other European network operators are only required to provide suppliers with next- and within-day forecasts for their NDM consumer portfolios. This study therefore focuses on developing improvements to such forecasts based on the availability of smart metered data, as current forecast models are limited by the applied manual meter reading frequency.

### 3.1.2 FAR Procedures

These procedures describe how estimation and booking of an individual NDM consumer’s daily gas consumption is managed by GNI from the day before it is consumed through to their next meter reading. The procedure begins with next- and within-day forecast models to estimate total (or top-down network) gas consumption for the NDM market for a given day (D), based on readily available daily market consumption data.

These top-down market estimates are in turn apportioned to energy suppliers using individual consumer (or bottom-up) forecast models, based on each consumer’s monthly or longer meter readings. This forecasting process is illustrated in Figure 3.2, where it is seen that each NDM market forecast governs the sum of supplier forecasts, as it feeds into the bottom-up forecasting models in the calculation of a forecast AWDD (weather) parameter and a scaling factor that are described in detail later in this section.
Figure 3.2: Bottom-up NDM market forecasting procedure.

Once the NDM market’s consumption is known for the given day (D), it is apportioned to energy suppliers using the same individual consumer forecast models, as illustrated in Figure 3.3. It can be seen that the only difference between this and the forecasting process in Figure 3.2 is that actual NDM market consumption and AWDD values are applied instead of forecast values.
**Figure 3.3:** Bottom-up NDM market allocation procedure.

Of these FAR procedures, this research is primarily concerned by the Individual Consumer Models (Equation 3.10) used to allocate (forecasted or metered) daily NDM market consumption between energy suppliers. However, before these models can be described the NDM market forecasts by which they are governed are described below.
3.1.2.1 NDM Market Forecasts

GNI currently forecasts daily NDM market consumption and apportions it between energy suppliers in the market, once on the day before (as next-day forecasts) and several times during each forecast day (as within-day forecasts) [1]. These daily NDM market forecasts are initially estimated using statistical models on the day before and on the morning of the forecast day, before they are re-estimated by local experts using near-time information such as metered supplies to the network during the forecast day.

The statistical model used for the next-day market forecast is described in the FAR procedures as a function of actual NDM market consumption and weather data for recent days, and forecasted weather data for the current day and the forecast day, as follows [21]:

\[ \hat{C}_{NDM,D} = f(C_{NDM,D-n}, \ldots, C_{NDM,D-2}, W_{D-n}, \ldots, W_{D-2}, \hat{W}_{D-1}, \hat{W}_D) \]  \hspace{1cm} (3.1)

where: \( \hat{C}_{NDM,D} \) is the NDM market’s consumption forecast for the given day (D); \( C_{NDM,D-n}, \ldots, C_{NDM,D-2} \) are actual market consumption values for recent days; \( W_{D-n}, \ldots, W_{D-2} \) are (unspecified) actual weather values for recent days; and \( \hat{W}_{D-1} \) and \( \hat{W}_D \) are forecasted weather values for the current and forecast day.

It is understood that the statistical model for the within-day forecasts, on the morning of the forecast day, is simply an improved version of this next-day model based on more up-to-date gas consumption data. Although the next-day model has not been described in detail in the FAR procedures, alternative models have been published and these are summarised later in this literature review for a general overview of market or network gas consumption modelling.
The next step in the FAR procedures is to apportion these NDM market forecasts between the energy suppliers in the market using individual consumer models. However, before these models can be described the AWDD parameter on which they are based is described below.

### 3.1.2.2 Adjusted Weighted Degree Days

GNI has developed an AWDD variable to provide a complete explanation of the variation in daily NDM gas market consumption [21]. Although the AWDD variable is a gas consumption estimator based on degree-days (°C·day), it is not given by temperature values. It is instead calculated by back-solving the degree-day values required in a linear model of the daily NDM market consumption, so that estimates from this model equal the modelled consumption series. This calculation method begins by estimating the following model of NDM market gas consumption:

\[
C_{NDM,D} = b_0 + b_1 \text{HDD}_D + b_2 S-\text{HDD}_D + \epsilon_D
\]  

(3.2)

where: \( C_{NDM,D} \) is market gas consumption (kWh) for a given day \((D)\) in the previous year, which has been adjusted to reflect changes in the population of domestic and SME consumers since it was recorded [21]; \( b_0 \) is an estimate of weather-independent gas consumption (kWh); \( b_1 \) is the HDD coefficient (kWh/°C·day), the product of which is an estimate of weather-dependent gas consumption (kWh); \( b_2 \) is the seasonal-HDD (S-HDD) coefficient (kWh/°C·day), the product of which is an estimate of seasonal gas consumption (kWh); \( S-\text{HDD}_D \) is the thirty-year average HDD value for the corresponding date, which can (for example) counteract the effect of unexpectedly high
HDDs for times of the year when heating systems are usually not operated; \(\varepsilon_D\) is the model error for a given day; and each HDD is given by:

\[
HDD_D = \max(0; 15.5^\circ C - \bar{T}_{O,D})
\]  

(3.3)

where: \(15.5^\circ C\) is the assumed base temperature (see Equation 4.7) commonly applied in the UK and Ireland, or the outdoor temperature above which heating is not required; and \(\bar{T}_{O,D} (^\circ C)\) is the average outdoor temperature for the day.

It can be seen that the HDD and S-HDD estimators in Equation 3.2 can be replaced by or used to define what is referred to as a Weighted-DD (WDD) variable as follows:

\[
c_{NDM,D} = b_0 + (b_1 + b_2)WDD_D + \varepsilon_D
\]  

(3.4)

\[
WDD_D = \frac{b_1}{(b_1+b_2)}HDD_D + \frac{b_2}{(b_1+b_2)}S\cdot HDD_D
\]  

(3.5)

AWDDs are the back-solved (or adjusted) WDDs required in Equation 3.4 that result in estimates equal to the modelled consumption series, and are given by [21]:

\[
AWDD_D = \frac{c_{NDM,D} - b_0}{b_1+b_2}
\]  

(3.6)

It is seen that zero residuals will result in the NDM market consumption model in Equation 3.4 if AWDDs are used in place of WDDs. The AWDD parameter therefore provides correct estimates of weather-dependent gas consumption for the NDM market.
Forecast-AWDDs on the other hand differ from the actual-AWDDs above, as they are calculated based on NDM market forecasts rather than actual NDM market consumption, as follows:

$$AWDD_D = \frac{\hat{C}_{NDM,D} - b_0}{b_1 + b_2}$$

(3.7)

where: $AWDD_D$ is the forecast-AWDD for a given day; $b_0$, $b_1$ and $b_2$ are the same coefficients in Equation 3.2; $\hat{C}_{NDM,D}$ is the NDM market consumption forecast for the day; and it is seen that any error in $AWDD_D$ is dictated by the error in $\hat{C}_{NDM,D}$.

### 3.1.2.3 Individual Consumer Models

Individual consumer models are required in the FAR process before each NDM market forecast can be apportioned between the market’s energy suppliers. These models are based on AWDDs and monthly or longer meter readings, and are used to estimate the daily gas consumption of a consumer between meter readings. In the FAR procedures these models are allocated to each energy supplier according to the current consumer-supplier register. The individual consumer modelling process begins by estimating the following regression model for each consumer in the NDM market:

$$C_M = b_0 Days_M + b_1 \sum AWDD_M + \epsilon_M$$

(3.8)

where: $C_M$ is the consumer’s metered (M) gas consumption (kWh) between each meter reading or for each metered period; $b_0$ is an estimate of the consumer’s daily base or weather-independent gas consumption (kWh/day); $Days_M$ is the number of days for each metered period; $\sum AWDD_M$ is the sum of AWDDs for each metered period and an
estimator of the consumer’s weather-dependent gas consumption for the metered period using the $b_1$ coefficient (kWh/°C·day); and $\varepsilon_M$ is the model error for each metered period.

It is seen that the $b_0$ and $b_1$ coefficients resulting from this model may be used to estimate the daily gas consumption for a consumer, as follows:

$$\hat{C}_D = b_0 + b_1 AWDD_D$$  \hspace{1cm} (3.9)$$

where: $\hat{C}_D$ is the estimate of a consumer’s gas consumption for a given day (D), and $AWDD_D$ is the AWDD forecast for the day given by Equation 3.7.

However, this assumes that a consumer’s gas requirement is the same irrespective of the day of week. This issue is addressed in the FAR procedures by the following model:

$$\hat{C}_D = \begin{cases} 
(b_0 + b_1 AWDD_D)(DoW_{WD})(SF_D); & \text{on weekdays,} \\
(b_0 + b_1 AWDD_D)(DoW_{WE/Hol})(SF_D); & \text{on weekends/holidays.}
\end{cases}$$  \hspace{1cm} (3.10)

where: $DoW_{WD}$ and $DoW_{WE/Hol}$ are day of week adjustment factors for weekdays and weekends or holidays, which are given as 0.96 and 1.10 respectively for domestic consumers, and 1.09 and 0.79 respectively for SME consumers [21]; and $SF_D$ is a scaling factor used to ensure that the sum of individual consumer forecasts equals each NDM market forecast, as seen in Figure 3.2 and as described later in Equation 3.12.
3.1.2.4 Supplier Forecasts

Each NDM market forecast is apportioned to each supplier as follows:

\[
\hat{C}_{SI,D} = \left\{ \left( \sum_{j=1}^{m} b_{0,j} + \left( \sum_{j=1}^{m} b_{1,j} AWDD_{D} \right) DoW_{Dom.} \right) \right. \\
\left. + \left( \sum_{k=1}^{n} b_{0,k} + \left( \sum_{k=1}^{n} b_{1,k} AWDD_{D} \right) DoW_{SME} \right) \right\} SF_{D} \tag{3.11}
\]

where: \( \hat{C}_{SI,D} \) is the forecasted gas consumption to be procured by supplier \((i)\) for a given day \((D)\); \( b_{0,j} \) and \( b_{1,j} \) are the model coefficients for domestic consumer \((j)\); \( m \) is the number of domestic consumers in the supplier’s portfolio; \( b_{0,k} \) and \( b_{1,k} \) are the model coefficients for SME \((k)\); \( n \) is the number of SMEs in the supplier’s portfolio; and \( DoW_{Dom.} \) and \( DoW_{SME} \) are the relevant domestic and SME day of week adjustment factors in Equation 3.10.

Such forecasts are issued to each supplier, who can then use this information to procure gas in the wholesale energy market to meet their customers’ daily gas consumption. The scaling factor applied in this model and the individual consumer model in Equation 3.10 is given by:

\[
SF_{D} = \frac{\hat{C}_{NDM,D}}{\sum_{i=1}^{n} \hat{C}_{SI,D}} \tag{3.12}
\]

where: it can be seen that the product of the resulting scaling factor and the aggregation of \((n)\) supplier forecasts equals the NDM market forecast. It is through this scaling
factor that the NDM market forecast governs the volume of gas delivered to the network each day.

### 3.1.2.5 Supplier Allocations

Once the NDM market’s consumption is known for the given day it is apportioned to each supplier as follows:

\[
\hat{C}_{Si,D} = \left\{ \left( \sum_{j=1}^{m} b_{0,j} + \left( \sum_{j=1}^{m} b_{1,j} \cdot AWDD_D \right) \right) \cdot DoW_{Dom.} \right\} 
\]

\[
+ \left\{ \left( \sum_{k=1}^{n} b_{0,k} + \left( \sum_{k=1}^{n} b_{1,k} \cdot AWDD_D \right) \right) \cdot DoW_{SME} \right\} \cdot SF_D \]

where: \( \hat{C}_{Si,D} \) is the gas consumption allocated to supplier \( (i) \) for a given day \( (D) \); \( b_{0,j}, b_{1,j}, m, b_{0,k}, b_{1,k}, n, DoW_{Dom.} \) and \( DoW_{SME} \) are as before in Equation 3.11; \( AWDD_D \) is the actual AWDD for the day given by Equation 3.6; and where \( SF_D \) is now given by:

\[
SF_D = \frac{C_{NDM,D}}{\sum_{i=1}^{n} \hat{C}_{Si,D}} \tag{3.14} \]

where: \( C_{NDM,D} \) is the NDM market’s gas consumption for the day.

### 3.1.2.6 Balancing Charges

GNI applies a balancing charge to negative differences in a supplier’s forecasted (Equation 3.11) and allocated (Equation 3.13) consumption for a given day. This is to
recoup the cost of the extra gas required to maintain network pressure during the day, and is charged to the supplier at either the system average price (SAP) (see Figure 3.1) and system marginal price (SMP) for the day in the UK NBP market, depending on the relative difference [1]. Similarly, positive differences in forecasted and allocated gas consumption are credited back to the supplier.

3.1.2.7 Reconciliation

Each time a meter reading is recorded for a consumer, the final step in these FAR procedures is to reconcile the difference between the consumption allocated to the consumer and their metered consumption since their previous meter reading. This difference is given by:

$$\Delta C_M = C_M - \sum_{D=1}^{N} (b_0 + b_1 AWDD_D) DoWD_D SF_D$$  \hspace{1cm} (3.15)

where: $\Delta C_M$ is the difference to be reconciled (kWh), and $N$ is the number of days in the metered period.

3.1.2.8 Reconciliation Charges

GNI applies a reconciliation charge to differences in an individual consumer’s allocated and metered consumption between meter readings. This difference is calculated using Equation 3.15, and the value of this gas is reimbursed to GNI or their energy supplier as appropriate at the mean-SAP in the metered period.
This research develops improvements to the individual consumer models in Equation 3.10 based on the availability of smart metered daily gas consumption data and the review of international practice in this regard later in Section 3.2.4. These research opportunities are investigated in Chapter 7 using the SME smart metering dataset. Such improved models will allow suppliers to purchase gas in the wholesale market more accurately and this can result in reduced balancing charges.

3.1.3 Peak-Day Forecasts

GNI currently forecasts peak-day gas consumption for the NDM market in advance of each winter. This estimate is used by GNI to assess the adequacy of their network for potential extreme cold weather periods, in order to inform year-ahead network operations and to fulfil regulatory requirements such as ‘winter outlook’ reports to relevant stakeholders including the energy regulator and energy suppliers.

In addition, this estimate is used to establish network capacity bookings and charges to energy suppliers for the delivery of the gas network infrastructure used to supply their consumers. In this regard, GNI has therefore developed a transparent peak-day gas consumption estimation methodology.

3.1.3.1 Peak-Day Forecast Procedure

This procedure begins by developing a composite weather variable (CWV) to model daily NDM market gas consumption for the previous year [11]. This CWV allows multiple weather effects on gas consumption to be estimated using a single parameter and is comprised of the following variables:
• a seasonal normal (30 year average) HDD,

• the difference between this value and the HDD,

• the same difference again, if the HDD is above a reference value,

• a wind chill (WC) function of HDD and wind speed,

• the multiple of seasonal normal HDD and HDD; and finally

• the HDD and its lagged values for the two days previous.

This single parameter or CWV is then calculated for each day in a long-term climate dataset and its 1-in-50 year extreme value is extrapolated from this series using a statistical model. The extreme value modelling method is described later in Section 6.2.2. The resulting extreme CWV value is used to quantity the coldest day in which gas supplies are to be maintained to NDM consumers. The corresponding peak-day gas forecast is estimated using this extreme CWV value and the latest regression model of daily NDM market gas consumption.

Because NDM market consumption data are available for the extreme cold weather periods in Ireland during January and December 2010, there is an opportunity to develop improvements to this peak day forecasting method, as the accuracy of the regression model can now be quantified for peak (or extreme cold weather) consumption values. This opportunity is used to develop an adapted HDD variable as an estimator of both normal and extreme cold weather gas consumption.
This variable is used to investigate the relative difference in peak supply capacity necessitated by alternative supply standards. For example, the EU’s supply standard differs to the above forecast procedure in a number of ways:

1. peak gas consumption is estimated by temperature rather than a CWV;

2. extreme weather is quantified by a 1-in-20 rather than a 1-in-50 year return level; and

3. the duration of the extreme weather event is seven days rather than a single day

This investigation is also used as an opportunity to review international practice with regard to peak supply capacity standards, weather parameters and the modelling methods applied by European network operators to estimate peak gas consumption.

### 3.2 International Practice

Literature relating to the above NDM market forecasting requirements is summarised in this section. This begins with an overview of the simple weather variables found in the literature, before a detailed description of the CWV applied in the UK. This is followed by a summary of modelling methods used to forecast daily NDM market or network gas consumption. Next the individual consumer models applied to the UK and other European markets are described. The review is then completed by a summary of European peak-day modelling methods.
3.2.1 Simple Weather Variables

There are numerous simple weather variables that can be used to estimate gas consumption. For example, the following estimators have been recommended by a gas industry publication: temperature, wind speed, cloud cover, (solar) radiation, snowfall and rainfall [22]; and the application or instances of such estimators has been recently summarised for academic publications [23]. This summary is shown for daily gas forecasting models in Figure 3.4, where it is seen that HDDs, recent temperatures and wind speeds are the most frequently applied estimators.

![Diagram of Simple Weather Variables](image)

**Figure 3.4:** Instances of simple weather variables in published daily forecast models [23], where: $T$ are temperature values for the forecast day (D) and previous days (D-1 and D-2), and $WS$ are wind speed values.

However, instead of numerous simple weather variables such as those reported above, CWVs have been developed in the UK and Ireland that allow multiple weather effects on gas consumption to be estimated using a single parameter. These CWVs can be easily applied to estimate daily, seasonal and peak NDM market consumption to inform
gas purchasing and network planning. The Irish CWV has been previously described in Section 3.1.3.1. The UK-CWV is described in detail below, because it has a linear relationship with gas consumption that greatly simplifies gas consumption modelling [24], and because it is used to develop improvements to the HDD variable in Chapter 4.

3.2.2 UK-CWV

The UK-CWV can be calculated using Equations 3.16 - 3.20. It should be noted that these formulae use modified nomenclature to that published [24] in order to ensure consistency with this thesis. The calculation of the UK-CWV begins with the following composite weather term [24]:

\[
CW_D = \omega_1 T_{EFF,D} + (1 - \omega_1) S_{EFF,D} - \gamma_2 W_C_D
\] (3.16)

where for each day (D): CW is a composite weather (temperature) value estimated using the coefficients \(\omega_1\) and \(\gamma_2\); \(T_{EFF,D}\), is the effective (outdoor) temperature; \(S_{EFF,D}\) is the seasonal effective temperature; and \(W_C\) is a wind-chill function.

Such effective temperatures are used to account for the lag in response of daily gas consumption to current and preceding days’ temperatures, and in the UK-CWV this temperature is given by [24]:

\[
T_{EFF,D} = 0.5T_D + 0.5T_{EFF,D-1}
\] (3.17)

where: \(T_D\) is the (weighted average) temperature for the (gas) day, \(T_{EFF,D-1}\) is the effective temperature for the previous day (D-1).
It can be seen by expanding this function over a short interval that it is an exponential filter or weighted sum of recent temperature values:

\[
T_{\text{EFF}, D} = 0.5T_D + 0.25T_{D-1} + 0.125T_{D-2} + 0.0625T_{D-3} + 0.03125T_{D-4} + \cdots
\] (3.18)

The seasonal effective temperature in the above composite weather term is calculated by averaging and smoothing the effective temperature for each day of the year over a number of years, and adjusting the resulting annual profile so that it has an improved correlation to past consumption and is more responsive to temperature warming in spring and cooling in the winter [24]. Such annual profiles can be used to account for the response of consumers’ to unseasonable weather conditions whereby seasonal or normal gas consumption levels are maintained for the time of year.

The wind-chill function in the above composite weather term is used to account for air-infiltration heat loss from buildings and is given by [24]:

\[
WC_D = \max(0, WS_D - WS_B)\max(0, T_B - T_D)
\] (3.19)

where: \(WS_D\) is the wind speed for the day; \(WS_B\) is a base wind speed parameter; and \(T_B\) is a base temperature parameter in a degree-day type variable.

The final stage in the calculation of this CWV is used to transform the non-linear relationship between daily gas consumption and the above composite weather term to a linear relationship, as follows [24]:
\[ CWV_D = \begin{cases} 
T_{HI} + s_2(T_{UPR} - T_{HI}) & \text{if } T_{UPR} < CW_D \text{ (summer cut-off)} \\
T_{HI} + s_2(CW_D - T_{HI}) & \text{if } T_{HI} < CW_D < T_{UPR} \text{ (transition)} \\
CW_D & \text{if } T_{LWR} \leq CW_D \leq T_{HI} \text{ (normal)} \\
CW_D + s_1(CW_D - T_{LWR}) & \text{if } T_{LWR} > CW_D \text{ (cold weather upturn)} 
\end{cases} \tag{3.20} \]

where: \( T_{LWR}, T_{HI} \) and \( T_{UPR} \) are used to transform each \( CW_D \), and are temperature parameters that specify ‘cold weather upturn’, ‘normal’, ‘transition’ and ‘summer cut-off’ regions in the relationship between gas consumption and \( CW_D \), as illustrated in Figure 3.5; and \( s_1 \) and \( s_2 \) are slope parameters that are also used to transform each \( CW_D \).

\[ \text{Figure 3.5: Interpretation of the UK-CWV’s transformation function using gas consumption plots adapted from [24].} \]
Although the estimation method for the (nine) $T_{LWR}$, $T_{HI}$, $T_{UPR}$, $s_1$, $s_2$, $T_B$, $WS_B$, $\omega_1$ and $\gamma_2$ parameters has not been found, it is known that they are estimated to establish an optimal linear relationship between the resulting CWV and gas network consumption [24]. This CWV is used in the UK’s peak gas consumption forecasting method [24] and individual consumer models [25]. Elements of this variable are accounted for in the development of the adapted HDD variables in Chapter 4.

### 3.2.3 NDM Market Forecasting

Instead of relying on a single model to forecast daily NDM market or network gas consumption, it has been found that some network operators apply combination models [26, 27]. Such models are used to estimate the optimum weighted average of numerous forecasts from alternative models. For example, daily gas consumption forecasts have been estimated in the UK using a weighted average of forecasts given by [26]:

- two Box-Jenkins models,

- a Bayesian model,

- either a winter or summer linear regression model,

- a neural network model, and

- an expert system.

The benefit of this approach is that if a model performs poorly it will only have a small influence on the final forecast [26]. Instead of an extensive review of such methods, this
general overview of NDM market forecasting presents example regression and neural network models as these are the most commonly published daily forecast models [23]. An expert system is also described to provide insight as to the methods employed by local experts to estimate gas network or NDM market forecasts, such as those referred to previously in Section 3.1.2.1. These reviews demonstrate the relative simplicity of CWVs compared to applying numerous gas consumption estimators. Example Box-Jenkins or autoregressive integrated moving average (ARIMA) time series models are described in the review of individual consumer models in Section 3.2.4

### 3.2.3.1 Regression Model

A multivariable regression model has been developed to forecast next-day gas consumption in Slovenia [28]. The model was developed using a stepwise regression procedure, which determined the optimal subset of estimators from a larger model of possible estimators. The selected estimators included: a weekly gas consumption index (or for consistency with this thesis, a DoW factor) for the next-day (day, D); measured consumption for the previous day (day, D-2) adjusted by the ratio of DoW factors for the next- and previous day; and hourly temperature forecasts for 12, 30, 36, 42 and 54 hours following the next-day forecast. These DoW factors are given by [28]:

\[
D_{oW_D} = \frac{1}{N} \sum_{n=1}^{N} \frac{C_{D+7n}}{\sum_{k=-3}^{3} C_{D+7n+k}}
\]

(3.21)

where: \(D_{oW_D}\) is the day of week factor for a given day, and \(N\) is the number of weeks in the model estimation data.
The model was estimated using daily consumption data for two heating periods during September 2005 to March 2006 and February 2007 to May 2007 (approximately). The in-sample error of this model was given as 1.5%, and was calculated using a modified-MAPE metric which applied the maximum capacity of the network as its denominator.

The stepwise regression method was also used to develop multivariate linear regression models to forecast daily gas consumption for a distribution network and an individual dwelling in Croatia [29]. It was found that the inclusion of solar radiation significantly improved the accuracy of both the distribution network and dwelling regression models. In both cases the most accurate model accounted for previous consumption values (or lagged dependent variables), and future and previous temperature and solar radiation values; while the network model also applied a DoW factor similar to Equation 3.21.

These models were estimated using daily consumption data for the heating period during November 2011 to April 2012. The in-sample error of the distribution network model was 1.17%, and was calculated using a modified-MAPE metric similar to that applied to the Slovenian model above. The in-sample error of the individual dwelling model was 3.25%, although the denominator for this modified-MAPE metric was not reported. These models compared favourably to a simple linear regression model based on temperature, which had in-sample errors of 4.46% and 5.81% for the distribution network and individual dwelling models, respectively.

Although this Croatian regression model accounts for the additional effect of solar radiation, neither this nor the above Slovenian regression model accounted for the non-linear relationship between gas consumption and weather as demonstrated in Figure 3.5 for the UK-CWV, because these models were applied to heating season rather than
annual data; nor do they account for important effects such as wind-speed and seasonal consumer behaviour that are accounted for by both the Irish and UK CWVs. Such effects and the additional effect of solar radiation are addressed in the derivation of the adapted HDD variable for gas network modelling in Chapter 4.

### 3.2.3.2 Neural Network Models

A neural network model has been developed to forecast daily gas consumption in Istanbul [30]. This model was estimated using a quick-propagation training algorithm and included the estimators: consumer population, each day of the week, workdays, holidays, month, year, consumption for the previous day, and minimum and maximum temperatures for the previous and forecast day. The in-sample MAPE for this model was 5.9%, and this was shown to be an improvement on several other models that were estimated using alternative training algorithms.

An alternative neural network model has been developed for the metropolitan region of Milwaukee in Wisconsin, US [31]. This model was used to estimate both the current and next-day’s gas consumption at a time ‘slightly’ before the current day. This two-day forecast model was estimated using a Kalman-Filter training algorithm and included the estimators: HDDs forecast for the current day (D); wind speed and sunshine forecasts for the current and next-day (days, D-1 and D); HDD and consumption values for previous days (days, D-2, D-6 and D-7); day of the week; day of the year; tap water temperature for the current day; and forecasts of an alternative quarter-HDD parameter (based on six hourly temperature values) for each quarter of the current and next-day. The accuracy of this model was shown to be an improvement on a linear regression model using the same estimators and two simpler neural network models.
Although it would appear that neural networks offer improved forecasting accuracy over regression modelling methods, this ‘black box’ approach is much less transparent than the regression methods applied in this research and by GNI to manage NDM market gas consumption. This is an important consideration when there are stakeholders such as the energy regulator and energy suppliers who are concerned with methodological transparency.

3.2.3.3  **Expert System**

An expert system (or programme) has been developed in the UK to emulate the forecasting methods of regional experts [32]. This system was developed based on structured interviews which found (amongst other information) that regional experts applied a regression model and an ‘effective temperature’ of some form (see Equation 3.17) and agreed that after temperature, the type of day and wind are the most important factors affecting gas consumption.

Based on the results of these interviews a model was developed to forecast next-day gas consumption, as follows: 1) consumption for the current day is adjusted by the forecasted difference in consumption for the next day, given by the corresponding difference in effective temperature and its slope coefficient from a separate linear consumption model; and then scaling the result by a series of percentages to account for: 2) the change in wind speed between the days; 3) the next-day’s type of day: weekday, Friday, Saturday, Holidays etc. and 4) ‘misery’ factors to account for effects such as snow, heavy rain and drizzle.

It was found that this expert system was approximately 10% and 47% more accurate than regional experts and the current regression model, respectively. The main benefit
of this method or regional experts’ forecasts is that they can easily account for such ‘misery’ factors, whereas it is difficult to estimate such effects using regression models. For example, there may be limited instances of snowfall in the modelling data in which to estimate a statistically significant coefficient for its effect on gas consumption. This also illustrates the benefit of the scaling factor applied in GNI’s FAR procedures in Equation 3.11 as this allows regional experts’ forecasts to be applied within the forecasted day.

3.2.4 Individual Consumer Models

Alternative models used to forecast the daily gas consumption of individual consumers are described in this section. This begins with by reviewing models applied by European network operators based on monthly or longer gas consumption data, followed by a description of the latest models based on daily gas consumption data.

3.2.4.1 Network Operator Models

In the UK, individual consumer forecasts are estimated using an annual load profile method [25]. In this method, the weather-corrected annual quantity of gas is calculated for a consumer using monthly or annual gas consumption meter readings, for the previous gas year, October to September [25, 33]. For an initial forecast of gas consumption, the mean daily value of this annual quantity is multiplied by the load profile value for the given day and for the relevant consumer category. These consumer categories are based on various annual gas consumption levels and the distribution of this consumption across the year [34]. Each load profile is calculated by dividing the seasonal demand series (or normal demand for each day of the year) for a consumer category by its mean daily value [25]. The initial forecast of gas consumption is then
modified by: 1) an adjustment factor for the given day and for the relevant consumer category, 2) a weather correction factor for the relevant region, and finally 3) a scaling factor, as similarly applied in the Irish method.

In the Czech Republic and Slovakia, individual consumer forecasts are estimated using a generalised additive model [35], which has been developed based on numerous modelling methods, previously developed for these markets [36, 37]. In this method [35], the expected mean daily quantity of gas is calculated for a consumer using annual gas meter readings for the previous three years. Daily forecasts are then estimated for this consumer, by multiplying this mean value by a time-varying function that models the effect of temperature, the type of day, and Christmas and Easter holiday periods for the relevant consumer category.

These consumer categories are distinguished by characteristics such as domestic or SME consumer types, and alternative gas requirements related to space heating, hot-water, cooking or industrial production. The required time-varying function is fitted separately using daily average gas consumption data for a cluster of consumers, previously grouped together from a sample of continuously metered consumers using hierarchical agglomerative clustering methods. In each time function, day-type effects are modelled using a categorisation scheme that defines each day, by workday or non-workday classifications for the previous, current and next day. Temperature effects are modelled using a response function that accounts for day-type, prevailing temperatures and a temperature transformation that is estimated using a LOESS (or local polynomial regression) smoother.
Each of these above modelling methods is limited by the meter reading frequency applied by the local TSO. In the UK, Czech and Slovakian gas markets these meter readings may only be once per year; and, as a consequence, these markets have developed load profiles or time-functions which assume that the daily gas consumption of a consumer will follow that expected for its relevant consumer category. Although at least four bi-monthly meter readings are recorded per year in Ireland (see Table 3-1), it was seen that the current FAR procedures requires that the same day of week factor is applied to all consumers in each domestic or SME consumer category (see Equation 3.10), irrespective of its applicability at an individual level. However, as a result of smart metering, such assumptions do not have to apply to the individual consumer since models can be estimated at an individual enterprise level.

### 3.2.4.2 Daily Data Models

Examples of individual consumers models based on daily gas consumption data are uncommon [38], possibly due to a paucity of suitable datasets. A stepwise regression model was previously described in Section 3.2.3.1 for a single (continuously heated) dwelling in Croatia, along with a similar distribution network model. And a non-linear model has been developed using daily consumption data recorded at an individual buildings in the Czech Republic [38]. This is a mixed-effects model that combines a conditional model of consumption with a model for the marginal probability of zero consumption.

The conditional model of consumption applies a Gumbel distribution function to account for the non-linear gas consumption response to outdoor temperatures for the current and previous day, and includes parameters to account for variation in gas
consumption for different days of the week – holidays are not addressed for parsimony and due to their limited occurrence during the modelled heating season. This model is based on multiplicative exponential terms, which it is understood, are fitted using a log-transformation that only permits positive consumption values. The probability of zero consumption is then estimated by a logistic regression model of temperature and the day of the week.

Substantial computational (or convergence) problems were encountered when an attempt was made to include a first-order autocorrelation error term in the conditional model for improved model accuracy. This mixed effect model delivered comparable results to benchmark models based on outdoor temperatures for the current and previous day and either lagged dependent (consumption) variables or an ARIMA error model to account for auto-correlated errors.

Although the Croatian regression model in Section 3.2.3.1 accounts for the effect of solar radiation, and the Czech mixed-effects model accounts for the non-linear relationship between gas consumption and recent outdoor temperatures, neither of these models account for both of these effects, nor do they account for the important effect of wind speed that is accounted for by both the Irish and UK CWVs.

3.2.5 Peak-Day Forecasting

Alternative methods applied by European network operators to quantify peak-day gas consumption are summarised in this section. This begins with a review of a sample of supply standards, followed by a description of current peak gas consumption modelling methods.
3.2.5.1 Supply Standards

In Table 3-3, the EU’s new supply standard referred to in Section 1.2.2 and a sample of European network operators’ supply standards have been summarised with respect to the criteria used to describe a peak consumption event, the weather parameter used to quantify this event, and additional information in relation to the methodology used to estimate the weather parameter’s return level and peak gas consumption.

The EU’s new gas supply standard was developed since the gas crisis in 2009 [39], when Central and Eastern European countries experienced significant gas shortages [40]. In the table, the peak gas consumption criterion of the standard has been summarised; two additional criteria in relation to storage capacity requirements complete the standard but these are not referred to here. The peak gas requirement in the EU standard is the minimum short-term quantity of gas supply that network operators in member states, must make available to ‘protected’ (mainly domestic) consumers.

However, it can be seen in Table 3-3 that some countries use a longer return period than the EU requirement but for shorter consumption periods. It is understood that this is the case in France [45], the Netherlands [46] and Ireland. In Ireland, the relatively high 50 year return period is used due to the country’s limited storage and poor interconnectivity with other gas networks.
Table 3-3: European peak gas supply standards

<table>
<thead>
<tr>
<th>Region</th>
<th>Consumption Criteria</th>
<th>Notes</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Union</td>
<td>Extreme temperatures during a 7 day peak period occurring with a 1-in-20 year probability.</td>
<td>These extreme temperatures have been quantified for the Belgian market as a 7-day average temperature that is equivalent to the 1-in-20 year 7-day heating degree day total.</td>
<td>[10, 41]</td>
</tr>
<tr>
<td>Belgium</td>
<td>5 consecutive days between -10 and -11°C.</td>
<td>-</td>
<td>[41]</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>5 consecutive days when the average daily temperature does not rise above -14°C</td>
<td>-</td>
<td>[42]</td>
</tr>
<tr>
<td>Denmark</td>
<td>3 consecutive days with a daily average temperature of down to -13°C i.e. the 1-in-20 year event</td>
<td></td>
<td>[43]</td>
</tr>
<tr>
<td>France</td>
<td>Very low temperatures over 3 consecutive days with a 1-in-50 year probability</td>
<td>These very low temperatures have been quantified by an effective daily temperature as follows:</td>
<td>[44, 45]</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{EFF,D}}=0.64\bar{T}<em>D+0.24\bar{T}</em>{D-1}+0.12\bar{T}_{D-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>where: $T_{\text{EFF,D}}$ is the effective temperature for a given day ($D$); and $\bar{T}<em>D$, $\bar{T}</em>{D-1}$ and $\bar{T}_{D-2}$ are the average temperatures for day ($D$), and the preceding days ($D-1$ and $D-2$), respectively.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A 1-in-50 year estimate of this effective temperature is provided by Météo France, which also accounts for climate change, using 30 years of temperature data.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In Table 3-3 it is seen that an ambient air temperature variable is the most common weather parameter used to estimate peak gas consumption. However, more complex estimators of gas consumption are also applied. For example, CWVs are applied in Ireland and the UK, and an effective temperature is applied in France that is calculated using a weighted temperature filter to account for the lag in response of daily network gas consumption using the current and two preceding days’ temperatures.

### 3.2.5.3 Peak Consumption Criteria

The supply standards in Table 3-3 use a variety of alternative definitions to quantify short-term peak consumption. Apart from in the UK, a peak consumption event is described by a weather parameter of some form, quantified by a return level for a single day or for several days.
The method chosen for estimating return levels depends on the specified weather parameter and the length of the available climate data – long-term climate data can allow simple empirical estimation, while shorter datasets may require statistical extreme value methods. In this regard, the use of temperature variables to estimate consumption has the advantage that long-term temperature datasets are readily available from local meteorological stations. However, in Ireland the 1-in-50 year CWV is extrapolated using an extreme value model, as records for both temperature and wind speed from the required weather station are only available since the 1940s.

### 3.2.5.4 Modelling Techniques

Although it is not clear in the literature, it appears that regression-based methods are used for the estimation of peak consumption in the French [45] and the UK [24] gas markets. For the UK gas market, peak gas consumption is forecasted as the mean of multiple 1-in-20 year return levels estimated by Gumbel-Jenkinson extreme value models of simulated long-term gas consumption series. These simulated gas consumption series are created using a model of daily gas consumption employing historical CWV values and random error terms [24].

### 3.3 Benchmarking

This section provides an overview of current methods used to benchmark the energy efficiency of buildings using metered energy consumption data, so that the new statistical benchmarking tool can be developed in Chapter 5. Although a wide variety of benchmarking tools are currently available, it has been found that these apply HDDs in almost all cases.
For example, the US Environmental Protection Agency (US-EPA) has developed an Energy Star Score system for a range of commercial buildings that applies a regression based benchmarking tool [47]. The first step in this scoring system calculates an energy efficiency ratio for a building by dividing its annual energy use intensity (both electricity and gas) by that predicted by a regression model for the building type [47]. For example, the regression model applied for multifamily housing (or apartment) buildings has been fitted using a reference dataset of such buildings and is based on the number of dwellings per 1000ft², the number of bedrooms per dwelling, the total HDDs and cooling degree days for the year, and the number of levels in each building [48]. The probability or percentile of the building’s energy efficiency ratio is then found using a lookup table developed using energy efficiency ratios for the reference dataset [47]. The Energy Star Score for the building is 100 minus this percentile value. For example, a building with an Energy Star Score of 75 is bettered by only 25% of the reference dataset.

Home Energy Yardstick is an online tool that has been developed as part of the US-EPA’s Energy Star program [49]. This tool benchmarks domestic building energy efficiency using a 1 to 10 scoring system, where a score of 10 represents a home with the best energy efficiency level [49, 50]. This score is based on a statistical method and requires users to provide utility bill consumption data for electricity and gas, and their building’s location, floor area and number of occupants [49, 50]. Energy suppliers in the US are encouraged to host this tool on their own websites [51].

In Europe, a Display Energy Certificate system is applied to large public buildings. These certificates are also based on metered energy consumption and building floor area and are used to present a building’s annual energy use intensity (kWh/m²/year) on an
A1 to G scale, where an E1 rating corresponds to a typical building in the relevant building class [52]. These energy intensities are based on building floor area. Such normalised energy consumption parameters are a very common way of benchmarking building energy efficiency [53].

Each of the above benchmarking tools is based on energy intensity parameters normalised by building floor area, which presupposes that floor area data are readily available. However, it has been observed that many householders are unable to provide their building’s floor area when surveyed – 75% in the case of a previous Irish housing quality survey [54] and 59% in the case of the smart metering survey applied in Chapter 5. Accurate area data would therefore be difficult to collect for an energy supply company. Moreover, many variables other than floor area contribute to household energy use; these include occupancy patterns, no. of occupants and dwelling type (detached, semi-detached, etc.). These, too, should be considered in a comprehensive gas consumption benchmarking method.

3.4 Conclusion

This literature review began with a detailed description of the current methods used by GNI to forecast daily NDM market gas consumption. This included a summary of GNI's FAR procedures, the AWDD parameter and individual consumer models used to allocate daily NDM market forecasts between energy suppliers. International practice in this regard was also described. It was found that the current methods are limited by monthly or longer consumption data and consequently apply consumer category profiles or adjustment factors irrespective of their applicability at an individual level; and although the latest model accounted for the additional effect of solar radiation it does
not account for wind speed nor the non-linear relationship between gas consumption and recent weather.

It was found that such wind speed and non-linear effects are accounted for by the UK-CWV and elements of this variable are accounted for in the development of the adapted HDD variables in Chapter 4. One of these variables is used in development of improved individual consumer models in Chapter 7. It is expected that such models will help reduce balancing charges between GNI and energy suppliers operating in the NDM market and this can help increase competiveness.

Next, the current method used by GNI to forecast peak-day gas consumption for the NDM market was reviewed. This included a summary of the Irish CWV gas consumption estimator and the differences between the Irish and European peak capacity supply standards. It was found that none of these methods apply a weather parameter that accounts for solar radiation, even though this has been recently been shown to be an important estimator of domestic and network gas consumption. This solar radiation effect is addressed in the development of the adapted HDD variables in Chapter 4. The first of these variables is used to develop a more accurate NDM market consumption model in Chapter 6.

This literature review was completed by a summary of domestic energy efficiency benchmarking methods based on metered energy consumption data. It was found that many of these methods are based on energy intensity parameters normalised by building floor area, even though many householders are unable to provide their building’s floor area when surveyed. Therefore, an alternative regression-based benchmarking method is
developed in Chapter 5 for the consideration of energy suppliers who are now required by the EU-EEOS to assist consumer energy savings.
CHAPTER 4

HEATING DEGREE DAYS
4 HEATING DEGREE DAYS

This chapter describes the HDD methods used in this research. The first section provides an overview of HDD theory, the internationally accepted HDD formulae and the formula selected from these to develop improvements to standard HDD methods in this research. Traditional HDD building energy modelling methods are then described, and a simple HDD regression modelling method based on daily gas consumption data is developed for the benchmarking tool in Chapter 5. Next, the HDD variable is derived and this is used as a basis in which to develop adapted HDD variables. The first of these is used to develop a NDM market gas consumption model for peak-day forecasting in Chapter 6. The second is used in the development of individual consumer models in Chapter 7.

4.1 HDD Overview

The HDD variable is a parameter based on outdoor temperature data that is used to model a building’s weather dependent fuel consumption. It is based on the concept that the instantaneous heat demand for a building may be estimated as the product of the building’s overall heat loss coefficient and the temperature differential between the heated space and the surrounding environment, as follows [55]:

\[ D = HLC (\Delta T) \]  \hspace{1cm} (4.1)

where: \( D \) is instantaneous heat demand (kW); \( \Delta T \) is the temperature differential (°C); and \( HLC \) is the building’s overall heat loss coefficient (kW/°C), which includes both a fabric loss and an air-infiltration coefficient, given by:
\[ HLC = \left( \sum UA + 0.33NV \right) / 1000 \]  

(4.2)

where: \( \Sigma UA \) is the building’s fabric loss coefficient (W/°C); \( U \) is the \textit{U-value} of each building fabric element (W/m\(^2\). °C); \( A \) is the area of each building fabric element (m\(^2\)); \( 0.33NV \) is the building’s air-infiltration coefficient (W/°C); \( N \) is the estimated or documented air-infiltration test value of the number of air changes per hour for the building; \( V \) is the volume of the heated space (m\(^3\)); 0.33 is a factor used to convert the units of \( NV \) (m\(^3\)/h) to the same units as the fabric loss coefficient (W/°C) – 0.33 is the product of the density (1.2 kg/m\(^3\)) and specific heat capacity (1000 J/kg°C) of air, divided by the number of seconds in an hour (3600 s/h); and 1000 is a denominator used to convert the units of overall heat loss coefficient from W/°C to kW/°C.

In a building, this temperature differential (Equation 4.1) will vary with changes in internal and external temperatures resulting in a proportional change in heat demand. HDDs are used to estimate the integral (or sum) of this temperature differential over time, so that the fuel consumption of the building’s heating system may be estimated, as follows:

\[ F = HLC \left( \sum_{i=1}^{n} HDD_i \right) (24) \left( \frac{1}{\eta} \right) \]  

(4.3)

where: \( F \) is fuel consumption (kWh); \( n \) is the number of days in the relevant time period; \( HDD \) is the heating degree day parameter (°C·day); 24 is a conversion factor from kW·day to kWh units; and \( \eta \) is a conversion factor to fuel consumption units that is given by the efficiency of the building’s heating system (%).
This fuel consumption model and the HDD variable have been developed from the following building energy model, which is derived later in Section 4.3.1 [55]:

\[ E = HLC \int (T_B - T_O) \, dt; \text{ for } T_O < T_B \]  \hspace{1cm} (4.4)

where: \( E \) is the heat energy consumption for a building over time; \( T_B \) is the building’s base temperature parameter (°C), which defines the outdoor temperature above which heating is not required; and \( T_O \) is the outdoor temperature (°C).

In this model, the integral is formally known as ‘degree-time’ [56], and the units of this integration (°C·day or °C·h, typically) define the units of the estimated energy consumption – usually kWh of fuel as shown for example in Equation 4.3. The HDD variable is an estimate of the degree-time integral over a day and is calculated using various formulae as follows.

### 4.1.1 HDD Formulae

In general, HDDs are calculated and published by the local meteorological service using the established formula and the traditional base temperature adopted for that nation – for example, 15.5°C in the UK [55] and Ireland. In the UK, HDDs are calculated and published using ‘Meteorological Office’ formulae that assume a quasi-sinusoidal diurnal outdoor temperature profile [55] based on daily maximum and minimum temperatures, as follows [55, 57] and as illustrated in Figure 4.1:

**Case 1**: \( T_{O,\text{MAX}} \leq T_B \)

\[ HDD = T_B - 0.5(T_{O,\text{MAX}} + T_{O,\text{MIN}}) \]  \hspace{1cm} (4.5a)
Case 2: $T_{O,MIN} < T_B$ and $(T_{O,MAX} - T_B) < (T_B - T_{O,MIN})$

$$HDD = 0.5(T_B - T_{O,MIN}) - 0.25(T_{O,MAX} - T_B) \quad (4.5b)$$

Case 3: $T_{O,MAX} > T_B$ and $(T_{O,MAX} - T_B) > (T_B - T_{O,MIN})$

$$HDD = 0.25(T_B - T_{O,MIN}) \quad (4.5c)$$

Case 4: $T_{O,MIN} \geq T_B$

$$HDD = 0 \quad (4.5d)$$

where for each day: $T_{O,MAX}$ is the maximum outdoor temperature (°C) and $T_{O,MIN}$ is the minimum outdoor temperature (°C).

Figure 4.1: Illustration of each case in the ‘Meteorological Office’ formulae in Equation 4.5 using a base temperature of 13°C and hourly outdoor temperature data from Dublin Airport during 6th - 9th October 2011.
Instead of this four-case formula, HDDs may be calculated more simply using discrete time interval methods based on either hourly or daily outdoor temperature data as follows [55, 57]:

Hourly temperatures formula:

\[
HDD = \sum_{h=1}^{24} \{\max(0; T_B - T_{O,h})\} \left( \frac{1}{24} \right)
\]  

Daily average temperature formula:

\[
HDD = \max\{0; T_B - 0.5(T_{O,MAX} - T_{O,MIN})\} \approx \max(0; T_B - \bar{T}_O)
\]  

where: \(T_{O,h}\) is outdoor temperature (°C) at hour (h) in the day; and \(24\) is a factor used to convert the summation of heating degree hours to a HDD; and \(\bar{T}_O\) is the average daily outdoor temperature (°C).

In Figure 4.2, HDD values for two sample days have been calculated using the hourly temperatures formula of Equation 4.6 and the daily average temperature formula of Equation 4.7. It is seen that the degree-time integral of Equation 4.4 is estimated more accurately using the hourly temperature formula than the daily average temperature formula.

However, in Figure 4.3 it is seen that the daily average temperature formula produces a more representative model of daily building fuel consumption than the ‘Meteorological Office’ and hourly temperatures formulae, in that it is the only linear model which results in a highly significant intercept \((b_0)\) parameter with a p-value less than 0.001 and because its slope \((b_1)\) parameter has the smallest standard error. This can be attributed to the daily average temperature formula accounting for consumer behaviour, in that it
assumes that heating systems are not operated when the average outdoor temperature exceeds the building’s base temperature [55]; whereas for example, the hourly temperature formula can estimate positive HDD values for the same days when overnight there may be only a few positive degree hours and during which there is no consumer response.

![Hourly and Daily Temperature Diagram](image)

**Figure 4.2:** HDD calculations using hourly (top) and daily (bottom) outdoor temperature data from Dublin Airport on 23rd and 24th December 2011.

The daily average temperature HDD formula in Equation 4.7 is also the simplest basis in which to develop improvements to the HDD method either in the estimation of building specific base temperatures as seen in Section 4.2.1 or additional weather effects such as solar radiation as seen in Section 4.3.2.1. This formula is therefore used
throughout this research in preference to the alternative ‘Meteorological Office’ and daily temperatures formulae in Equations 4.5 and 4.6, respectively. It should also be noted that Equation 4.7 is generally used to calculate HDDs in countries outside of the United Kingdom and Ireland [55].

Figure 4.3: HDD regression models using ‘Meteorological Office’ (top), hourly temperatures (middle) and daily average temperature (bottom) formulae with an assumed base temperature of 15.5°C, and aggregated gas consumption data for the smart-metered domestic sample in Chapter 5.
4.1.2 Base Temperature

In the HDD formulae, it can be seen that the internal temperature record for the building is not required to estimate the temperature differential driving heat consumption. Instead, a base temperature is estimated as a constant parameter and is defined for an intermittently heated building as follows [55]:

\[ T_B = \bar{T}_I - T_G \]  

(4.8)

where: \( \bar{T}_I \) is the average daily internal temperature of the building (°C), and \( T_G \) is the equivalent temperature due to heat gains in the building (°C).

In Equation 4.7, this base temperature parameter is used to indicate the days when heating is not required and to transform outdoor temperature data so that the fuel consumption of buildings can be estimated using a linear model. However for simplicity, it is typically assumed that a building’s base temperature is given by the traditional value used by the local meteorological service to calculate published HDDs, rather than estimating it using the ‘trial and error’ methods such as those described in Section 4.2.1.

However, as the accuracy of HDD regression models are dependent on the base temperature parameter, HDDs are calculated in this thesis using outdoor temperature data and building specific base temperatures that are calculated using the NLS method described in Section 5.1.2.
4.2 HDD Building Energy Modelling

The HDD variable can be used to model the thermal energy performance of a building using direct and/or indirect modelling methods:

- **Direct modelling methods** are used to estimate the expected heat consumption of a building at design or renovation stages, using engineering models. Typically these models utilise the region’s HDD total for a normal year and data from design calculations such as the building’s overall heat loss coefficient, heating system efficiency, temperature set-point and time control settings for the heating system.

- **Indirect modelling methods** are used to estimate the thermal properties of an occupied building, such as the building’s overall heat loss coefficient, heating system efficiency and mean internal temperature, using regression based models. Typically these models are based on the building’s metered fuel consumption data and the corresponding HDD series for the region. These methods are described in further detail in Section 4.2.1 and are applied in Chapter 5 to develop the domestic gas end-use efficiency benchmarking tool.

- Together direct and indirect models can be used to inform building renovation options, in order to reduce fuel consumption. Indirect modelling methods are used to estimate thermal properties of the building, and these estimates are used by direct modelling methods to assess the potential impact of various upgrades to the building and its heating system.
4.2.1 Simple HDD Regression Models

In this study, simple HDD regression models are estimated for the smart metered domestic consumer sample and the resulting parameter estimates and their distributions are used to develop a gas end-use efficiency benchmarking tool in Chapter 5. Traditionally, such models have been based on the monthly or quarterly gas meter readings provided to the consumer by their utility supplier and can be estimated using an equation similar to the AWDD model in Equation 3.8, as follows:

\[ C_M = b_0 Days_M + b_1 \sum HDD_M + \varepsilon_M \]  

(4.9)

where the terms in this model have similar interpretations to those in Equation 3.8.

Such HDD regression models are generally fitted using published HDD data. However, if instead outdoor temperature data are applied and the individualised base temperature for the building is estimated, a more representative building energy model will result. Many calls have been made in this regard for the adoption of building-specific base temperatures [55].

Traditionally, the true base temperature for a building has been estimated using alternative ‘trial and error’ techniques for monthly or daily metered fuel consumption data [55]. For monthly data, a quadratic HDD regression model is applied that estimates a building’s base temperature by the value which yields a zero squared-HDD coefficient [58]. For daily metered data, a building’s base temperature is estimated either by: 1) visually identifying the point of inflection in a scatter plot of fuel consumption vs. outdoor temperature; or 2) the upper temperature limit in the data that yields the
maximum coefficient of determination ($R^2$ value) for a linear model of fuel consumption based on the lower temperatures [55].

However, daily data is widely available for domestic consumers from smart meters. Therefore, this study has developed a more direct method to estimate the $b_0$, $b_1$ and $T_B$ parameters of the HDD building energy model using daily metered data. This method is based on the following simple HDD regression model:

$$C_D = b_0 + b_1 HDD + \epsilon_D = b_0 + b_1 \max(0; T_B - \bar{T}_D) + \epsilon_D$$

(4.10)

where: $C_D$ is daily gas consumption (kWh) and $\epsilon_D$ is the model error for each day (D).

This expression permits the use of the iterative NLS computational method in Section 5.1.2 to estimate the base temperature for a building within (rather than separately to) a HDD regression model. This model is estimated for over 500 consumers in the smart metered domestic consumer sample in Chapter 5. The resulting $b_0$, $b_1$ and $T_B$ parameters can be interpreted as follows.

**4.2.1.1 Intercept parameter**

The intercept ($b_0$) parameter is the building’s daily base gas consumption, and for domestic consumers this is typically used for hot water and cooking purposes. Therefore, the $b_0$ parameter may be used to identify buildings in need of a hot water heating system upgrade or a reduction in hot water set-point temperature [59], or buildings using electrical water heating systems that could switch to gas fuelled systems.
4.2.1.2 Slope parameter

Based on Equation 4.3, the slope \( b_1 \) parameter is related to the building’s heat loss coefficient and heating system efficiency as follows [55]:

\[
b_1 \approx HLC \left( \frac{24}{\eta} \right)
\]  

(4.11)

and may be used to identify buildings in need of a building fabric or heating system upgrades [59].

4.2.1.3 Base temperature parameter

The \( T_B \) parameter is related to the average indoor temperature and useful heat gain in the building, as shown in Equation 4.8. This average temperature is in turn related to the building’s heating system set-point temperature, as illustrated in Figure 4.4. In this profile, it is assumed that the indoor temperature decreases during the building’s unoccupied period and increases to the required set-point temperature following a pre-heat period, before the return of the building’s occupants. Based on this profile, the average temperature in an intermittently heated building is given by [55]:

\[
\bar{T}_I \approx T_{SP}(On) + \frac{\sum_h^{(24-On)} T_{I,h}}{24}
\]  

(4.12)

where: \( T_{SP} \) is the heating system’s set-point temperature (°C) – which is assumed to be representative of the building’s indoor temperature during occupied periods; \( On \) is the number of heating system operating hours each day or the total pre-heating and
occupancy time; and $T_{I,h}$ is the indoor temperature at hour $h$ in the day when the heating system is off during unoccupied periods.

![Idealised indoor temperature profile for intermittently heated building](image)

**Figure 4.4:** Idealised indoor temperature profile for intermittently heated building [55].

The $T_B$ parameter may be used to assess a consumer’s thermal comfort requirement, as buildings with high base temperatures must respond to more HDDs during each heating season than those with lower base temperatures. This may either be the result of increased set-point temperatures and heating system operating hours or poor heat retention by the building. Such buildings are targets for behavioural programmes or improved heating system control systems, for example programmable thermostats [59].

### 4.3 HDD Parameter Derivations

Because the HDD parameter has been developed to model monthly or longer building fuel consumption using simple regression models, and is used to estimate annual
building fuel consumption using engineering models, it does not directly account for
daily building heat consumption dynamics, such as variations in:

1. solar heat gain,

2. air-infiltration, and

3. building fabric thermal storage, for example.

Instead, the HDD parameter addresses these effects using a number of simplifying
assumptions. This is an acceptable limitation in monthly or longer HDD regression
models, as these effects are not pronounced over such large time steps and may be
simply subsumed within either the estimated HDD coefficient or the base temperature
parameter, without a significant impact on modelling accuracy. The derivation of the
HDD variable in Section 4.3.1 is used to illustrate where simplifying assumptions have
been made in the standard method, so that adapted HDD variables can be developed
later in this chapter.

4.3.1 HDD Derivation

The derivation of the HDD variable is based on the following instantaneous heating
system energy balance, as illustrated in Figure 4.5:

\[ Q_{HS} = Q_{BF} + Q_{AI} + Q_{TS} - Q_{SG} - Q_{IG}; \text{ for } Q_{HS} > 0 \]  \hspace{1cm} (4.13)

where: \( Q_{HS} \) is the instantaneous heat gain from the building’s heating system (kW); \( Q_{BF} \)
is the heat loss through the building fabric (kW); \( Q_{AI} \) is the heat loss due to air
infiltrating/exiting the building (kW); $Q_{TS}$ is the heat gain/loss due to building thermal storage effects (kW); $Q_{SG}$ is the solar gain through windows (kW); and $Q_{IG}$ is the internal heat gain from lights, people and appliances (kW).

The derivation begins by estimating the combined effect of the instantaneous building fabric and air-infiltration heat loss components on heat demand, using an initial estimate of the temperature differential across the building’s envelope, as follows:

$$Q_{HS} = HLC(T_I - T_O) + Q_{TS} - Q_{SG} - Q_{IG}; \text{ for } Q_{HS} > 0$$  \hfill (4.14)

where: $T_I$ is the instantaneous indoor temperature ($^\circ$C) and $T_O$ is the instantaneous outdoor temperature ($^\circ$C).

**Figure 4.5:** Instantaneous energy balance of a building’s heating system.
However, the overall heat loss coefficient in this initial model is given as a constant value for each day, although it includes the effect of air-infiltration (see Equation 4.2), which is dependent on external wind speeds and is therefore a variable component of a building’s daily heat consumption. This simplification is the first of the primary assumptions used in the derivation of the HDD variable that will be addressed in order to develop improved estimators of daily gas consumption.

The derivation continues by integrating this energy balance model, to give an initial estimate of building heat consumption over time, as follows:

\[
E = \int Q_{HS} \, dt; \quad \text{for } Q_{HS} > 0
\]

\[
= HLC \int (T_I - T_O) \, dt + \int Q_{TS} \, dt - \int Q_{SG} \, dt - \int Q_{IG} \, dt; \quad \text{for } Q_{HS} > 0
\]

(4.15)

where the overall heat loss coefficient is estimated as a constant value and is taken outside of the integral of the building heat loss term.

### 4.3.1.1 Continuously Heated Buildings

The model is then simplified for the case of a continuously heated building, where over time it is assumed there will be no thermal storage effects, and the internal temperature is given by the building’s set-point temperature, as follows:

\[
E = \int Q_{HS} \, dt; \quad \text{for } Q_{HS} > 0
\]

\[
= HLC \int (T_{SP} - T_O) \, dt - \int Q_{SG} \, dt - \int Q_{IC} \, dt; \quad \text{for } Q_{HS} > 0
\]

(4.16)

In this model, not all of the heat loss estimated by the indoor-outdoor temperature differential is supplied by the building’s heating system. Over time, some of this heat
loss is supplied by solar and internal heat gains. To account for these effects, the heat loss temperature differential is simply adjusted by the equivalent temperature effect of these gains. These temperatures are calculated by dividing each heat gain inside the temperature differential by the heat loss coefficient, as follows:

\[ E = \int Q_{HS} \, dt = HLC \int \left( T_{SP} - \frac{Q_{SG}}{HLC} - \frac{Q_{IG}}{HLC} - T_o \right) dt; \text{ for } Q_{HS} > 0 \]  \hspace{1cm} (4.17)

\[ T_{SG} = \frac{Q_{SG}}{HLC} \] \hspace{1cm} (4.18)

\[ T_{IG} = \frac{Q_{IG}}{HLC} \] \hspace{1cm} (4.19)

\[ E = \int Q_{HS} \, dt = HLC \int (T_{SP} - T_{SG} - T_{IG} - T_o) \, dt; \text{ for } Q_{HS} > 0 \] \hspace{1cm} (4.20)

where: \( T_{SG} \) is the equivalent temperature effect of solar heat gain (°C) and \( T_{IG} \) is the equivalent temperature effect of internal heat gain (°C).

However, as these equivalent temperatures cannot be accounted for separately in HDD models based on monthly (or longer) fuel consumption data, they are combined together in the HDD method, as follows:

\[ E = \int Q_{HS} \, dt = HLC \int (T_{SP} - T_G - T_o) \, dt; \text{ for } Q_{HS} > 0 \] \hspace{1cm} (4.21)

\[ T_G = T_{SG} + T_{IG} \] \hspace{1cm} (4.22)
where: $T_G$ is the equivalent temperature heat gain due to both solar and internal heat gains ($^\circ$C).

The HDD variable is used to estimate a simplified form of the integral of this temperature differential over a day, using the building’s base temperature parameter, as follows:

$$E_D = \int_D Q_{HS} \, dt = HLC \int_D (T_B - T_O) \, dt; \text{ for } Q_{HS} > 0$$  \hspace{1cm} (4.23)

$$T_B = T_{SP} - T_G$$  \hspace{1cm} (4.24)

$$HDD = \max(0; T_B - \bar{T}_O) \approx \int_D (T_B - T_O) \, dt$$  \hspace{1cm} (4.25)

where: the subscript $D$ is used to denote a day.

In this HDD formula, the base temperature parameter is estimated as a constant value and represents the limiting temperature for the building for which the indoor-outdoor temperature differential is positive and heat is required to maintain the building’s set-point temperature. The main benefit of the base temperature concept is that the energy demand of the building can be estimated using only outdoor temperature data from the nearest weather station.

However, it is seen in Equations 4.22 and 4.24 that a constant base temperature implies that the equivalent temperature effect of solar gains are also constant for each day, even though such gains are a variable component of a building’s daily fuel consumption.
series. This simplifying assumption will also be addressed in the development of an improved HDD estimator.

### 4.3.1.2 Intermittently Heated Buildings

The HDD variable is then redefined for the more common case of an intermittently heated building. In such buildings, the thermal storage component of the original energy demand model in Equation 4.15 will have an effect on the heat consumption, as the internal temperature in the building is allowed to vary across each day. However, instead of addressing this thermal storage component directly, the HDD variable is simply redefined by a revised base temperature, as follows:

\[ T_B = \bar{T}_I - T_G \]  

(4.26)

where it is seen that the building’s set-point temperature, used in the definition of the base temperature in Equation 4.24, is simply replaced by the building’s average internal temperature because it is not continuously heated.

This concept of accounting for the thermal storage effects by defining the base temperature as a function of the average indoor temperature, rather than the set-point temperature, is used in order to prevent the over-estimation of heat consumption using direct modelling methods (see Section 4.2). It is also important to distinguish between the alternative base temperature definitions of Equation 4.24 and Equation 4.26 when interpreting the estimated base temperature using indirect regression modelling methods. Although, in general terms it may be assumed that the building is heated intermittently there may be exceptions such as hospitals, for example.
The adjusted base temperature is based on an idealised indoor temperature profile for a building on an average day during the heating season, as illustrated in Figure 4.4. In this approach, the effect of thermal storage influences the rate of change of the internal temperature of the building as it cools from or is heated to the set-point temperature over the unoccupied period. This in turn influences the average indoor temperature of the building.

In direct HDD modelling methods, the average internal temperature may be calculated based on this indoor temperature profile using thermodynamic formulae that account for the building fabric’s heat loss coefficient and effective thermal capacity, while assuming a constant outdoor temperature for the unoccupied period. In indirect HDD regression modelling methods the average internal temperature may be simply inferred from the estimated base temperature.

It is seen that Equation 4.26 does not take into account the inertial effect of thermal storage. This effect is not apparent in fuel consumption data with large time steps such as months. However, it is apparent in the daily consumption data of intermittently heated buildings, and should therefore be accounted for in the development of an improved HDD estimator.

### 4.3.2 Weather Adjusted HDD

The weather adjusted HDD (HDD\text{WA}) parameter derived below improves upon the HDD method by accounting for important causes of daily variability in heat consumption related to weather conditions. In the derivation below, the HDD\text{WA} variable results from a series of incremental improvements to the HDD variable, in order to address daily variation in:
1. heat gain due to solar radiation;

2. air-infiltration due to wind speed; and

3. building fabric thermal storage due to intermittent heating.

### 4.3.2.1 Solar Radiation

The first revision to the HDD parameter is to account for the effects of variable solar heat gain by extracting the equivalent temperature effect of solar gains from the base temperature in the consumption model of Equation 4.23, as follows:

\[
E = \int Q_{HS} \, dt = HLC \int (T_B - T_{SG} - T_O) \, dt
\]  
(4.27)

\[
T_B = T_{SP} - T_{IG}
\]  
(4.28)

In this energy consumption model, it is seen that the HDD parameter of Equation 4.25 can be adjusted to account for the variable effect of solar gains. This adjustment is estimated using the following parameter:

\[
HDD_{SA} = \max\left(0; T_B - \gamma GR - T_{O,AVG}\right) \approx \int_D (T_B - T_{SG} - T_O) \, dt
\]  
(4.29)

where: \(HDD_{SA}\) is the solar adjusted HDD parameter (°C·day), which accounts for the variable effect of solar heat gain using \(GR\), the daily global radiation value, and the coefficient \(\gamma\) to estimate the required temperature adjustment. Importantly, it has been found that global radiation is not significantly correlated with Irish outdoor
temperatures and as a result multi-collinearity between these variables in the above solar adjusted parameter is not a concern.

4.3.2.2 Wind Speed

The next step is to address the assumption that the effects of air-infiltration can be estimated using a constant air-infiltration coefficient. This modification is developed based on the concept that the overall heat loss coefficient in the building fuel consumption model of Equation 4.3 may be modified to become a heat loss variable that accounts for daily variation in wind speed, as follows:

\[
F = \frac{24}{\eta} (HLV)(HDD_{SA})
\]  

(4.30)

\[
HLV = \left( \sum UA + 0.33N_D V \right)/1000
\]

(4.31)

where: \( HLV \) is the heat loss variable (kW/°C); and \( N_D \) is the mean number of air changes per hour for the building for a given day.

However, the original HDD method is based on the concept that the heat consumption for a building may be estimated as the product of a heat loss coefficient by a variable temperature differential. This in turn allows the HDD variable to be used in linear regression fuel consumption models, where the suitability of the estimated model can be assessed by:

- the standard error of the HDD coefficient, and
• the $R^2$ (coefficient of determination) value of the model.

So that the effect of variable air-infiltration can be accounted for within the HDD$_{WA}$ parameter and in a simple linear model of building fuel consumption, this effect is used to adjust the HDD$_{SA}$ parameter by rearranging the fuel consumption model of Equation 4.30, as follows:

\[
F = \frac{24}{\eta} \frac{(\sum UA + 0.33NDV)}{1000} (HDD_{SA})
\]  
(4.32)

\[
F = \frac{24 \sum UA}{\eta} (HDD_{SA}) + \frac{24 \times 0.33NDV}{\eta} (HDD_{SA})
\]  
(4.33)

\[
F = \frac{24 \sum UA}{\eta} \left( HDD_{SA} + \frac{\frac{24 \times 0.33NDV}{\eta}}{24 \sum UA} (HDD_{SA}) \right)
\]  
(4.34)

\[
F = \frac{24 \sum UA}{\eta} \left( HDD_{SA} + \frac{0.33NDV}{\sum UA} (HDD_{SA}) \right)
\]  
(4.35)

In this fuel consumption model, the HDD$_{SA}$ parameter of Equation 4.29 is adjusted to account for the effects of air-infiltration. This adjustment is estimated using the following parameter:

\[
HDD_{S&WA} = HDD_{SA} + \gamma_2 (\overline{WS}) (HDD_{SA}) \approx HDD_{SA} + \frac{0.33NDV}{\sum UA} (HDD_{SA})
\]  
(4.36)
\[ HDD_{S\&WA} = (1 + \gamma_2 \overline{WS})HDD_{SA} \]  

(4.37)

where: \( HDD_{S\&WA} \) is the solar and wind adjusted HDD parameter (°C·day), which includes the \( HDD_{SA} \) parameter of Equation 4.29 and accounts for the effects of air-infiltration heat loss, using \( \overline{WS} \) the daily average wind speed value with the coefficient \( \gamma_2 \) to estimate the required HDD adjustment.

The complete form of this parameter is given as follows by substituting Equation 4.29 for \( HDD_{SA} \):

\[ HDD_{S\&WA} = (1 + \gamma_2 \overline{WS})\max(0; T_B - \gamma_1 GR - \overline{T}_0) \]  

(4.38)

### 4.3.2.3 Intermittently Heated Buildings

The HDD parameter is revised again for the more common case of an intermittently heated building. For such buildings, the HDD parameter was simply modified by redefining the base temperature of a building to include the average daily internal temperature, rather than the set-point temperature. However, this simplification does not take into account that an intermittently heated building’s daily heat consumption may depend on the climatic conditions of previous days, in the form of building fabric thermal storage effects.

In a building the effect of thermal storage is to absorb or release heat, as a function of the building’s effective thermal capacity and the rate of change of the building fabric temperature, as follows [60]:

87
\[ Q_{TS} = c_{\text{eff}} \frac{dT_{BF}}{dt} \]  \hfill (4.39)

where: \( c_{\text{eff}} \) is the effective heat capacity of the building fabric (kJ/°C), and \( dT_{BF}/dt \) is rate of change of the building fabric temperature.

Based on this model, the average rate of heat transfer to/from the building fabric over the day may be estimated, as follows [55]:

\[ \bar{Q}_{TS} = \frac{c_{\text{eff}}(\Delta T_{BF})}{24(3600)} \]  \hfill (4.40)

where: \( \bar{Q}_{TS} \) is the average rate of heat transfer to/from the building fabric over the day; \( \Delta T_{BF} \) is the change in building fabric temperature over the day (°C/day); and the average rate of change of this temperature, is given by the denominator.

This heat demand may be incorporated into Equation 4.38 as the equivalent temperature effect of thermal storage, as follows:

\[ HDD_{WA,D} = (1 + \gamma_2 W_S) \max(0; T_B - \gamma_1 G_R - T_{O,D} + T_{TS,D}) \]  \hfill (4.41)

\[ T_{TS,D} = \frac{c_{\text{eff}}(\Delta T_{BF,D})}{24(3600)(HLV_D)} \]  \hfill (4.42)

where: \( HDD_{WA} \) is the weather adjusted HDD parameter (°C·day) that denotes the HDD parameter adjusted for the variable effects of solar heat gain, air-infiltration heat loss and thermal storage; \( T_{TS} \) is the equivalent temperature effect of thermal storage (°C); and \( HLV_D \) is the heat loss variable for the day, which is used to convert the average
thermal storage heat demand to an equivalent temperature – as similarly applied in Equation 4.17 using the building’s overall heat loss coefficient.

Although some estimate of the daily change in building fabric temperature is required for Equation 4.42 no data are available. However, it has been proposed that the daily change in building fabric temperature can be estimated by the daily change in average outdoor temperature [61], as illustrated in Figure 4.6.

**Figure 4.6**: Idealised temperature distribution across a building fabric element on consecutive days.

In this figure, an idealised temperature distribution is shown for a building fabric element from the outside environment to the heated space on consecutive days. For both days the building fabric’s temperature increases from the outer to the inner surfaces. It can be seen that the daily change in building fabric temperature can be estimated by the daily change in average outdoor temperature, if it is assumed that the (unavailable) indoor temperature is relatively constant compared to the outdoor temperature.
This concept has been used to estimate a building heating system’s daily fuel consumption due to thermal storage ($F_{TS,D}$) [61], as follows:

$$F_{TS,D} = c_{eff} \Delta T_{BF,D} \quad (4.43)$$

$$\Delta T_{BF,D} \approx \bar{T}_{O,D} - \bar{T}_{O,D-1} \quad (4.44)$$

This approximation of the daily change in building fabric temperature is used in Equation 4.41 to improve the HDD parameter, as follows:

$$HDD_{W,AD} \approx (1 + \gamma_2 WS_D) \max \left(0; T_B - \gamma_1 GR_D - \bar{T}_{O,D} + \alpha_1 \Delta T_{BF} \right) \quad (4.45)$$

$$HDD_{W,AD} \approx (1 + \gamma_2 WS_D) \max \left[0; T_B - \gamma_1 GR_D - \bar{T}_{O,D} + \alpha_1(\bar{T}_{O,D} - \bar{T}_{O,D-1}) \right] \quad (4.46)$$

$$HDD_{W,AD} \approx (1 + \gamma_2 WS_D) \max \left[0; T_B - \gamma_1 GR_D - [(1 - \alpha_1)\bar{T}_{O,D} + \alpha_1\bar{T}_{O,D-1}] \right] \quad (4.47)$$

where, if the building is continuously heated the thermal storage ($\alpha_1$) parameter will reduce to zero, and the revised HDD parameter reduces to the form given by Equation 4.38.

The outdoor temperature terms in Equation 4.47 introduce the concept of building thermal memory in that the temperature of a building can be influenced by the outdoor temperature for the previous day. For example, if yesterday’s outdoor temperature ($\bar{T}_{O,D-1}$ = 5°C) is colder than today’s ($\bar{T}_{O,D}$=10°C), the thermal memory term ($(1-\alpha_1)\bar{T}_{O,D}+\alpha_1\bar{T}_{O,D-1}$) accounts for greater heat consumption than would have been the case if no thermal memory term was included (and vice-versa), as the resulting temperature is colder than
today’s outdoor temperature \((0.5\bar{T}_{O,D}+0.5\bar{T}_{O,D-1})=7.5^\circ C\), assuming \(a_1=0.5\). However, the temperature of a building can be influenced by outdoor temperatures over previous days, not just one day as indicated in Equation 4.47. For example, in building cooling literature, it has been found that the temperature distribution in a 300mm deep concrete slab depends on daily average outdoor temperatures up to one month previously [62].

In Equation 4.44 and Figure 4.6, it was assumed that the indoor temperature is relatively constant and the daily change in building fabric temperature may be estimated by the daily change in outdoor temperature. However, based on building cooling literature, it is the temperature at the outer region of the building fabric that is mostly influenced by daily changes in outdoor temperatures, see Figure 4.7. The temperature at the mid- and core region of the building fabric are influenced by longer term outdoor temperature fluctuations. The extent of this thermal memory relates to the depth and in turn the thermal capacity of the building fabric.

![Figure 4.7: Temperature variation across a building fabric element.](image)

HDD estimation should therefore be extended to account for the effect of additional days, as follows:
\[ HDD_{WA} = (1 + \gamma_2 W_S) \max \left( 0; T_B - \gamma_1 GR_D - \bar{T}_{0,D} + \alpha_1 \Delta T_{BF,D} \right) \]  
(4.48)

\[ \Delta T_{BF,D} \approx \bar{T}_{0,D} - T_{EFF \cdot O,D-1} \]  
(4.49)

\[ HDD_{WA} = (1 + \gamma_2 W_S) \max \left[ 0; T_B - \gamma_1 GR_D - \bar{T}_{0,D} + \alpha_1 (\bar{T}_{0,D} - T_{EFF \cdot O,D-1}) \right] \]  
(4.50)

\[ HDD_{WA} = (1 + \gamma_2 W_S) \max \left[ 0; T_B - \gamma_1 GR_D - [(1 - \alpha_1)\bar{T}_{0,D} + \alpha_1 T_{EFF \cdot O,D-1}] \right] \]  
(4.51)

where: \( T_{EFF \cdot O,D-1} \) is the effective outdoor temperature for previous days (°C) and is initially defined as follows:

\[ T_{EFF \cdot O,D-1} \equiv \varphi_1 \bar{T}_{0,D-1} + \varphi_2 \bar{T}_{0,D-2} + \varphi_3 \bar{T}_{0,D-3} + \cdots \; \text{for} \; \varphi_1 > \varphi_2 > \varphi_3 > \cdots, \text{and } \sum_{i=1}^{n} \varphi_i = 1 \]  
(4.52)

using decreasing \( \varphi \) factors of outdoor temperature to reflect their decreasing influence on the current day’s fabric temperature.

However, this method introduces an unknown number of \( \varphi \) factors required to estimate the effective outdoor temperature in Equation 4.52. In order to limit the number of parameters required to estimate the HDD_{WA} variable, the effective outdoor temperature is based on that used in the UK’s CWV (see Equation 3.17), as follows:

\[ T_{EFF \cdot O,D} = (1 - \alpha_1)\bar{T}_{0,D} + \alpha_1 T_{EFF \cdot O,D-1}; 0 \leq \alpha_1 \leq 1 \]  
(4.53)

and as a result the HDD_{WA} variable in Equation 4.51 is now given by:
\[
HDD_{WA,D} = (1 + γ_2 \bar{WS}_D) \max(0; T_B - γ_1 GR_D - T_{EFF,O,D})
\]  
(4.54)

By expanding Equation 4.53, over several days it is seen that by using \( \alpha_1 \) (thermal storage/memory parameter) that the redefined prevailing temperature is equivalent to the initial function of Equation 4.52:

\[
T_{EFF,O,D} = (1 - \alpha_1) \bar{T}_{O,D} + \alpha_1(1 - \alpha_1) \bar{T}_{O,D-1} + \alpha_1^2(1 - \alpha_1) \bar{T}_{O,D-2} + \cdots; \text{ for } 0 \leq \alpha_1 \leq 1
\]  
(4.55)

where the weights applied result in successive terms decreasing approximately exponentially to a limit of zero over time, for example:

\[
T_{EFF,O,D}(\alpha_1 = 0.7) = 0.3 \bar{T}_{O,D} + 0.21 \bar{T}_{O,D-1} + \cdots + 0.0353 \bar{T}_{O,D-6} + \cdots
\]  
(4.56)

This HDD\(_{WA}\) variable may be used to model the daily gas consumption of individual buildings. It accounts for the additional effects of solar radiation, wind speed and building thermal memory and is an improvement on the HDD variable which only accounts for base and outdoor temperatures. However, it is not applied in this research as does not account for further effects on daily NDM market or SME gas consumption, such as seasonal consumption behaviour in the case of the market or variable heating schedules for each day of week in the case of SMEs. Although, it is the basis upon which the adjusted HDDs referred to previously are derived in the following sections.
4.3.3 Climate Adjusted Network Degree Day

A climate adjusted network degree day (NDD$_{CA}$) is an adapted HDD parameter that can be used to model network (or NDM market) gas consumption. It is used in this regard in Chapter 6. It is based on two adjustments to the HDD$_{WA}$ parameter in Equation 4.54 to account for additional effects on network gas consumption.

The first of these adjustments is to address the effect of base temperature variation across the domestic and SME building population in the market. In the HDD and the HDD$_{WA}$ variables, the base temperature parameter serves two functions: it indicates the temperature above which heating is not used; and it is used to define a temperature transformation that allows the heat consumption response of a building to be estimated linearly. However, the base temperature appropriate to each building will vary across buildings in the NDM market, due to individual consumer behaviour and building thermal efficiency. In Figure 4.8, this effect is illustrated for three domestic consumers from the smart metered sample using the HDD transformation in Equation 4.10.
Figure 4.8: Relationship between weekday gas consumption and daily average outdoor temperature for three sample domestic gas consumers and the NDM gas market for the gas year: October 2009 - September 2010.
In Figure 4.8, the cumulative effect of individual variations in base temperatures and consumption responses at a network level is illustrated using weekday NDM market consumption data. It is seen that there is a curve in the gas consumption response to mild temperatures when varying numbers of consumers switch on their heating system. This non-linear response is modelled using both an upper $T_{B,\text{upr}}$ and lower $T_{B,\text{lwr}}$ base temperature parameters in the following three-case transformation model [63]:

$$C_{\text{NDM,WD}} = b_0 + \begin{cases} 0; & \text{if } T_{0,D} > T_{B,\text{upr}} \\ b_1 \left( \frac{(T_{0,D} - T_{B,\text{upr}})^2}{2(T_{B,\text{upr}} - T_{B,\text{lwr}})} \right); & \text{if } T_{B,\text{lwr}} < T_{0,D} \leq T_{B,\text{upr}} \\ \varepsilon_D \left( \frac{T_{B,\text{upr}} + T_{B,\text{lwr}}}{2} - T_{0,D} \right); & \text{if } T_{0,D} \leq T_{B,\text{lwr}} \end{cases}$$

(4.57)

In Equation 4.57, the transformation function reduces to the HDD transformation function in Equation 4.10, if the estimated $T_{B,\text{upr}}$ and $T_{B,\text{lwr}}$ base temperature parameters are equal, and is also seen in Figure 4.8 to model a curve in the gas consumption response between these limits if required. This transformation function is linear for temperatures below $T_{B,\text{lwr}}$, and is selected in preference to the UK-CWV’s transformation function in Equation 3.20, which also accounts for an observed non-linear increase in the rate of gas consumption during very cold weather. It can be seen in Figure 4.8 that such an increase in gas consumption rates is not observed for the Irish NDM market.

The second adjustment to the HDD$\text{WA}$ variable made by the NDD$\text{CA}$ parameter is to address the effect of seasonal consumption behaviour, or the response of consumers’ to unseasonable weather conditions whereby seasonal or normal gas consumption levels are maintained for that time of year irrespective of prevailing weather conditions. This
is due to the effect of inefficient consumers who do not reduce their heating needs during mild temperature periods in the wintertime. Such effects are accounted for by the NDD_{CA} parameter using seasonal degree-day values.

The ‘climate-adjusted NDD (NDD_{CA})’ parameter is so called because it accounts for the additional network level effect of base temperature variation and seasonal consumer behaviour using long-term climate or seasonal degree-day values, and is given by:

\[
NDD_{CA,D} = (1 - \omega_1)NDD_{WA,D} + \omega_1 SS-NDD_{WA,D}
\]  

(4.58)

where: \(NDD_{WA,D}\) is given by Equations 4.59-4.62, and is the weather adjusted network degree day for the given day (D) that is based on the HDD_{WA} in Equation 4.54 and the alternative transformation function in Equation 4.57; and \(SS-NDD_{WA}\) is given by Equations 4.63 and 4.64, and is the corresponding smoothed seasonal value of the \(NDD_{WA}\) parameter, which is used with the coefficient \(\omega_1\) to account for the effect of seasonal consumption.

The NDD_{WA} required as part of the NDD_{CA} parameter is given by:

\[
NDD_{WA,D} = (1 + \gamma_2 \overline{WS}_D) \times \begin{cases} 
0; & \text{if } T_{S&TM,D} > T_{B,upr} \\
(T_{S&TM,D} - T_{B,upr})^2 / 2(T_{B,upr} - T_{B,lwr}); & \text{if } T_{B,lwr} < T_{S&TM,D} \leq T_{B,upr} \\
(T_{B,upr} + T_{B,lwr})/2 - T_{S&TM,D}; & \text{if } T_{S&TM,D} \leq T_{B,lwr} 
\end{cases}
\]  

(4.59)

where: the alternative transformation function in Equation 4.57 is applied; and \(T_{S&TM,D}\) (°C) is a temperature used to account for the combined effect of solar gain and thermal memory that is given by:
\[ T_{S\&TM,D} = T_{SG,D} + T_{EFF-O,D} \]  

(4.60)

where: \( T_{SG,D} \) (°C) is given by Equation 4.61, and is the equivalent temperature effect of solar gains; and \( T_{EFF-O,D} \) (°C) is given by Equation 4.62, and is the effective outdoor temperature.

\[ T_{SG,D} = \gamma_1 G_R D \]  

(4.61)

\[ T_{EFF-O,D} = (1 - \alpha_1) \bar{T}_{O,D} + \alpha_1 T_{EFF-O,D-1}; 0 \leq \alpha_1 \leq 1 \]  

(4.62)

The \( SS-ND\_WA \) required as part of the \( NDD\_CA \) parameter is given by:

\[
SS-ND\_WA,D = \begin{cases} 
\frac{1}{m} \sum_{i=-j}^{j} S-ND\_WA_{d+i} & \text{for } D \neq \text{Feb’29th} \\
0.5( SS-ND\_WA_{D-1} + SS-ND\_WA_{D+1} ) & \text{for } D = \text{Feb’29th}
\end{cases}
\]

(4.63)

\[
S-ND\_WA_D = \frac{1}{n} \sum_{i=0}^{n-1} NDD\_WA_{d+365i}; \text{Feb’29th } \notin \text{ climate data}
\]

(4.64)

where: \( d \) is the corresponding day of year number for a non-leap year; \( m \) is the order of the moving average filter, \( j = (m - 1) / 2 \); \( S-ND\_WA,D \) (°C·day) is the seasonal value of the \( NDD\_WA \) variable; and \( n \) is the number of years of climate data used to calculate each seasonal value; and where \( m \) and \( n \) are manually specified parameters.
4.3.4 Weather and Day-Type Adjusted HDD

A weather and day-type adjusted HDD (HDD\textsubscript{WDA}) is a parameter that can be used to model the daily gas consumption of smart-metered domestic and SME buildings. It is used in this regard to model the daily gas consumption of SMEs in Chapter 7. It is based on a single adjustment to the HDD\textsubscript{WA} parameter in Equation 4.54 to account for the effect of variable heating schedules in a building for each day of the week. For example, an office building may only be occupied on weekdays and may employ reduced heating system temperature set-points and operating hours at the weekend. The HDD\textsubscript{WDA} parameter accounts for such effects by estimating different base temperatures for each day of week, as follows:

\[
HDD_{WDA} = (1 + \gamma_{2} \bar{WSD}) \max(0; T_{B,DoW} - \gamma_{3} GR_{D} - T_{EFF-O,D})
\]  \hspace{1cm} (4.65)

where: \(T_{B,DoW}\) is the base temperature for a given day of the week (\(^\circ\)C).

4.4 Conclusion

This chapter began with a description of HDD theory and how the parameters of simple HDD regression models can be interpreted. Such models and these interpretations are used to develop multinomial logistic regression (MLR) models for the domestic gas end-use efficiency benchmarking tool in Chapter 5.

The derivation of the HDD parameter was then described and opportunities to develop upon this were identified. This resulted in the development of HDD\textsubscript{WA} parameter that may be used to address the additional effects of solar radiation, wind-speed and thermal...
memory in models of daily gas consumption for individual buildings. This parameter was used to develop the alternative NDD_{CA} and HDD_{WDA} parameters.

The NDD_{CA} parameter is used to model daily gas consumption for the NDM market in Chapter 6. Extreme values of the parameter are also estimated in Chapter 6, and these are used to forecast year-ahead peak-day gas consumption for the market to alternative supply standards. The HDD_{WDA} parameter is used in the development of individual consumer models in Chapter 7. It is used to assess the practicality of individualised state-of-the-art gas consumption estimators for each consumer in the NDM market, as part of the portfolio forecasting process.
CHAPTER 5

DOMESTIC ENERGY EFFICIENCY BENCHMARKING TOOL
5 DOMESTIC ENERGY EFFICIENCY BENCHMARKING TOOL

This chapter develops a statistical benchmarking tool that can be used by energy suppliers to infer the gas end-use efficiency of buildings in their domestic portfolio, in order to deliver improved energy management services to consumers and to fulfil commitments made under the EU’s ‘Energy Efficiency Obligations Scheme’. This benchmarking tool is based on the simple HDD linear regression model in Equation 4.10, and a multinomial logistic regression (MLR) modelling method that is described later in this chapter. The main steps in the method are summarised in Figure 5.1.

Figure 5.1: Summary of the domestic energy efficiency benchmarking method.

The method begins by using NLS to estimate a HDD regression model, including intercept \(b_0\), slope \(b_1\) and base temperature \(T_B\) parameters for each consumer in the
sample. The resulting intercept, slope and base temperature parameter distributions are then presented. The relationship between these parameters and the thermal energy performance of a building were previously described in Chapter 4. It was stated that: 1) the intercept parameter relates to gas consumption for cooking and hot water purposes; 2) the slope parameter relates to the overall heat loss coefficient of a building and the efficiency of its heating system; and 3) the base temperature parameter relates to the average temperature inside a building and the equivalent temperature effect of useful heat gain. However, as these parameters are dependent on factors such as dwelling size and occupancy, it is difficult for suppliers to identify appropriate energy saving measures for individual consumers based on their HDD model parameter estimates alone.

This issue is addressed in the benchmarking method using multinomial logistic regression models. These models are used to characterise the intercept, slope and base temperature parameter distributions resulting from the HDD models. The MLR models are estimated using survey data such as the number of occupants and bedrooms for each dwelling in the sample. Such household characteristics were generally known by the consumer sample when surveyed and this data can be easily gathered by energy suppliers when applying the benchmarking tool. The resulting MLR models can be used to estimate the probability that an individual consumer’s HDD model parameter estimates are higher or lower than expected when compared to similar households. This allows energy efficiency measures which are likely to be appropriate for the consumer to be identified.

The benefit of this benchmarking approach is that it estimates the relative end-use gas consumption for each customer compared to other similar households. The methods
previously described in Section 3.3 compare buildings based on floor area, even though many householders are unable to provide this building measurement when surveyed. In addition to the benefits to suppliers in identifying poorly-performing customers for possible demand side management measures, the assessment method can also be used to provide consumers with benchmarks for their own energy management needs. These benefits and examples of this method are demonstrated later in this chapter.

5.1 Methodology

This section begins with a summary of the data used to develop the domestic energy efficiency benchmarking tool. Next, the NLS method used to estimate HDD regression models for the domestic consumer sample is described. The section is completed by a description of the multinomial regression method used to characterise the intercept, slope and base temperature parameter distributions resulting from the HDD modelling process.

5.1.1 Data

The development of the benchmarking tool is based on domestic smart-metered gas consumption and household survey data for a sample of over 500 dwellings which formed the control group in the Irish smart meter trials, as described in Section 2.1.1. The HDD models are estimated using temperature data from Dublin Airport, as described in Section 2.3. The household variables used are shown in Table 5-1, where their relationships to the intercept, slope and base temperature parameters of the HDD regression model in Equation 4.10 are also described.
The survey also collected data on building floor area, wall insulation and building occupancy. However, it was found that a significant proportion of consumers did not provide information in this regard. For example, 59% did not know their building’s floor area, 27% did not know whether or not wall insulation was present in their building, and 26% did not state whether or not their building was occupied by adults during the day. Therefore these explanatory factors were not used in the development of the MLR models, as their inclusion would severely limit the usable sample size.

Similarly, data relating to the presence of attic insulation have not been used in the MLR models, as this explanatory factor does not apply to all dwelling types (for example, mid-level apartments), and a question in this regard was not included in the survey.

**Table 5-1**: Applied survey data and their relationship to the HDD regression model.

<table>
<thead>
<tr>
<th>Survey Data Collected</th>
<th>Relationship to the HDD Regression Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
<td><strong>Categories</strong></td>
</tr>
<tr>
<td>No. of adults</td>
<td>1, 2, 3, 4, 5 or ≥ 6.</td>
</tr>
<tr>
<td>No. of children</td>
<td>0, 1, 2, 3, 4, 5 or ≥ 6.</td>
</tr>
<tr>
<td>Hot water system</td>
<td>Timed gas fuelled, untimed gas fuelled or alternatively fuelled system. (^{(a)})</td>
</tr>
<tr>
<td>Cooking system</td>
<td>Gas fuelled or alternatively fuelled system. (^{(a)})</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>1, 2, 3, 4 or ≥ 5.</td>
</tr>
<tr>
<td>Dwelling type</td>
<td>Apartment, terrace, semi-detached, detached or bungalow.</td>
</tr>
</tbody>
</table>
### Table 5-1: continued

<table>
<thead>
<tr>
<th>Survey Data Collected</th>
<th>Relationship to the HDD Regression Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construction year</strong></td>
<td><strong>Slope (b₁)</strong> These construction years generally relate to increasing levels of insulation as required by Irish building standards. And this relates to the required fabric U-values used to determine a building’s heat loss coefficient.</td>
</tr>
<tr>
<td>Pre-1935, 1935-1979, 1980-1999, 2000-2004 or 2005-2010.(^{b)})</td>
<td>This relates to heating system efficiency, which in turn is used to determine a building’s heat loss coefficient.</td>
</tr>
<tr>
<td><strong>Boiler service frequency</strong></td>
<td><strong>Base temperature (T₀)</strong> This set-point together with the heating system operating hours is related to a building’s average indoor temperature, and this in turn is related to the building’s base temperature.</td>
</tr>
<tr>
<td>Annually, every 2-3 years or never.</td>
<td>This relates to heating system operating hours and whether or not a consumer can control the set-point temperature in different zones of their building in order to facilitate decreased average indoor temperatures. All of which is related to the building’s base temperature.</td>
</tr>
<tr>
<td><strong>Temperature set-point</strong></td>
<td><strong>Timer control</strong> Separate zones, single zone, not known by the consumer, or either the timer system is not used or no time control system is present.(^{a)})</td>
</tr>
<tr>
<td>&lt; 18°C, 18-20°C, 21°C, 22-24°C, &gt;24°C, not known by the consumer, or no thermostat control system.</td>
<td>This relates to heating system operating hours and whether or not a consumer can control the set-point temperature in different zones of their building in order to facilitate decreased average indoor temperatures. All of which is related to the building’s base temperature.</td>
</tr>
<tr>
<td><strong>Timer control</strong></td>
<td><strong>Operating hours</strong> See temperature set-point description above.</td>
</tr>
<tr>
<td>Separate zones, single zone, not known by the consumer, or either the timer system is not used or no time control system is present.(^{a)})</td>
<td>0 &lt; h/day ≤ 8, 8 &lt; h/day ≤ 10, 10 &lt; h/day ≤ 12 or 12 &lt; h/day ≤ 24.(^{c)})</td>
</tr>
<tr>
<td><strong>Operating hours</strong></td>
<td></td>
</tr>
<tr>
<td>0 &lt; h/day ≤ 8, 8 &lt; h/day ≤ 10, 10 &lt; h/day ≤ 12 or 12 &lt; h/day ≤ 24.(^{c)})</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

(a) The levels (or categories) of this explanatory factor incorporate alternative categories or answers allowed in the survey questionnaire. For example, there were three alternative answers in the survey which described a gas fuelled hot water system: 1) central heating system, 2) combination boiler (no hot water cylinder) or 3) gas fuelled system.

(b) Construction year is reported in the survey either by the actual construction year or by the categories given in the table, thus any actual construction years reported in the survey have been also been categorised.

(c) Heating system operating hours have been determined using each consumer’s hourly resolution smart-metered gas consumption data. For simplicity, this metric has been evaluated for each consumer by the average daily number of gas consumption hours during the second week of January. During this week, it is assumed that buildings are likely to be occupied and heating systems are likely to consume gas during each timed operating hour. Any suspected pilot light consumption in the sample has been accounted for by applying a nominal 0.5 kWh gas consumption threshold to the hourly gas consumption data.
5.1.2 HDD Model Estimation

The HDD model in Equation 4.10 is estimated for each consumer in the sample using the Levenberg-Marquardt non-linear least-squares (nlsLM) algorithm, available in the statistical computing software, R [64]. This local NLS method was used in preference to a global NLS (nls) algorithm that is also available in R, as it is more robust to stochastic changes in the modelled series – such as significant increases or decreases in daily gas consumption values for no apparent reason. Another reason for this method over the global NLS algorithm was that it allows logical limits to be specified for each parameter in the model.

Each HDD model is estimated using daily gas consumption values for the final year in the smart-meter trial, as a single heating season is required to estimate the most recent base temperature parameter for each consumer. To help convergence to a local NLS solution, starting values and limits have been stipulated for each parameter as shown in Table 5-2. Alternative starting values were trialled to assess the sensitivity of the HDD models, but this resulted in a slight decrease in the number of successfully converged models and no observable change to the resulting intercept, slope and base temperature parameter distributions.

Table 5-2: Parameter starting values and limits

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Starting Value</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($b_0$)</td>
<td>0</td>
<td>0</td>
<td>None</td>
</tr>
<tr>
<td>Slope ($b_1$)</td>
<td>0</td>
<td>0</td>
<td>None</td>
</tr>
<tr>
<td>Base temperature ($T_B$)</td>
<td>15.5</td>
<td>5</td>
<td>25</td>
</tr>
</tbody>
</table>
Beginning with these starting values this NLS regression method minimises the sum of squared-residuals in Equation 4.10 by solving the optimum intercept ($b_0$), slope ($b_1$) and base temperature ($T_B$) estimates within the lower and upper limits of this three (parameter) dimensional space. This NLS regression method also reports standard errors for each parameter estimate, including the non-linear base temperature parameter, which can be used to assess parameter significance when analysing the estimated HDD model.

5.1.3 MLR Modelling

The MLR modelling method used to characterise each of the HDD model parameter distributions is described in its general form for the categorical dependent variable $Y$, as follows [65]:

\[
\log \left( \frac{P(Y = 1)}{P(Y = K)} \right) = \beta_{0,1} + \beta_{1,1}X_1 + \cdots + \beta_{n,1}X_n
\]

(5.1)

\[
\log \left( \frac{P(Y = 2)}{P(Y = K)} \right) = \beta_{0,2} + \beta_{1,2}X_1 + \cdots + \beta_{n,2}X_n
\]

(5.2)

\[
\log \left( \frac{P(Y = K - 1)}{P(Y = K)} \right) = \beta_{0,K-1} + \beta_{1,K-1}X_1 + \cdots + \beta_{n,K-1}X_n
\]

(5.3)

where: $Y$ includes the categories 1, 2, ..., $K$; $K$ is the specified reference category; and the sub-models describe the log-odds of the other $K$-1 categories compared to the
reference category using separate $\beta_0$ constants and $\beta_1, \beta_2, \ldots, \beta_n$ coefficients for each $X_1, X_2, \ldots, X_n$ explanatory variable.

The exponential of each $\beta$ coefficient are known as odds-ratios and these describe the change in odds for one-unit change in the predictor [66]. These values are used to interpret the effect of each explanatory factor in the MLR model. The coefficients of this model can be used to estimate the probability of each $Y$ variable category, as follows [65]:

$$P(Y=1) = \frac{\exp(\beta_{0,1} + \beta_{1,1}X_1 + \ldots + \beta_{n,1}X_n)}{1 + \exp(\beta_{0,1} + \beta_{1,1}X_1 + \ldots + \beta_{n,1}X_n) + \ldots + \exp(\beta_{0,K-1} + \beta_{1,K-1}X_1 + \ldots + \beta_{n,K-1}X_n)}$$

(5.4)

$$P(Y=2) = \frac{\exp(\beta_{0,2} + \beta_{1,2}X_1 + \ldots + \beta_{n,2}X_n)}{1 + \exp(\beta_{0,1} + \beta_{1,1}X_1 + \ldots + \beta_{n,1}X_n) + \ldots + \exp(\beta_{0,K-1} + \beta_{1,K-1}X_1 + \ldots + \beta_{n,K-1}X_n)}$$

(5.5)

$$\ldots$$

$$P(Y=K-1) = \frac{\exp(\beta_{0,K-1} + \beta_{1,K-1}X_1 + \ldots + \beta_{n,K-1}X_n)}{1 + \exp(\beta_{0,1} + \beta_{1,1}X_1 + \ldots + \beta_{n,1}X_n) + \ldots + \exp(\beta_{0,K-1} + \beta_{1,K-1}X_1 + \ldots + \beta_{n,K-1}X_n)}$$

(5.6)

$$P(Y=K) = \frac{1}{1 + \exp(\beta_{0,1} + \beta_{1,1}X_1 + \ldots + \beta_{n,1}X_n) + \ldots + \exp(\beta_{0,K-1} + \beta_{1,K-1}X_1 + \ldots + \beta_{n,K-1}X_n)}$$

(5.7)

This MLR modelling method is used in the benchmarking tool to characterise low, medium and high categories of each of the resulting intercept, slope and base temperature parameter distributions. Each of these MLR models comprises low and high sub-models based on a medium reference category. The most frequently occurring categorical explanatory variable (see categories in Table 5-1) has been specified as a
reference category. Small sample categories of some explanatory variables have been
merged into alternative categories or removed from the MLR models were appropriate.
Each MLR model is fitted using the MLR (multinom) algorithm in R [64].

5.2 Results and Discussion

The results of this chapter begin with a presentation of the $R^2$ distribution resulting from
the individual HDD models for the consumer sample. Models which poorly fit the data
are removed. A $R^2$ value threshold of 0.6 was chosen to ensure at least a moderately-
strong relationship between the models and the data, and this resulted in the removal of
66 or 13% of the dwellings. The resulting intercept, slope and base temperature
parameter distributions were then characterised using the MLR models; these were then
used to infer the relative energy efficiencies of buildings based on their intercept, slope
and base temperature parameter estimates.

5.2.1 HDD Models

The distribution of $R^2$ values resulting from the HDD models is shown in Figure 5.2.
From this distribution it has been found that 15% and 72% of the HDD models have
strong and moderately-strong $R^2$ values above 0.8 and between 0.6 and 0.8 respectively.
However, 13% of the $R^2$ values are weak to moderate between 0 and 0.6, and as result
these models or consumers have been eliminated from the subsequent HDD model
analysis. These consumers gas consumption was frequently zero during the heating
season, indicating they were either unoccupied, or intermittently occupied. Consequently, they would not represent a good opportunity for energy savings. In
addition, two consumers from the total sample (524) are not included in the $R^2$
distribution in Figure 5.2 or in the subsequent HDD model analysis, as the NLS
algorithm failed to converge using these consumers’ gas consumption series. Again, both of these consumers had numerous zero consumption days during wintertime.

**Figure 5.2:** Distribution of $R^2$ values for the HDD models (522 sample size).

**Figure 5.3:** Boxplot, categories and distribution of the intercept ($b_0$) parameter for HDD models with an $R^2 \geq 0.6$ (456 sample size).
Figure 5.4: Boxplot, categories and distribution of the slope ($b_1$) parameter for HDD models with an $R^2 \geq 0.6$ (456 sample size).

Figure 5.5: Boxplot, categories and distribution of the base temperature ($T_B$) parameter for HDD models with an $R^2 \geq 0.6$ (456 sample size).
The distribution of the intercept, slope and base temperature parameters for the HDD models for the retained consumer sample are shown above in Figures 5.3, 5.4 and 5.5. Each of these parameter distributions have been categorised by low and high quartiles and a medium interquartile range. These categories are shown using boxplots in the figures and are used as a basis in which to develop the following MLR models. This limitation to quartiles allows simple classifications of each distribution and reduces the size of the resulting MLR models in Tables 5.3, 5.4 and 5.5.

The mean value of this base temperature parameter distribution was 14.23°C. This is over a degree lower than the 15.5°C traditionally assumed for HDD modelling in the UK and Ireland. This is unsurprising as this 15.5°C value was recommended in 1934 [55], since then improvements have been made to heating control systems and building insulation standards.

### 5.2.2 Multinomial Logistic Regression Models

The MLR models for the intercept \((b_0)\), slope \((b_1)\) and base temperature \((T_B)\) parameter distributions are shown in Tables 5.3, 5.4 and 5.5. Likelihood ratio (or \(\chi^2\)) tests for these models show that each model rejects the test’s null hypothesis (see Note (a) in Table 5.3), and that most explanatory factors are significant in this regard. However, some explanatory factors did not significantly contribute to their respective MLR models, including: the number of children, boiler service frequency and temperature set-point. By comparing the pseudo-\(R^2\) value (see Note (d) in Table 5.3) for each model, it is seen that the slope and base temperature MLR models have the best and weakest overall fits, respectively.
Table 5.3: Multinomial logistic regression model for the intercept parameter ($b_0$)

<table>
<thead>
<tr>
<th>Intercept Model</th>
<th>$\chi^2$ test of -2LL (df)</th>
<th>Pseudo-$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Model</td>
<td>74.5 (16) ***</td>
<td>0.22</td>
</tr>
<tr>
<td>Explanatory Factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of Adults</td>
<td>17.34 (4) **</td>
<td></td>
</tr>
<tr>
<td>No. of Children</td>
<td>2.73 (6)</td>
<td></td>
</tr>
<tr>
<td>Hot Water</td>
<td>27.95 (4) ***</td>
<td></td>
</tr>
<tr>
<td>Cooking</td>
<td>30.9 (2) ***</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sub-models (c)</th>
<th>Low</th>
<th>Med.(f)</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>$\beta$</td>
<td>SE</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.72</td>
<td>0.37</td>
<td>***</td>
</tr>
<tr>
<td>No. of Adults:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (j)</td>
<td>61</td>
<td>0.22</td>
<td>0.56</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>-0.94</td>
<td>0.37</td>
</tr>
<tr>
<td>≥ 3</td>
<td>14</td>
<td>0.22</td>
<td>0.40</td>
</tr>
<tr>
<td>No. of Children:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 (j)</td>
<td>43</td>
<td>0.52</td>
<td>0.40</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>-0.16</td>
<td>0.58</td>
</tr>
<tr>
<td>≥ 3</td>
<td>18</td>
<td>0.21</td>
<td>0.35</td>
</tr>
<tr>
<td>Hot Water:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Untimed gas system (i)</td>
<td>26</td>
<td>1.13</td>
<td>0.38</td>
</tr>
<tr>
<td>Timed gas system</td>
<td>24</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>Alternative system</td>
<td>15</td>
<td>1.43</td>
<td>0.31</td>
</tr>
<tr>
<td>Cooking:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gas cooker (j)</td>
<td>25</td>
<td>1.43</td>
<td>0.31</td>
</tr>
<tr>
<td>Alternative system</td>
<td>55</td>
<td>1.43</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Notes:

(a) Chi-squared ($\chi^2$) test to ascertain the significance of the decrease in unexplained variance from an intercept only model to the overall model [66], based on the null hypothesis that each regression coefficient in the model is zero [67]. The -2 log likelihood (-2LL) statistic used in this test is given by -2(LL(intercept model) - LL(overall model)) [66, 67]. This $\chi^2$ test is based on model’s corresponding degrees of freedom (df) [67].

(b) $\chi^2$ test to ascertain the significance of explanatory factors to the overall model [66]. This -2LL statistic is given by -2(LL(overall model) - LL(overall model without the factor under test) [66]. This $\chi^2$ test is based on explanatory factor’s corresponding degrees of freedom (df).

(c) See notes (a) and (b).

(d) Nagelkerke’s pseudo-$R^2$ statistic is a measure of the improvement in fit of the overall model compared to a model with no independent variables. This statistic has a range of 0 to 1 and is analogous to the coefficient of determination ($R^2$) statistic used in ordinary least squares regression [67].

(e) Sub-model categories: 0 ≤ Low < Q1 and Q3 < High ≤ Max; where Q1, Q3 and Max are the first and third quartiles and the maximum value of the modelled distribution.

(f) Reference category level. (g) Sample size (n). (h) Coefficient ($\beta$). (i) Standard Error (SE). (j) Reference factor level.

*, ** and *** significance at 0.1, 0.05, 0.01 and 0.001 levels respectively.
Table 5.4: Multinomial logistic regression model for the slope parameter ($b_1$)

<table>
<thead>
<tr>
<th>Slope Model</th>
<th>$\chi^2$ test of -2LL (df) (c)</th>
<th>pseudo-$R^2$ (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Model (a)</td>
<td>157.4 (24) ***</td>
<td>0.37</td>
</tr>
<tr>
<td>Explanatory Factors (b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedrooms</td>
<td>62.57 (6) ***</td>
<td></td>
</tr>
<tr>
<td>Dwelling Type</td>
<td>32 (6) ***</td>
<td></td>
</tr>
<tr>
<td>Construction Year</td>
<td>46.34 (8) ***</td>
<td></td>
</tr>
<tr>
<td>Boiler Service Freq.</td>
<td>5.2 (4)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sub-models (e)</th>
<th>Low</th>
<th>Med.(f)</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (g)</td>
<td>$\beta$ (h)</td>
<td>SE (i)</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>-0.82</td>
<td>0.28</td>
</tr>
<tr>
<td>Bedrooms:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤2</td>
<td>67</td>
<td>1.52</td>
<td>0.47</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>-0.91</td>
<td>0.35</td>
</tr>
<tr>
<td>≥5</td>
<td>1</td>
<td>-1.12</td>
<td>1.13</td>
</tr>
<tr>
<td>Dwelling Type:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-detached</td>
<td>54</td>
<td>0.06</td>
<td>0.30</td>
</tr>
<tr>
<td>Apartment/Terrace</td>
<td>37</td>
<td>0.06</td>
<td>0.30</td>
</tr>
<tr>
<td>Detached</td>
<td>8</td>
<td>-0.18</td>
<td>0.49</td>
</tr>
<tr>
<td>Bungalow</td>
<td>3</td>
<td>-0.45</td>
<td>0.79</td>
</tr>
<tr>
<td>Construction Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1935-1979 (j)</td>
<td>31</td>
<td>0.32</td>
<td>0.51</td>
</tr>
<tr>
<td>&lt;1935</td>
<td>11</td>
<td>0.32</td>
<td>0.51</td>
</tr>
<tr>
<td>1980-1999</td>
<td>34</td>
<td>0.56</td>
<td>0.32</td>
</tr>
<tr>
<td>2000-2004</td>
<td>18</td>
<td>0.79</td>
<td>0.39</td>
</tr>
<tr>
<td>2005-2010</td>
<td>8</td>
<td>0.92</td>
<td>0.56</td>
</tr>
<tr>
<td>Boiler Service Freq.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annually (j)</td>
<td>63</td>
<td>-0.40</td>
<td>0.28</td>
</tr>
<tr>
<td>2-3 years</td>
<td>32</td>
<td>-0.40</td>
<td>0.28</td>
</tr>
<tr>
<td>Never</td>
<td>7</td>
<td>-0.53</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Notes:
(a) Chi-squared ($\chi^2$) test to ascertain the significance of the decrease in unexplained variance from an intercept only model to the overall model [66], based on the null hypothesis that each regression coefficient in the model is zero [67]. The -2 log likelihood (-2LL) statistic used in this test is given by -2(LL(intercept model) - LL(overall model)) [66, 67]. This $\chi^2$ test is based on model’s corresponding degrees of freedom (df) [67].
(b) $\chi^2$ test to ascertain the significance of explanatory factors to the overall model [66]. This -2LL statistic is given by -2(LL(overall model) - LL(overall model without the factor under test)) [66]. This $\chi^2$ test is based on explanatory factor’s corresponding degrees of freedom (df).
(c) See notes (a) and (b).
(d) Nagelkerke’s pseudo-$R^2$ statistic is a measure of the improvement in fit of the overall model compared to a model with no independent variables. This statistic has a range of 0 to 1 and is analogous to the coefficient of determination ($R^2$) statistic used in ordinary least squares regression [67].
(e) Sub-model categories: 0≤Low<Q1 and Q3<High≤Max; where Q1, Q3 and Max are the first and third quartiles and the maximum value of the modelled distribution.
(f) Reference category level. (g) Sample size (n). (h) Coefficient ($\beta$). (i) Standard Error (SE).(j) Reference factor level.
*(, ** and *** significance at 0.1, 0.05, 0.01 and 0.001 levels respectively.*
### Table 5.5: Multinomial logistic regression model for the base temperature parameter ($T_b$)

<table>
<thead>
<tr>
<th>Base Temperature Model</th>
<th>$\chi^2$ test of -2LL (df) (c)</th>
<th>Pseudo-$R^2$(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Model (a)</td>
<td>40.26 (22) *</td>
<td>0.11</td>
</tr>
<tr>
<td>Explanatory Factors (b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature Set-point</td>
<td>10.28 (12)</td>
<td></td>
</tr>
<tr>
<td>Timer Control</td>
<td>10.07 (4) *</td>
<td></td>
</tr>
<tr>
<td>Operating Hours</td>
<td>17.91 (6) **</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sub-models (e)</th>
<th>Low</th>
<th>Medium (f)</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n°(g)</td>
<td>β(h)</td>
<td>SE(i)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.50</td>
<td>0.29</td>
<td>0.61</td>
</tr>
<tr>
<td>Temperature Set-point:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 20°C (j)</td>
<td>33</td>
<td>51</td>
<td>26</td>
</tr>
<tr>
<td>&lt; 18°C</td>
<td>8</td>
<td>0.08</td>
<td>0.54</td>
</tr>
<tr>
<td>21°C</td>
<td>6</td>
<td>-0.61</td>
<td>0.53</td>
</tr>
<tr>
<td>22 - 24°C</td>
<td>7</td>
<td>-0.14</td>
<td>0.53</td>
</tr>
<tr>
<td>&gt; 24°C</td>
<td>2</td>
<td>-0.50</td>
<td>0.87</td>
</tr>
<tr>
<td>No Thermostat</td>
<td>32</td>
<td>-0.29</td>
<td>0.31</td>
</tr>
<tr>
<td>Unknown</td>
<td>14</td>
<td>-0.14</td>
<td>0.41</td>
</tr>
<tr>
<td>Timer Control:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Zone (j)</td>
<td>58</td>
<td>130</td>
<td>71</td>
</tr>
<tr>
<td>Separate Zones</td>
<td>20</td>
<td>0.65</td>
<td>0.36</td>
</tr>
<tr>
<td>No Timer/Not Used</td>
<td>24</td>
<td>0.29</td>
<td>0.31</td>
</tr>
<tr>
<td>Operating Hours:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 &lt; hours/day ≤ 8 (j)</td>
<td>48</td>
<td>81</td>
<td>23</td>
</tr>
<tr>
<td>8 &lt; hours/day ≤ 10</td>
<td>28</td>
<td>0.00</td>
<td>0.31</td>
</tr>
<tr>
<td>10 &lt; hours/day ≤ 12</td>
<td>13</td>
<td>-0.36</td>
<td>0.39</td>
</tr>
<tr>
<td>12 &lt; hours/day ≤ 24</td>
<td>13</td>
<td>-0.47</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Notes:

(a) Chi-squared ($\chi^2$) test to ascertain the significance of the decrease in unexplained variance from an intercept only model to the overall model [66], based on the null hypothesis that each regression coefficient in the model is zero [67]. The -2 log likelihood (-2LL) statistic used in this test is given by: $-2(LL(\text{intercept model}) - LL(\text{overall model}))$ [66, 67]. This $\chi^2$ test is based on model’s corresponding degrees of freedom (df) [67].

(b) $\chi^2$ test to ascertain the significance of explanatory factors to the overall model [66]. This -2LL statistic is given by: $-2(LL(\text{overall model}) - LL(\text{overall model without the factor under test}))$ [66, 67]. This $\chi^2$ test is based on explanatory factor’s corresponding degrees of freedom (df).

(c) See notes (a) and (b).

(d) Nagelkerke’s pseudo-$R^2$ statistic is a measure of the improvement in fit of the overall model compared to a model with no independent variables. This statistic has a range of 0 to 1 and is analogous to the coefficient of determination ($R^2$) statistic used in ordinary least squares regression [67].

(e) Sub-model categories: $0 \leq \text{Low} < Q1$ and $Q3 < \text{High} \leq \text{Max};$ where Q1, Q3 and Max are the first and third quartiles and the maximum value of the modelled distribution.

(f) Reference category level. (g) Sample size (n). (h) Coefficient ($\beta$). (i) Standard Error (SE). (j) Reference factor level.

*, *, ** and *** significance at 0.1, 0.05, 0.01 and 0.001 levels respectively.
It can be seen that each statistically significant coefficient ($\beta$) estimate in the MLR models is consistent with the domestic gas consumption dynamics described in Table 5-1. This is confirmed by the following characterisations of the intercept, slope and base temperature parameter distributions:

- Dwellings with low intercepts ($b_0$) are unlikely to be occupied by three or more adults, given this factor’s low odds-ratio ($\text{Exp}(\beta)$) value, and are highly likely to use alternative hot water and cooking systems, given these factors high odds-ratios. Those with high intercepts are unlikely to be occupied by a single adult and to use an alternative hot water system.

- Dwellings with low slopes ($b_1$) are likely to have no more than two bedrooms, and to have been built since 1980. Those with high slopes are likely to have four or more bedrooms, are likely to be detached dwellings rather than apartment or terrace type dwellings and are unlikely to have been built since 1980.

- Dwellings with low base temperatures ($T_B$) are likely to use zoned time control systems. High base temperature dwellings are unlikely to use zoned time control systems, and are likely to have their heating systems operated for over eight hours each day, although this characterisation is not statistically significant for the ten to eleven hours category.

5.2.3 Energy Efficiency Assessments

In this section the MLR models presented in Tables 5.3, 5.4 and 5.5 are used to compare the relative energy end-use levels of consumers with the same household characteristics
in order to identify buildings with unexpectedly high intercept, slope and base temperature estimates.

In Table 5.6, intercept parameters are presented for three sample consumers – Consumer No. 1, 2 and 3. It is seen that these consumers have low, medium and high intercept parameter estimates, respectively, even though they share the same household characteristics. Based on these characteristics, 9%, 58% and 33% probabilities have been predicted for the low, medium and high intercept categories, respectively, using the MLR probability formulae in Equations 5.4-5.7 and the relevant $\beta$ coefficients in Table 5.3. Therefore, Consumer No. 3 has an unexpectedly high intercept parameter estimate; thus indicating unusually high hot water and cooking consumption. This may be due to an inefficient hot water heating system, poor hot water cylinder insulation, or high hot water consumption by the occupants, relative to the other consumers in the Table. Energy saving opportunities should be explored for this consumer in this regard. For example, this consumer could: 1) decrease the number of operating hours set by their hot water system’s timer, 2) upgrade their hot water cylinder’s insulation, and/or 3) decrease its temperature set-point, if such a control system is present. In addition, it is estimated that Consumer No. 3 spends approximately €425/year on cooking and hot water (14.51kWh/day (intercept) x 365days/year x €0.08/kWh) at current Irish gas market rates. This estimate may be used to assess the viability of installing a solar hot water heating system or boiler upgrade based on current cost estimates.
In Table 5.6, the estimated slope parameters are presented for another three consumers – Consumer No. 4, 5 and 6. It is seen that Consumer No. 6 has an unexpectedly high slope parameter estimate. This indicates that this dwelling may have an inefficient space heating system or a building fabric with poor thermal insulation levels, relative to the other consumers in the Table. Therefore, this consumer may benefit from a boiler or building fabric upgrade. It is estimated that this consumer spent approximately €660 on space heating for the previous year \((7.63 \text{kWh/°C·day}) \times 1078.72 \text{°C·day/year} \times €0.08/\text{kWh}\), where the total HDDs for the year is estimated using the dwelling’s base temperature). This estimate may be used to assess the viability of boiler or building fabric upgrades based on current cost estimates.

In Table 5.6, the estimated base temperature parameters are presented for another three consumers – Consumer No. 7, 8 and 9. It is seen that Consumer No. 9 has an unexpectedly high base temperature parameter estimate, relative to the other consumers in the Table. Such consumers could be targeted with behavioural change programmes or zoned heating control systems. If for example, behavioural change or zoning results in a

---

### Table 5.6: Energy efficiency assessments

<table>
<thead>
<tr>
<th>Consumer No.</th>
<th>No. 1</th>
<th>No. 2</th>
<th>No. 3</th>
<th>No. 4</th>
<th>No. 5</th>
<th>No. 6</th>
<th>No. 7</th>
<th>No. 8</th>
<th>No. 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Intercept ((b_0))</td>
<td>Slope ((b_1))</td>
<td>Base temperature ((T_B))</td>
<td>Intercept ((b_0))</td>
<td>Slope ((b_1))</td>
<td>Base temperature ((T_B))</td>
<td>Intercept ((b_0))</td>
<td>Slope ((b_1))</td>
<td>Base temperature ((T_B))</td>
</tr>
<tr>
<td>Estimate</td>
<td>2.73</td>
<td>8.21</td>
<td>14.51</td>
<td>3.18</td>
<td>5.84</td>
<td>7.63</td>
<td>11.66</td>
<td>14.0</td>
<td>15.54</td>
</tr>
<tr>
<td>Standard Error</td>
<td>1.72</td>
<td>1.87</td>
<td>2.65</td>
<td>0.12</td>
<td>0.16</td>
<td>0.26</td>
<td>0.31</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>Category</td>
<td>Low</td>
<td>Med.</td>
<td>High</td>
<td>Low</td>
<td>Med.</td>
<td>High</td>
<td>Low</td>
<td>Med.</td>
<td>High</td>
</tr>
<tr>
<td>Characteristics</td>
<td>2 adults</td>
<td>3 bedrooms</td>
<td>18-20°C temp. set-point</td>
<td>0 children</td>
<td>Semi-detached</td>
<td>Single zone timer</td>
<td>Timed gas fuelled hot</td>
<td>1980-1999 construct. year</td>
<td>0 - 8 operating hours</td>
</tr>
<tr>
<td>Category</td>
<td>Low</td>
<td>Probability</td>
<td>9%</td>
<td>41%</td>
<td>32%</td>
<td>Medium</td>
<td>58%</td>
<td>53%</td>
<td>52%</td>
</tr>
</tbody>
</table>

---
nominal 1°C reduction in base temperature, a saving of approximately €140 was possible in the modelled year for this consumer (5.53 kWh/°C·day (slope) x (2365.45-2050.41)°C·day/year x €0.08/kWh, where the reduction in HDDs is estimated using the total HDDs for a 1°C reduction in the dwelling’s base temperature parameter).

5.3 Conclusion

This chapter presents a NLS method to estimate individual HDD building energy models using daily gas consumption data. Such data will soon be readily available as smart metering infrastructure is deployed across Europe. The method was used to the estimate individual HDD regression models for a representative sample of Irish domestic dwellings.

The chapter also demonstrated a MLR modelling method based on the resulting intercept, slope and base temperature distributions that can be used to compare the inferred gas end-uses of individual dwellings. These MLR models have been presented as an alternative to energy intensity metrics based on building floor area; as it was found that a large proportion of consumers in the sample did not know the area of their dwelling. By way of example, the MLR models were used to compare the energy efficiency of similar buildings based on their intercept, slope and base temperature estimates.

However, the MLR models were limited to low, medium and high categories for each of the intercept, slope and base temperature parameter distributions. This limitation was applied for simple classifications of each parameter distribution and to reduce the size of the MLR models in Tables 5.3, 5.4 and 5.5. It is recommended that the number of
categories for each parameter distribution is increased for larger consumer samples. This increase can allow the energy efficiency of buildings at lower regions in each distribution to be compared.
CHAPTER 6

PEAK DAY FORECASTING
6 PEAK DAY FORECASTING

This chapter develops and implements a methodology to forecast year-ahead peak day gas consumption for NDM gas markets. This can be used by network operators to establish the peak supply capacity of their network to alternative supply standards. This peak-day forecasting method is based on the NDD_{CA} parameter given by Equations 4.58-4.64, and multiple linear regression and generalised extreme value (GEV) modelling methods that are described later in this chapter. The main steps in the method are summarised in Figure 6.1.

![Diagram of the peak-day forecasting method]

**Figure 6.1:** Summary of the peak-day forecasting method.
The method begins by using NLS to estimate a multiple linear regression model of weekday gas consumption for the Irish NDM gas market. Only weekday consumption is modelled because this is when most gas is consumed by the market due to increased commercial activity and therefore when peak-day gas consumption is most likely to occur. The resulting multiple linear regression model including the coefficients of weather variables within the NDD\textsubscript{CA} parameter are then presented. These coefficients are then used to calculate long-term (>30 years) NDD\textsubscript{CA} data for extreme value modelling. The resulting extreme value models are then used to extrapolate various return levels (e.g. 1-in-50 year values) of the NDD\textsubscript{CA} variable for peak-day forecasting. These extreme NDD\textsubscript{CA} values and the multiple linear regression model are then used to forecast year-ahead peak day gas consumption for the Irish NDM gas market to alternative supply standards.

The development of this method is also used as an opportunity to assess the benefit of the new NDD\textsubscript{CA} parameter compared to other parameters applied by European network operators to forecast peak gas consumption. These alternative parameters are compared to the NDD\textsubscript{CA} parameter in Table 6.1. It is seen the UK-CWV is the closest alternative parameter, but does not account for solar radiation. The benefit of the NDD\textsubscript{CA} parameter, in terms of modelling accuracy, is assessed in this chapter by a comparative analysis of each of the effects accounted for by the parameter.
<table>
<thead>
<tr>
<th>Effect</th>
<th>Use elsewhere?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outdoor temperature</td>
<td>- EU supply standard [10].</td>
</tr>
<tr>
<td></td>
<td>- Belgium [41].</td>
</tr>
<tr>
<td></td>
<td>- Czech Republic [42].</td>
</tr>
<tr>
<td></td>
<td>- Denmark [43].</td>
</tr>
<tr>
<td></td>
<td>- France [45].</td>
</tr>
<tr>
<td></td>
<td>- Ireland, see Section 3.1.3.1.</td>
</tr>
<tr>
<td></td>
<td>- The Netherlands [46].</td>
</tr>
<tr>
<td></td>
<td>- The UK, see Section 3.2.2.</td>
</tr>
<tr>
<td>Wind speed</td>
<td>- Irish, using a wind chill function as described in Section 3.1.3.1.</td>
</tr>
<tr>
<td></td>
<td>- The UK, using a wind chill function given by Equation 3.19.</td>
</tr>
<tr>
<td>Solar radiation</td>
<td>Solar radiation is not modelled in the alternative European weather parameters.</td>
</tr>
<tr>
<td>Thermal memory</td>
<td>- France, using the effective temperature given in Table 3-3.</td>
</tr>
<tr>
<td></td>
<td>- Ireland, using a weighted HDD filter as described in Section 3.1.3.1.</td>
</tr>
<tr>
<td></td>
<td>- The UK, using the effective temperature given by Equation 3.17.</td>
</tr>
<tr>
<td>Variable base temperatures</td>
<td>- The UK, using the transformation function given by Equation 3.20 and illustrated in Figure 3.5.</td>
</tr>
<tr>
<td>Seasonal consumption</td>
<td>- Ireland, using seasonal HDDs as described in Section 3.1.3.1.</td>
</tr>
<tr>
<td></td>
<td>- The UK, using a ‘pseudo’ seasonal effective temperature, see Equation 3.16 and Section 3.2.2.</td>
</tr>
</tbody>
</table>
6.1 Methodology

This section begins with a summary of the data used to develop the peak day gas consumption forecasting method. Next, the multiple linear regression model of weekday NDM market consumption is described, including the NLS method used to estimate both the parameters of the model and the NDDCA variable. The section is completed by a description of the GEV modelling method used to extrapolate extreme values of the NDDCA variable for peak day forecasting.

6.2 Data

The NDM market gas consumption data used to develop the peak day gas consumption forecasting method were previously described in Section 2.2. Corresponding climate data were also described in Section 2.3.

6.2.1 NDM Market Gas Consumption Model

The parameters of the NDDCA variable are estimated within the following gas consumption model using the NLS method in R [64].

\[
\begin{align*}
C_{\text{NDM,WD}} &= b_0 + \Delta b_{0,\text{Oct'10-Sept'11}}DV_{\text{Oct'10-Sept'11}} + \Delta b_{0,\text{Oct'09-Sept'10}}DV_{\text{Oct'09-Sept'10}} + \\
&\quad b_1NDD_{\text{CA,WD}}(T_{B,\text{upr}}, T_{B,\text{lwr}}, \omega_1, \alpha_1, \gamma_1, \gamma_2, m, n) + \\
&\quad \Delta b_{1,\text{Oct'10-Sept'11}}DV_{\text{Oct'10-Sept'11}}NDD_{\text{CA,WD}}(T_{B,\text{upr}}, T_{B,\text{lwr}}, \omega_1, \ldots, n) + \\
&\quad \Delta b_{1,\text{Oct'09-Sept'10}}DV_{\text{Oct'09-Sept'10}}NDD_{\text{CA,WD}}(T_{B,\text{upr}}, T_{B,\text{lwr}}, \omega_1, \ldots, n) + \\
&\quad \varepsilon_{\text{WD}}
\end{align*}
\]

where: \(C_{\text{NDM,WD}}\) is NDM market gas consumption for a normal weekday (WD), excluding public holidays and the Christmas period: 24th December to the day before the first working day in the New Year;
$b_0$ is the model intercept for the most recent gas-year: Oct’11-Sept’12, and an estimate of the market’s average daily weather-independent gas consumption;

$\Delta b_{0i}$ are differential intercepts required for the other gas-years;

$DV$s are dummy variables to indicate the other gas-years, which have a value of 1 if a data point is in the designated year, otherwise the value is zero, thus allowing different linear models to apply to each of the three gas-years examined;

$b_1$ is the NDD$_{CA}$ coefficient for the most recent gas-year: Oct’11-Sept’12, the product of which is an estimate of the market’s daily weather-dependent gas consumption;

$\Delta b_{1i}$ are differential NDD$_{CA}$ coefficients required for the other gas-years;

$T_{B,upr}$, $T_{B,lwr}$, $\omega_1$, $\alpha_1$, $\gamma_1$ & $\gamma_2$ are non-linear parameters in the NDD$_{CA}$ variable, see Equations 4.58 - 4.64;

$m$ and $n$ are manually specified parameters in the SS-NDD$_{CA}$ variable (Equation 4.63), which represent the moving average window width and the number of years used to calculate each S-NDD$_{WA}$ value (Equation 4.64), here they are given by 13 days and 30 years, respectively, however they could be profiled separately if required; and finally

$\varepsilon_{WD}$ is the residual (model error) for a given weekday.
Alternative day-types such as weekends, public holidays and the Christmas period are excluded in the modelling process as they would each require additional differential intercepts and slopes, thus adding unnecessary complication to the model.

The model is estimated using consumption data for the three gas years between October 2009 and September 2012. This was to allow the extremely cold winter periods of January and December 2010 to be used to evaluate the accuracy of the NDD\textsubscript{CA} parameter for such extreme weather. However, as this consumption is over three years it may also be affected by a number of influences such as: increased consumer numbers; improved building fabric standards; or housing energy efficiency programs. Over time, such effects can impact on the estimated intercept and slope of a linear gas consumption model. This is accounted for in Equation 6.1 using differential intercepts and slopes for each of the preceding gas years to that of the most recent gas year. To help convergence to a global NLS solution, starting values are stipulated for each parameter as shown in Table 6.2.

**Table 6.2:** NDM market model parameter starting values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Starting Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>5</td>
</tr>
<tr>
<td>$b_1$</td>
<td>5</td>
</tr>
<tr>
<td>$\Delta b_{0\text{+Oct'10-Sept'11}}$, $\Delta b_{0\text{+Oct'09-Sept'10}}$, $\Delta b_{1\text{+Oct'10-Sept'11}}$ and $\Delta b_{1\text{+Oct'09-Sept'10}}$</td>
<td>0</td>
</tr>
<tr>
<td>$T_{B,upr}$</td>
<td>20</td>
</tr>
<tr>
<td>$T_{B,low}$</td>
<td>15</td>
</tr>
<tr>
<td>$\omega_1$, $\gamma_1$, $\gamma_2$ and $\alpha_1$</td>
<td>0</td>
</tr>
</tbody>
</table>
The $b_0$ and $b_1$ linear parameters estimated for the most recent gas year from this NLS solution are then used to estimate peak daily gas consumption for the following heating season, as follows:

$$\hat{C}_{NDM,WD,Oct'12-Sept'13} = b_0 + b_1 NDD_{CA} \quad (6.2)$$

where $\hat{C}_{NDM,WD,Oct'12-Sept'13}$ is the forecasted peak day gas consumption for the next gas-year: Oct'12-Sept'13, using $NDD_{CA}$ return levels quantified by a separate extreme value model.

### 6.2.2 Extreme Value Modelling

The generalised extreme value (GEV) model is used in this chapter to estimate various return levels of the $NDD_{CA}$ parameter. The cumulative distribution function (CDF) of this model is given by [68]:

$$F(x) = \begin{cases} 
\exp \left\{ - \left[ 1 - \frac{k(x - \mu)}{\sigma} \right]^{1/k} \right\}, & \text{for } k \neq 0, \\
\exp \left\{ - \exp \left[ - \frac{(x - \mu)}{\sigma} \right] \right\}, & \text{for } k = 0.
\end{cases} \quad (6.3)$$

where: $\mu$, $\sigma$, and $k$ are the location, scale and shape parameters, respectively.

In this study, this model is fitted to observed gas-year or block maxima $NDD_{CA}$ values, so that various return levels for the parameter can be extrapolated. These return levels are estimated using the inverse distribution function of the above CDF [68]:
\( x(F) = \begin{cases} \mu + \frac{\sigma}{k}[1 - \{-log(F)\}^k], & \text{for } k \neq 0, \\ \mu - a\log\{-\log(F)\}, & \text{for } k = 0. \end{cases} \) \hspace{1cm} (6.4)

where \( x(F) \) is the return level of a parameter for a given return period \( (P) \), i.e. the level estimated to be exceeded on average once every \( P \) years, and \( F = 1 - 1/P \).

Each of the GEV models required in this study are estimated by the probability weighted moments method [69].

### 6.2.2.1 Goodness-of-Fit Test

The goodness of fit of these models is assessed using Anderson-Darling (AD) and modified-AD (upper-tail) tests. These tests are performed due to limited years of climate data available to calculate the \( \text{NDD}_{CA} \) parameter (see Section 2.3), and are used to ensure that the fitted GEV models are appropriate and can be used to extrapolate the required return levels. The AD and modified-AD tests are given by the following computational formulae [70]:

\[
A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)\left[\log\{F(x_i)\} + \log\{1 - F(x_{n+1-i})\}\right] \hspace{1cm} (6.5)
\]

\[
AU_n^2 = \frac{n}{2} - 2 \sum_{i=1}^{n} F(x_i) - \sum_{i=1}^{n} \left\{2 - \frac{2i - 1}{n}\right\}\log\{1 - F(x_i)\} \hspace{1cm} (6.6)
\]

The AD test is used initially to test if the empirical CDF follows the hypothesised distribution. Its statistic \( A^2 \), is the sum of the squares of differences between the hypothesised distribution \( F(x) \) to the empirical CDF, \( F_n(x) \) over the ordered sample \( x_1 \),
\( x_1, \ldots, x_n \); using a weighting function that emphasises differences at both tails. The similar, modified-AD test is then applied since it uses a weighting function that gives greater emphasis to deviations at the upper tail of the distribution; this relates to the high return period region (\( \geq 20 \) years) of the GEV distribution where this study is primarily concerned.

The fitted GEV models pass these AD tests if the \( A^2 \) and \( AU^2 \) statistics are less than their respective critical values for a specified significance level. Critical values are used to define the limiting value below which the null hypothesis \( (H_0) \) that the empirical data follow the fitted distribution is accepted. Such critical values for these AD tests must be calculated for the specific distribution under test and can be estimated using Monte Carlo methods. However in this study, critical values are reported for the closest sample size from tabulated data published for GEV models [69].

### 6.3 Results and Discussion

The results of this chapter begin with the implementation of the weekday gas consumption model, including the estimation of the NDD\(_{CA}\) parameters. The accuracy of this model is assessed using standard metrics and results are then presented for the estimation of peak day gas consumption based on alternative supply standards.

#### 6.3.1 Gas Consumption Model

In this section the gas consumption model (Equation 6.1) and the NDD\(_{CA}\) variable (Equations 4.58 - 4.64) are described in terms of: 1) the parameter estimates; 2) the in-sample model accuracy; and 3) a comparative analysis of the effects modelled by the NDD\(_{CA}\) parameter with respect to model accuracy.
The parameter estimates for the gas consumption model and the NDD\textsubscript{CA} variable are shown in Table 6.3. It is seen that the intercept ($b_0$) and slope ($b_1$) parameters of the linear model for the most recent gas year are highly significant ($p < 0.001$), while the difference in intercepts ($\Delta b_0$s) and slopes ($\Delta b_1$s) for previous gas years relative to the most recent gas year are significant for the intercepts ($p<0.02$) and highly significant for the slopes. It is also seen that each of the NDD\textsubscript{CA}’s internal parameters have insignificant standard errors.

The in-sample modelling accuracy of this weekday gas consumption model is shown in the upper plot of Figure 6.2. This plot shows the weekday gas consumption values for the three years examined (October 2009 to September 2012), together with the fitted linear models for each of the three years. It is seen that NDD\textsubscript{CA} weather parameter estimates consumption for each gas year in a multiple linear regression model with a strong coefficient of determination ($R^2$) of 0.9859 and a mean absolute percentage error (MAPE) of 7.81\%. It is also seen that a MAPE value of 3.53\% is obtained for the upper 5\% of consumption values, above 70.6GWh - the majority of which were observed during the extremely cold winter periods in January and December 2010.
Table 6.3: NLS solution summary for the gas consumption model and the NDD$_{CA}$ parameter.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>5.983</td>
<td>0.677</td>
<td>8.838</td>
<td>7.19E-18 ***</td>
</tr>
<tr>
<td>$\Delta b_{0,Oct'10-Sept'11}$</td>
<td>-0.955</td>
<td>0.4089</td>
<td>-2.335</td>
<td>0.0198 *</td>
</tr>
<tr>
<td>$\Delta b_{0,Oct'09-Sept'10}$</td>
<td>-0.9329</td>
<td>0.3945</td>
<td>-2.365</td>
<td>0.0183 *</td>
</tr>
<tr>
<td>$b_1$</td>
<td>4.749</td>
<td>0.1303</td>
<td>36.44</td>
<td>1.65E-166 ***</td>
</tr>
<tr>
<td>$\Delta b_{1,Oct'10-Sept'11}$</td>
<td>0.3329</td>
<td>0.06131</td>
<td>5.429</td>
<td>7.70E-08 ***</td>
</tr>
<tr>
<td>$\Delta b_{1,Oct'09-Sept'10}$</td>
<td>0.2423</td>
<td>0.05968</td>
<td>4.059</td>
<td>5.45E-05 ***</td>
</tr>
<tr>
<td>$T_{B,upr}$</td>
<td>22.11</td>
<td>0.4506</td>
<td>-</td>
<td>- (a)</td>
</tr>
<tr>
<td>$T_{B,low}$</td>
<td>7.046</td>
<td>0.5559</td>
<td>-</td>
<td>- (a)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.001774</td>
<td>6.61E-05</td>
<td>26.85</td>
<td>4.64E-111 ***</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.01799</td>
<td>0.001444</td>
<td>12.46</td>
<td>1.86E-32 ***</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.4967</td>
<td>0.01287</td>
<td>38.6</td>
<td>1.41E-178 ***</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.3408</td>
<td>0.01279</td>
<td>26.65</td>
<td>7.32E-110 ***</td>
</tr>
</tbody>
</table>

Significance Levels: ‘.’ $p < 0.1$, ‘*’ $p < 0.05$, ‘**’ $p < 0.01$, ‘***’ $p < 0.001$

Note (a): standard significance tests are inappropriate for the base temperature parameters.
Figure 6.2: Linear relationship between gas consumption for both the NDD$_{CA}$ and the HDD parameters.
In Table 6.4 a comparative analysis of each of the effects modelled by the $\text{NDD}_{\text{CA}}$ parameter is shown with respect to the in-sample modelling accuracy of the weekday gas consumption model. In this assessment, incrementally adjusted HDD parameters are used in place of the $\text{NDD}_{\text{CA}}$ parameter in the weekday gas consumption model of Equation 6.1. These parameters include: the HDD variable in Equation 4.7, the $\text{HDD}_{\text{WA}}$ variable in Equation 4.54 with various permutations of its parameters set to zero, and the $\text{NDD}_{\text{WA}}$ variable in Equation 4.59.

In Table 6.4 it is seen that the in-sample model accuracy of the gas consumption model improves with each of the incremental adjustments to the HDD variable, and that the $\text{NDD}_{\text{CA}}$ variable is the most accurate estimator overall. This is shown by the model’s $R^2$ and MAPE values, in relation to the overall accuracy of the model and for upper consumption values. It can also be seen that the solar radiation and the effective outdoor temperature parameters both account for the largest increase in model accuracy from that observed for the HDD parameter, as $R^2$ increases from 0.8452 to 0.9372 and 0.9214 respectively when these parameters are included separately, to 0.9692 when they are included together. The in-sample modelling accuracy of weekday gas consumption models using these incrementally adjusted HDD variables are also shown in Figures A.1 to A.9 in the Appendix.
Table 6.4: NLS solution and in-sample accuracy summaries of incrementally updated HDD based parameters.

<table>
<thead>
<tr>
<th></th>
<th>HDD(a)</th>
<th>HDD_{WA}^{(b)} (γ_1&amp;α_1=0)</th>
<th>HDD_{WA}^{(c)} (γ_2&amp;α_1=0)</th>
<th>HDD_{WA}^{(d)} (γ_1&amp;γ_2=0)</th>
<th>HDD_{WA}^{(e)} (γ_1=0)</th>
<th>HDD_{WA}^{(f)} (γ_2=0)</th>
<th>HDD_{WA}</th>
<th>NDD_{WA}</th>
<th>NDD_{CA}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NLS Solution Summary:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( T_B )</td>
<td>14.01</td>
<td>14.54</td>
<td>17.91</td>
<td>13.57</td>
<td>13.66</td>
<td>16.33</td>
<td>16.07</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( T_{B,upr} )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>20.24</td>
<td>22.11</td>
</tr>
<tr>
<td>( T_{B, lw} )</td>
<td>-</td>
<td>-</td>
<td>0.0036</td>
<td>-</td>
<td>-</td>
<td>0.0023</td>
<td>0.0021</td>
<td>0.0021</td>
<td>0.0018</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>-</td>
<td>0.0398</td>
<td>-</td>
<td>-</td>
<td>0.0248</td>
<td>-</td>
<td>0.0128</td>
<td>0.0125</td>
<td>0.018</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>-</td>
<td>-</td>
<td>0.7041</td>
<td>0.6575</td>
<td>0.6547</td>
<td>0.6127</td>
<td>0.6154</td>
<td>0.4967</td>
<td></td>
</tr>
<tr>
<td>( \omega_1 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.3408</td>
<td></td>
</tr>
<tr>
<td><strong>In-Sample Model Accuracy:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.8452</td>
<td>0.8749</td>
<td>0.9372</td>
<td>0.9214</td>
<td>0.9359</td>
<td>0.9692</td>
<td>0.9734</td>
<td>0.9755</td>
<td>0.9859</td>
</tr>
<tr>
<td>MAPE(^{\text{g)}},</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>26.75</td>
<td>24.19</td>
<td>16.93</td>
<td>17.87</td>
<td>16.43</td>
<td>11.82</td>
<td>11.28</td>
<td>10.16</td>
<td>7.805</td>
</tr>
</tbody>
</table>

Independent Variable Notes:

a) Outdoor temperature only (see Equation 4.7).
b) Wind speed and outdoor temperature only (see Equation 4.54).
c) Solar radiation and outdoor temperature only (see Equation 4.54).
d) Effective outdoor temperature (thermal memory) only (see Equation 4.54).
e) Wind speed and effective outdoor temperature only (see Equation 4.54).
f) Solar radiation and effective outdoor temperature only (see Equation 4.54).

Model Accuracy Note:

\( g \) MAPEs are reported for the linear model: overall and for the upper 5% of consumption values above 70.6 GWh (see Figure 6.2).
6.3.2 Peak Consumption Estimates

In this study, peak daily gas consumption is estimated from two alternative supply standard perspectives, based on: 1) a peak day supply standard that applies a 1-in-50 year NDD_{CA} return level – similar to Irish supply standard; and 2) a peak week supply standard that applies a 1-in-20 year 7-day average NDD_{CA} return level – similar to the EU’s new supply standard. These estimates are used to illustrate the relative difference in peak daily consumption levels required by such alternative standards.

6.3.2.1 Peak Day Supply Standard

In this section, peak daily gas consumption is estimated for the Irish domestic and SME gas market for the year ahead, based on a 1-in-50 year NDD_{CA} (or NDD_{CA,0.02}) return level extrapolated from a GEV model.

Table 6.5 presents the results for the GEV model fitted to gas-year or block maxima NDD_{CA} values that have been sampled from a NDD_{CA} series calculated using the NDD_{CA} parameter estimates of Table 6.3 and climate data since 1976 (see Section 2.3). The results of the Anderson-Darling goodness of fit tests in Table 6.5 are used to compare the GEV distribution to the empirical data. It is seen that a positive result has been found for both tests, which confirms the suitability of this GEV distribution as an appropriate model of the observed block maxima NDD_{CA} series. This GEV model fit can be further assessed by the density distribution and return level plots shown in Figures 6.3 and 6.4.
Table 6.5: Block maxima NDD_{CA} GEV model solution summary.

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13.74</td>
<td>1.333</td>
<td>-0.05375</td>
</tr>
</tbody>
</table>

Goodness of Fit: $A^2(n=36)$ $AU^2(n=36)$

<table>
<thead>
<tr>
<th></th>
<th>$A^2(n=36)$</th>
<th>$AU^2(n=36)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.3548</td>
<td>0.2147</td>
</tr>
</tbody>
</table>

Critical values [69]: $A^2(n=35)$ $AU^2(n=35)$

<table>
<thead>
<tr>
<th></th>
<th>$A^2(n=35)$</th>
<th>$AU^2(n=35)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.572</td>
<td>0.2754</td>
</tr>
</tbody>
</table>

Null Hypothesis ($H_0$) Accepted for both tests

Goodness of Fit Notes:

$H_0$: the empirical data follow the fitted GEV distribution.

Significance level: 0.05

Critical region: $H_0$ is rejected if goodness of fit result is greater than the critical values reported.

Figure 6.3: Block maxima NDD_{CA} histogram and the GEV model’s density distribution.
Figure 6.4: Empirical (points) and the GEV model’s (black line) NDD\textsubscript{CA} return level estimates.

Figure 6.4 also reports a 1-in-50 year NDD\textsubscript{CA} return level of 19.53°C·day, which was calculated using the inverse distribution function in Equation 6.4 and the GEV model parameter estimates in Table 6.5. Based on this return level, the 1-in-50 year peak day gas consumption for the next gas year (2012-’13) is estimated as 98.72 GWh, using the gas consumption forecast model of Equation 6.2 and the $b_0$ and $b_1$ parameters reported in Table 6.3.

Although the gas consumption model was estimated using weekday gas consumption, this distinction was not applied to the long-term NDD\textsubscript{CA} series as part of the GEV modelling process. An adjustment is required to account for the probability of an extreme NDD\textsubscript{CA} value occurring at a weekend or on a holiday when gas consumption is lower than on weekdays. The proportion of normal-weekdays to the number of days in the long-term climate data series is used to scale down the 1-in-50 year return period to
an effective return period which can be used to calculate a 1-in-50 year ‘weekday \( \text{NDD}_{\text{CA}} \)’ return level from the above GEV model. Based on this method, the effective return period is approximately 34 years, with a 1-in-50 year ‘weekday \( \text{NDD}_{\text{CA}} \)’ return level of 18.89°C·day. From this, the 1-in-50 year peak weekday consumption is reduced from 98.72 GWh to 95.7 GWh.

### 6.3.2.2 Peak Week Supply Standard

In this section, a peak week supply standard is applied to the Irish NDM gas market to illustrate: 1) the relative difference in peak daily consumption levels required by this standard compared to a peak day supply standard; and 2) to identify any change in GEV goodness of fit results for the 7-day average \( \text{NDD}_{\text{CA,AVG,7D}} \) variable compared to the daily \( \text{NDD}_{\text{CA}} \) variable.

Table 6.6 presents the results for the GEV model fitted to gas-year or block maxima \( \text{NDD}_{\text{CA,AVG,7D}} \) values that have been sampled from the \( \text{NDD}_{\text{CA,AVG,7D}} \) series calculated using the \( \text{NDD}_{\text{CA}} \) series since 1976 and a 7-day moving-average filter. The goodness of fit results for this GEV model in Table 6.6 are better than those observed for the GEV model in Table 6.5. It is seen that the \( A^2 \) and \( AU^2 \) statistics have decreased from 0.3548 to 0.119 and 0.2147 to 0.04578, respectively, and are further away from the corresponding critical values of 0.572 and 0.2754. This improvement in GEV model fit can also be seen by comparing the density distribution and return level plots in Figures 6.5 and 6.6 to Figures 6.3 and 6.4.
### Table 6.6: Block maxima NDD\textsubscript{CA,AVG,7D} GEV model solution summary.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12.61</td>
<td>1.28</td>
<td>0.03035</td>
</tr>
</tbody>
</table>

**Goodness of Fit:**

<table>
<thead>
<tr>
<th>Test</th>
<th>$A^2(n=36)$</th>
<th>$AU^2(n=36)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.119</td>
<td>0.04578</td>
</tr>
</tbody>
</table>

**Critical values [69]:**

<table>
<thead>
<tr>
<th>Test</th>
<th>$A^2(n=35)$</th>
<th>$AU^2(n=35)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.572</td>
<td>0.2754</td>
</tr>
</tbody>
</table>

**Null Hypothesis ($H_0$)** Accepted for both tests.

**Goodness of Fit Notes:**

- $H_0$: the empirical data follow the fitted GEV distribution.
- Significance level: 0.05
- Critical region: $H_0$ is rejected if goodness of fit result is greater than the critical values reported.

### Figure 6.5: Block maxima NDD\textsubscript{CA,AVG,7D} histogram and the GEV model’s density distribution.
**Figure 6.6:** Empirical (points) and the GEV model’s (black line) NDD\textsubscript{CA,AVG,7D} return level estimates.

Figure 6.6, also reports a 1-in-20 year NDD\textsubscript{CA,AVG,7D} return level of 16.25°C·day. Using this return level, peak day gas consumption for the next gas year is estimated as 83.14 GWh. This estimate is approximately 13% lower than the 95.7 GWh estimated to the 1-in-50 year supply standard, in Section 6.3.2.1.

### 6.4 Conclusion

The NDD\textsubscript{CA} variable derived in Chapter 4 is shown to be an accurate estimator of NDM market gas consumption, accounting for numerous weather effects and gas consumption dynamics, in the form of a composite weather variable.

Parameters for the variable have been estimated using a gas consumption series of three years duration when wide ranging weather patterns may have occurred. Regardless of
this, an $R^2$ value of almost 0.99 has been reported, for the variable in a multiple linear regression model.

To estimate these parameters a non-linear least squares model was developed that included differential intercepts and slopes to account for gas market conditions of previous gas years. Using this methodology each parameter in the NDD$_{CA}$ variable has also been quantified, and found to be significant. A comparative analysis of these parameters found that accounting for the effect of solar radiation, and building thermal memory using an effective outdoor temperature, both contributed to the most significant improvements in model accuracy.

Although the addition of solar radiation contributed to a significant reduction in the number of years of climate data in which to develop an extreme value model, it was shown using goodness-of-fit tests that GEV models are an appropriate representation of the block maxima NDD$_{CA}$ series.

It was also shown that the year-ahead estimate of peak day gas consumption based on a 1-in-20 year peak-week supply standard is approximately 13% lower than that based on a 1-in-50 year peak day supply standard. However, it was also found that the fit of the GEV model used to assess the 1-in-20 year peak-week supply standard is an improvement on the model used to assess a 1-in-50 year peak-day supply standard.

The methodology presented in this chapter can be used by network operators to inform plans to safeguard against diminished supply capacity during extreme cold weather conditions.
CHAPTER 7

INDIVIDUAL SME CONSUMER MODELS
This chapter develops models of daily gas consumption for individual consumers in the NDM market. These are based on availability of daily gas consumption data from smart meters and are an improvement on a current industry model limited by monthly meter reads. The resulting models are for the consideration of network operators currently in the process of installing smart metering infrastructure and who now must provide forecasts to energy suppliers of their NDM portfolio’s daily gas consumption. Although the new models are developed using daily gas consumption data for the SME smart metering sample, they can be applied to both domestic and SME consumers in NDM gas markets. SME rather than domestic daily gas consumption was used to develop the models because it is relatively more difficult to forecast given different industries’ diverse gas requirements and significant variation in this consumption for different days of the week.

The new models are based on either the HDD\textsubscript{WDA} parameter given by Equation 4.65 or the AWDD market consumption estimator described in Section 3.1.2.2. Because the HDD\textsubscript{WDA} parameter must be estimated by NLS, the new models based on this parameter are referred to as non-linear least squares models. And because the AWDD market consumption estimator allows simpler OLS methods, the new models based on this parameter are referred to as ordinary least squares models. This AWDD market consumption estimator is also used to replicate the model currently applied to SMEs in Ireland given by Equation 3.10. This model is used to assess the improvement in accuracy given by the new models and is referred to as the ‘Industry Model’. A summary of this Industry Model and each of the new models is given in Figure 7.1.
Figure 7.1: Summaries of the Industry, OLS, OLS<sub WD</sub>, NLS and NLS<sub WD</sub> Models, where arrows are used to illustrate which models will be compared to one another in terms of modelling accuracy.

Each of the models in Figure 7.1 can be used to forecast the daily gas consumption of consumers in the NDM market. However, they differ in a number of ways in order to assess alternatives in the individual consumer modelling process. The Industry Model is the only model that is fitted using monthly data and because of this it can only differentiate between weekdays and weekends (including holidays) for its daily estimates, using the same day of week adjustment factors applied to SMEs in Equation 3.10 irrespective of their applicability to the individual enterprises.

The ordinary and non-linear least squares models are fitted using daily gas consumption data and as a result they both have daily coefficients to distinguish between each day of
the week and holidays for their daily estimates. The ‘OLS Model’ is used to assess the benefit of daily modelling coefficients rather than only weekday and weekend adjustment factors as in the case of the Industry Model. The ‘NLS Model’ is also used to assess if there is any benefit in estimating individualised HDD\textsubscript{WDA} parameters for each consumer in the sample; this contrasts with the OLS Model and Industry Model which apply the same AWDD market consumption estimator to each enterprise, thus assuming that the weather-dependent daily gas consumption for each consumer follows that of the overall market. Such an assumption is no longer necessary with smart-metering data.

The ‘OLS\textsubscript{WD} Model’ and ‘NLS\textsubscript{WD} Model’ are estimated using the same daily gas consumption data as the OLS Model and NLS Model but improve upon these models by addressing first-order autocorrelation in their residual error series. This refers to the correlation between consecutive residuals or differences in the modelled series and fitted values. If there is a strong correlation between consecutive residuals in the daily gas consumption models, the residual error or consumption for each previous day can be used as an explanatory variable to improve the accuracy of each estimate. However for this method to be applied by the gas industry for portfolio forecasting, real-time smart metering data must be available, whereby data is downloaded on a daily basis by the network operator at an additional cost compared to less frequent monthly downloads, for example.

The OLS\textsubscript{WD} Model and NLS\textsubscript{WD} Model assume the availability of such real-time smart metering data and are used to assess the improvement in modelling accuracy achievable by addressing residual autocorrelation. These models can be used for within-day estimates of gas consumption for a given day when the previous day’s consumption
value becomes available, hence the WD subscript in their names. The OLS Model and NLS Model can be used for both next- and within-day estimates but do not benefit from real-time smart metering data. The additional value which might arise from any improvement in the accuracy of the $\text{OLS}_{WD}$ Model and $\text{NLS}_{WD}$ Model could be used as a basis to justify the additional telecommunication costs necessitated by real-time smart-metering data.

Figure 7.1 also illustrates which models are to be compared to one another in terms of modelling accuracy. Each of the new models are compared to the Industry Model to demonstrate the improvement in modelling accuracy possible with daily rather than monthly data. The NLS Model is compared to the OLS Model to investigate if an improvement in accuracy is possible with individualised HDD$_{WDA}$ parameters instead of AWDD market consumption estimators. And the within-day models are compared to their initial models to investigate if the availability of real-time smart-metering data to address residual autocorrelation is justified.

These comparisons are made both for the individual models and then for their aggregated estimates. The assessment of the individual models is used to investigate the significance of each modelling parameter. The assessment of the aggregated estimates is used to demonstrate the accuracy of the models using the SME sample as a portfolio of consumers, in order to simulate forecasts for a hypothetical energy supplier.

### 7.1 Methodology

This section begins with a description of the Industry Model used to represent the current model applied to monthly-metered SMEs in Ireland. The ordinary least squares
models are developed next. The OLS Model addresses the Industry Model’s inability to use daily smart metering data, and the OLS\textsubscript{WD} Model assumes real-time smart-metering data is available to address residual autocorrelation. The non-linear least squares models are developed last. The NLS Model applies the HDD\textsubscript{WDA} parameter and the NLS\textsubscript{WD} Model also addresses residual autocorrelation. Finally, the model estimation methods are then described along with the applied modelling accuracy metric.

7.1.1 Industry Model

The current model applied to monthly-metered SMEs in Ireland is described in Equation 3.10. This applies a scaling factor to ensure that the sum of individual consumer forecasts equals a separate NDM market forecast. This scaling factor cannot be applied here because the Industry Model is only estimated for a small sample of smart-metered SMEs rather than the entire population of Irish NDM consumers. The scaling factor is omitted and the Industry Model is therefore given by:

\[
\hat{C}_D = \begin{cases} 
(b_0 + b_1 \text{AWDD}) \text{DoW\textsubscript{WKD}}; & \text{on weekdays,} \\
(b_0 + b_1 \text{AWDD}) \text{DoW\textsubscript{WE/Hol}}; & \text{on weekends/holidays.} 
\end{cases} 
\]  

(7.1)

where: \(\hat{C}_D\) is the estimate of consumption for a given day \((D)\); \(b_0\) and \(b_1\) are given by Equation 3.8 using monthly meter readings – which for the purpose of this study are calculated as monthly gas consumption values using the available smart-metering data; \(\text{DoW\textsubscript{WKD}}\) and \(\text{DoW\textsubscript{WE/Hol}}\) are day of week adjustment factors for weekdays and weekends or (observed) public holidays, which were given as 1.09 and 0.79 for SMEs in Equation 3.10.
The consequence of omitting the scaling factor from the Industry Model is assessed later in Section 7.2.4.

### 7.1.2 Ordinary Least Squares Models

The OLS Model is an extension on the simple principle in Equation 7.1 that SMEs in general have different gas consumption requirements on weekdays and weekends/public holidays. In this new model, it is assumed that each SME can have different gas consumption requirements for each type of day. For example, a restaurant may be busier on holidays and on days leading up to and including the weekends. The OLS Model is given by:

\[
C_D = b_0 + \Delta b_{0,Tue}.DV_{Tue.} + \cdots + \Delta b_{0,Sun}.DV_{Sun.} + \Delta b_{0,Hol}.DV_{Hol.} + \\
\quad b_1AWDD_D + \Delta b_{1,Tue}.DV_{Tue.AWDD_D} + \cdots + \Delta b_{1,Hol}.DV_{Hol.AWDD_D} + \varepsilon_D
\]  

(7.2)

where: \(C_D\) is the consumer’s gas consumption (kWh) for a given day \((D)\); \(b_0\) is the model’s intercept parameter and an estimate of the consumer’s base consumption on a Monday (kWh/day); \(\Delta b_{0S}\) are the differences in base gas consumption for Tuesdays-Sundays and public holidays (kWh/day); \(DV_{s}\) are dummy variables to indicate Tuesdays-Sundays and public holidays; \(b_1\) is the model’s slope parameter and an estimate of the building’s response to AWDDs on a Monday (kWh/°C·day); \(\Delta b_{1S}\) are the differences in this AWDD response for Tuesdays-Sundays and public holidays; and \(\varepsilon_D\) is the model error for a given day.

However, the OLS Model in Equation 7.2 does not account for the potential for real-time smart metering data, whereby a consumer’s gas consumption series is available up to the previous day on which their next- or within-day forecast is made. This availability
of recent consumption data allows any autocorrelation in the residual error series to be accounted for in the model. This is addressed by the following OLSwd Model:

\[
C_D = b_0 + \Delta b_{0,Tue}.DV_{Tue.} + \cdots + \Delta b_{0,Sun}.DV_{Sun.} + \Delta b_{0,Hol}DV_{Hol.} + b_1AWDD_D + \Delta b_{1,Tue}.DV_{Tue.}AWDD_D + \cdots + \Delta b_{1,Hol}DV_{Hol}AWDD_D + b_2C_{D-1} + \varepsilon_D
\]  

(7.3)

where: \(C_{D-1}\) is the consumer’s gas consumption (kWh) for the previous day (D-1) and is a lagged dependent variable in the model, or a simple ordinary least squares method to account for autocorrelation in the residual error series; and \(b_2\) is the coefficient for this lagged dependent variable.

Although these ordinary least squares models address the limitation of only weekday and weekend/holiday adjustment factors in Industry Model, they still assume that each consumer’s response to weather is given by the AWDD estimator of NDM market gas consumption; therefore these models do not account for each building’s unique response to various weather effects. This AWDD parameter is also greater than zero for each day in the modelled period; consequently, neither the Industry nor the ordinary least squares models can estimate zero weather-dependent consumption for a consumer in the summertime when their heating system is not operated. The OLS Model and OLSwd Model apply different \(b_1\) parameters for each day type. However, in degree day models this coefficient should be fixed as it is related to the unvarying thermal properties of the building. These limitations are addressed by the following non-linear least squares models.
7.1.3 Non-linear Least Squares Models

The NLS Model assumes that each SME can have different gas consumption requirements for each type of day and is given by:

\[
C_D = b_{0,\text{Mon}}DV_{\text{Mon}} + b_{0,\text{Tue}}DV_{\text{Tue}} + \cdots + b_{0,\text{Sun}}DV_{\text{Sun}} + b_{0,\text{Hol}}DV_{\text{Hol}} \\
+ b_1HDD_{WDA,D}(T_{\text{B,Mon}}, T_{\text{B,Tue}}, \ldots, T_{\text{B,Sun}}, T_{\text{B,Hol}}, \gamma_1, \gamma_2, \alpha_1) + \epsilon_D
\]  

(7.4)

where: \(b_0\) terms are base consumption values for each day-type, Monday-Sunday and public holidays (kWh/day), rather than \(\Delta b_0\) parameters as applied in the OLS models – this is to permit logical zero value lower limits for each \(b_0\) parameter or base consumption value in the NLS model estimation algorithm, as shown later in Table 7.1; \(T_B\) terms are base temperature values for each day-type in the HDD\(_{WDA}\) parameter; and \(\gamma_1, \gamma_2\) and \(\alpha_1\) are the solar radiation, wind-speed and thermal memory parameters of the HDD\(_{WDA}\) parameter.

However, the NLS Model in Equation 7.4 does not account for the potential for real-time smart metering data and autocorrelation in the residual error series. This is accounted for by the following NLS\(_{WD}\) Model:

\[
C_D = b_{0,\text{Mon}}DV_{\text{Mon}} + b_{0,\text{Tue}}DV_{\text{Tue}} + \cdots + b_{0,\text{Sun}}DV_{\text{Sun}} + b_{0,\text{Hol}}DV_{\text{Hol}} \\
+ b_1HDD_{WDA,D}(T_{\text{B,Mon}}, T_{\text{B,Tue}}, \ldots, T_{\text{B,Sun}}, T_{\text{B,Hol}}, \gamma_1, \gamma_2, \alpha_1) + \epsilon_D
\]  

(7.5.1)

\[
\epsilon_D = \rho \epsilon_{D-1} + \nu_D
\]  

(7.5.2)
where: the additional first-order autoregressive error or AR(1) model is used to account for autocorrelation in the residual error series; $\rho$ is its autoregressive parameter; and $v_D$ are random errors.

Such first-order autoregressive error models are estimated using iterative methods and are preferred in this model to the previous lagged dependent variable method in Equation 7.3. This is because, unlike lagged dependent variables, they do not affect coefficient interpretation and because the NLS Model in Equation 7.4 must also be solved using iterative methods. The NLS Model and NLS\(_{WD}\) Model are solved using the non-linear least squares method described in the following section.

### 7.1.4 Model Estimation

Each of these models have been estimated using R [64]. Coefficients for the Industry Model, OLS Model and OLS\(_{WD}\) Model have been estimated using this software’s linear model (lm) package. The NLS Model and NLS\(_{WD}\) Model are estimated using R’s Levenberg-Marquardt non-linear least-squares (nlsLM) algorithm.

To help this algorithm to converge, starting values and limits have been stipulated for each parameter in the non-linear least squares models, as shown in Table 7.1. These starting values force the nlsLM algorithm to begin with a simpler HDD model before iterating to an optimal solution.
### Table 7.1: Parameter starting values and limits for the non-linear least squares models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Starting Value</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$ terms</td>
<td>$b_0$ terms are given by a simpler HDD model with $T_B$, $\gamma_1$, $\gamma_2$, $\alpha_1$ fixed to the starting values below.</td>
<td>0 kWh</td>
<td>None</td>
</tr>
<tr>
<td>$b_1$</td>
<td>$b_1$ is given by a simpler HDD model with $T_B$, $\gamma_1$, $\gamma_2$, $\alpha_1$ fixed to the starting values below.</td>
<td>0 kWh/°C·day</td>
<td>None</td>
</tr>
<tr>
<td>$T_B$ terms</td>
<td>15.5°C (a)</td>
<td>0°C (b)</td>
<td>30°C (c)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0 °C/°J/cm²</td>
<td>0 °C/°J/cm²</td>
<td>0.005 °C/°J/cm² (d)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0 °C·day/knot</td>
<td>0 °C·day/knot</td>
<td>0.01 °C·day/knot (e)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0</td>
<td>0</td>
<td>0.7 (f)</td>
</tr>
<tr>
<td>$\rho$ (g)</td>
<td>$\rho$ is given by a simpler HDD model with $T_B$, $\gamma_1$, $\gamma_2$, $\alpha_1$ fixed to the starting values above.</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes:

a) The base temperature commonly assumed in the UK (and Ireland) [55].
b) To model the day(s) on which heating is not required and to facilitate standard parameter significance tests.
c) To model the day(s) on which heating is always required.
d) Limits the temperature effect of solar-radiation to 15°C, see Equation 4.54.
e) Limits the equivalent HDD effect due to wind-speed to a 30% increase in HDDs, see Equation 4.54.
f) Limits the ‘thermal memory’ effect to approximately a week, see Equation 4.56.
g) Relates to the NLS_{WD} Model only.

#### 7.1.5 Model Accuracy

In this study, the accuracy of each individual SME model is quantified by the following mean normalised absolute percentage error (MnAPE) metric:

$$MnAPE = \frac{100}{n} \sum_{D=1}^{n} \frac{|C_D - \hat{C}_D|}{0.5(C_{D,MAX} + C_{D,MIN})}$$

(7.6)
where: \( n \) is the number of days in the modelled series; and \( C_{D,\text{MAX}} \) and \( C_{D,\text{MIN}} \) are the maximum and minimum daily gas consumption values in the modelled series.

This metric is based on a mid-range consumption denominator as an alternative to the daily gas consumption denominator applied in the standard mean absolute percentage error (MAPE) metric. This alternative denominator is required because numerous zero daily consumption values were found in the individual SME consumption data. It is used instead of the maximum capacity denominator applied in the modified-MAPE reported in the NDM market forecasting literature (see Section 3.2.3.1), as some of the individual SMEs had large maximum consumption values which would have resulted in unrepresentative small percentage errors, and vice-versa. It is calculated using only maximum and minimum consumption values, because some of the individual SMEs had many zeroes or low consumption values which would have resulted in low average consumption denominators and unrepresentative high percentage errors. This mid-range consumption denominator was similarly applied to an alternative coefficient of variation metric in a previous building energy modelling study [71].

The accuracy of the aggregated (or portfolio) estimates for the alternative models is quantified in Section 7.2.4 using both the MnAPE and nAPE formulae in Equation 7.6. nAPEs are used to quantify the accuracy of the daily estimates given by the alternative models and to highlight dates or times of the year when aggregated estimates are least accurate. MnAPEs are used to compare the overall accuracy of aggregated estimates for each of the alternative models.
7.1.6 Data

Each of the alternative models is estimated using daily gas consumption data for each member of the smart-metered SME sample (Section 2.1.2), recorded during the gas year: 1st Oct.’10 - 30th Sept.’11. The ordinary least squares models are estimated using AWDD data provided by GNI, and the non-linear least squares models are estimated using outdoor temperature, global radiation and wind-speed data for the most representative single location, Dublin Airport, as no climate data were recorded from the vicinity of the buildings and because the SMEs were no further than approx. 300km from this location. One consumption outlier was identified and removed for a single SME before model estimation.

7.2 Results and Discussion

The results of this study begin with observations from the applied model estimation methods, and an analysis of coefficient estimates for the OLS Model and NLSWD Model. This is followed by an assessment of first-order autocorrelation for the residual error series for each type of ordinary and non-linear least squares model. The accuracy of these models and Industry Model is first illustrated for a single SME, and then for the SME sample, initially for the individual models and then for their aggregated (or portfolio) estimates.

7.2.1 Model Estimation

It was found that the NLS Model and NLSWD Model failed to converge for seven SMEs in the sample, including three consumers common to both models. Of the eleven non-converging consumers, eight had largely weather-independent gas consumption, and
three had largely uniform gas requirements during the heating season. These non-converging consumers represent approximately 20% of the SME sample and were removed from the remainder of this study, as trials of alternative starting values to those in Table 7.1 proved to be impractical for these non-converging consumers. This was not the only difficulty observed for this non-linear least squares modelling method, as it was also found such models require significantly more processing time than the ordinary least squares models to estimate model coefficients. For example, 23 minutes was required to estimate the successfully converged NLSWD Models and 0.15 seconds were required to estimate the corresponding OLSWD Models, using a computer with a 2.2 GHz dual core processor and 4GB of RAM. Such orders of magnitude in time difference may be important when models must be estimated for potentially millions of consumers.

7.2.2 Model Coefficients

Boxplots of coefficient p-values for the individual OLS Models and NLSWD Models are shown in Figures 7.2 and 7.3, respectively. These results are presented for the NLSWD Models rather than the NLS Models, as later in Figure 7.4 it is seen that the residual series for the NLS Models present significant first-order autocorrelation, and therefore underestimate the coefficient standard errors required to calculate these p-values [72]. The p-values are presented for the OLS Models rather than the OLSWD Models, since the coefficients from the OLS Models have a simpler interpretation and because the effect of residual autocorrelation on coefficient standard errors could be addressed using standard methods [73] available with R.
In Figure 7.2, it is seen that the $\Delta b_0$ differences (to Monday’s base gas consumption) and the $\Delta b_1$ differences (to Monday’s AWDD response) are generally significant to 5% for Saturday and 10% for Sunday; thus indicating different base gas consumption requirements at the weekend for the SME sample. Similarly, Saturday and Sunday’s $\Delta b_1$ coefficients are generally significant to 5%; thus indicating different heating requirements at the weekend for the SME sample.

**Figure 7.2**: Boxplots of coefficient p-values for the OLS Models – calculated using standard errors that have been corrected for autocorrelation, and vertical lines below which the estimated coefficients are significant at 10% and 5% probability levels, respectively.
In Figure 7.3, it is seen that the $b_0$ and $b_1$ coefficients and the $T_B$, $\gamma_1$ (solar radiation) and $\alpha_1$ (thermal memory) parameters in most cases are significant to 5% and therefore greater than zero, thus indicating the significance of these effects in the alternative NLS models. Whereas the $\gamma_2$ (wind speed) parameter is only significant to 10% for less than half of the sample.

**Figure 7.3:** Boxplots of coefficient p-values for the NLS\textsubscript{WD} Models – calculated using a one-tailed test that the coefficient is greater than zero [74], and vertical lines below which the estimated coefficients are significant at 10% and 5% probability levels, respectively.
7.2.3 Residual Autocorrelation

Boxplots of the first-order autocorrelation coefficient for the residual error series from the individual OLS Models and NLS Models and their alternative within-day versions are shown in Figure 7.4. It is seen that this autocorrelation coefficient is significant (or outside the ±0.05 critical region) for all of the OLS Models and NLS Models. It is also seen that many of the residual error series for the $\text{OLS}_{\text{WD}}$ Models and $\text{NLS}_{\text{WD}}$ Models have insignificant first-order autocorrelation coefficients and therefore have been reduced to approximately ‘white noise’. This is because these models are based on real-time smart-metering data which permits the use of a lagged dependent variable in the case of the $\text{OLS}_{\text{WD}}$ Models, or a first-order autocorrelation error model in the case of the $\text{NLS}_{\text{WD}}$ Models.

![Boxplots](image)

**Figure 7.4:** Boxplots (and means, in descending order) of $r_1$, the first-order autocorrelation coefficient for the residual error series from the individual ordinary and non-linear least squares models, and a critical region between ±0.05 (given by ±√ no. of residuals [72]) where $r_1$ values are insignificant.
7.2.4 Model Accuracy

In Figure 7.5, the relative accuracy of the alternative models is illustrated for a SME chosen from the sample to represent a consumer with regular daily and weather-dependent gas consumption, so that the benefit of the new models can be easily demonstrated. These accuracy comparisons are made using correlation plots of actual and estimated gas consumption and the MnAPE accuracy metric. This SME has a lower gas requirement at the weekend. This lower weekend requirement is the main source of error for the Industry Model, as illustrated by its correlation plot, where it is seen that many low consumption values have been significantly overestimated. These days have been addressed by the NLS Model and OLS Model and this has resulted in improved MnAPE values in their correlation plots, compared to the Industry Model. It is also seen that although the OLSWD Model is the most accurate, it is only marginally more accurate than the OLS Model. It is also seen that some models produce illogical negative estimates. However, these are permitted in GNI’s FAR procedures as only negative portfolio estimates are corrected to zero [21].

In Figure 7.6, the accuracy of estimates from the individual models is compared for the alternative model types using MnAPE boxplots. The NLS Model and OLS Model improve upon the Industry Model while the NLSWD Model and OLSWD Model provide further improvements in modelling accuracy. The means of these MnAPEs (shown in parenthesis above each boxplot) indicate that the ordinary least squares models are more accurate than their non-linear least squares alternatives. However, as the notches overlap for the OLS and NLS boxplots and for the OLSWD and NLSWD boxplots the difference in the corresponding medians are not statistically significant.
Figure 7.5: Daily gas consumption for a sample SME between Oct.’10 and Sept.’11 (top left); and in order of accuracy, correlation plots for this consumption series (y-axis) based on the corresponding estimates given by each of the alternative models (x-axes).
Figure 7.6: In order of accuracy, boxplots of MnAPE values for daily gas consumption estimates, given by the alternative models for each consumer in the SME sample, with the mean values of each boxplot in parenthesis.
In Figure 7.7, the accuracy of the aggregated (or portfolio) estimates is compared for the alternative model types using boxplots of nAPEs for each daily estimate in the modelled year. Means of these nAPEs (or MnAPEs) are also reported in parenthesis above each boxplot, and these illustrate that the ordinary least squares models are more accurate than the alternative non-linear least squares models. It is seen that the accuracy of the Industry Model is comparable to the NLS Model, even though it accounts for fewer day-type effects. It is also seen that the NLSWD Model is an improvement on the Industry Model, and that real-time smart-metering data is also important for this model type as it is a significant improvement on its corresponding NLS Model. However, this is not the case for the ordinary least squares models, as it is seen that the simpler OLS Model is more accurate than the OLSWD Model which relies on real-time smart metering data. Although it was seen for the individual SME models in Figure 7.6 that the OLS Model was less accurate than the NLSWD Model and OLSWD Model, it is the most accurate here.

In Figure 7.7 it is also seen that the largest nAPEs for each model occurred mostly during the Christmas holiday period (24th Dec.’10 – 3rd Jan.’11). In this study, only the official public holidays were applied as holiday day-types in the individual models; for example, the public holidays for Christmas Day (Sat. 25th Dec.’10), St. Stephen’s Day (Sun. 26th Dec.’10) and New Year’s Day (Sat. 1st Jan.’10) were observed on the 27th and 28th Dec. and the 3rd Jan., respectively. Attempts made to model the Christmas period either with or separately to the applied holiday parameters resulted in increased convergence problems in the non-linear least squares models or insignificant holiday parameters in the ordinary least squares models. However, this is an example of why the scaling factor omitted from the Industry Model (see Section 7.1.1) is important. This factor can address large daily errors such as those highlighted in Figure 7.7, as it used to
ensure that aggregated individual consumer forecasts equal a separate NDM market forecast. These knowledge-based market forecasts can address Christmas holidays more easily than individual consumer models presented with irregular gas consumption during such periods.

![Boxplot Figure 7.7](image)

**Figure 7.7**: In order of accuracy, boxplots of nAPE values for daily aggregated (or portfolio) estimates given by the alternative model types, with MnAPE values in parenthesis.

### 7.3 Conclusion

This chapter presented alternative ordinary and non-linear least square methods to model daily SME gas consumption. A benchmark Industry Model was described which
applies monthly meter readings and an AWDD estimator of NDM market consumption to individual enterprises. The OLS Model, improved on the Industry Model by estimating coefficients for each type of day rather than just for weekdays and weekends/holidays. The NLS Model was developed later and was used to assess the practicality of applying HDD_{WDA} estimators of building heat consumption unique to each enterprise. Finally, the OLS_{WD} Model and NLS_{WD} Model were developed to assess the benefit of using real-time smart-metering data.

Each of these alternative models was then estimated for the SME sample. The resulting MnAPEs were used to quantify the improvement in daily gas consumption estimates for a single SME based on the new models and smart-metering data, compared to the Industry Model limited by monthly meter readings. nAPEs were used to quantify the accuracy of aggregated gas consumption estimates for the SME sample portfolio. It was shown in Figures 7.6 and 7.7 that the percentage errors for the most accurate individual consumer models and aggregated portfolio estimates were approximately half that of the Industry Model.

It was also found that at an individual SME level the ordinary least squares models are only marginally more accurate than their corresponding non-linear least squares models, and that a significant improvement in accuracy is possible for both modelling methods using real-time smart-metering data. However, it was shown in Figure 7.7 that at an aggregated portfolio level the availability of real-time smart-metering data only improves the accuracy of the non-linear least squares models, and that the ordinary least squares models were again the most accurate modelling method.
Although it was found that the ordinary least squares models are more accurate than their non-linear least squares alternatives, this is based on in-sample model estimates only and an AWDD parameter that exactly estimates past NDM market consumption. However in practice, the accuracy of the ordinary least squares models for out-of-sample forecasts is dependent on the accuracy of AWDD forecasts and in turn NDM market consumption forecasts (see Equation 3.7). The non-linear least squares models do not rely on such NDM market consumption forecasts because they apply individualised HDD_{WDA} parameters. However, it was shown that these non-linear least squares models are less accurate, rely on real-time smart-metering data, suffer from non-convergence and have processing times orders of magnitude greater than the ordinary least squares models. The OLS Model in Equation 7.2 was found to be the most accurate at an aggregated portfolio level and therefore may be the most suited to portfolio forecasting in the future when smart-metering data are widely available.
CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS
8 CONCLUSIONS, RECOMMENDATIONS AND AFTERWORD

8.1 Introduction

This research originated as a result of GNI’s ongoing commitment to the continuous improvement of its gas management processes and developments at an EU level that aim to improve building energy efficiency, security of gas supplies, and energy market integration.

One of these developments is the installation of smart-metering infrastructure across the EU and this is expected to help accelerate improvements in building energy efficiency. In recognition of this and with the support of GNI, smart-metering data from a sample of Irish domestic consumers was used in this research in Chapter 5: Domestic Energy Efficiency Benchmarking, to develop a new benchmarking tool for energy suppliers, which now must assist in promoting energy savings among their consumers as part of the ‘Energy Efficiency Obligations Scheme’ recently introduced in the EU.

New EU regulations require network operators to ensure the security of gas supplies during extreme cold weather. Consequently, a state-of-the art peak day forecasting method was developed in order to assess the difference in the maximum network capacity required by alternative gas supply standards similar to those applied in Ireland and elsewhere in the EU. Daily gas consumption for the Irish NDM market was used for this purpose in Chapter 6: Peak Day Forecasting.

EU regulations also require that network operators provide forecasts to new and existing energy suppliers in order to improve energy market integration across the EU. These forecasts must provide estimates of next- and within-day gas consumption for each
energy supplier’s portfolio of NDM consumers. In this regard, GNI identified that
smart-metering data can also be used to improve the accuracy of their current forecast
method. Sample smart-metering data from Irish SME consumers was used to develop
new individual consumer models that can be used for this purpose in Chapter 7:
Individual SME Consumer Models

8.2 Conclusions

This research developed upon HDD theory in order to meet the aims and objectives set
out in Section 1.3. These were fulfilled through the development of a gas end-use
benchmarking tool in Chapter 5; the development of the new state-of-the-art climate
adjusted network degree day (NDDCA) parameter in Chapter 4, so that an improved
peak-day NDM market gas consumption forecasting method could be developed in
Chapter 6; and finally the development of the similar weather and day-type adjusted
HDD (HDDWDA) parameter to assess the practicality of non-linear least squares models
compared to ordinary least squares models, so that an improved individual consumer
forecasting method could be developed in Chapter 7.

HDD parameters and NLS methods were first applied in this research to estimate
individual building energy models for the domestic smart-metering sample in Chapter 5.
Such HDD models were estimated using smart-metering data and an NLS algorithm for
the first time. The resulting intercept \(b_0\), slope \(b_1\) and base temperature \(T_B\)
parameter distributions were presented and it was found that the average base
temperature of Irish dwellings is over 1°C less than the 15.5°C value commonly
assumed in Ireland and the UK. These parameter distributions were also used with
household survey data to develop multinomial logistic regression (MLR) models that
can be used to compare the inferred gas end-uses of dwellings with similar household characteristics. These MLR models were presented as an alternative to current approaches such as energy intensity metrics based on building floor area; and by way of example, were used to compare the energy efficiency of similar buildings based on their intercept, slope and base temperature parameter estimates.

Similar NLS methods were then used to estimate the coefficients of the new $\text{NDD}_{\text{CA}}$ parameter in Chapter 6. This parameter was shown to be an accurate estimator of NDM market gas consumption and accounted for the effect of solar radiation within the HDD (indoor-outdoor) temperature differential for the first time. Alternative $\text{NDD}_{\text{CA}}$ values were used to estimate peak-day gas consumption for the year-ahead to 1-in-50 year peak day and 1-in-20 year peak week supply standards similar to the current Irish and EU supply standards, respectively. It was found that a 1-in-50 year peak day standard requires 13% additional maximum network capacity than the less demanding 1-in-20 year peak week standard. It was also shown that the main difference between the $\text{NDD}_{\text{CA}}$ parameter and previous state-of-the-art parameters is that solar radiation effects are not accounted for by the latter. Therefore, the increase in accuracy due to solar radiation was quantified. It was found that solar radiation accounts for a significant increase in accuracy, as $R^2$ values for a NDM market gas consumption model increased from 0.8452 to 0.9372 when the model’s HDD parameter was adjusted to account for solar radiation (see Table 6.4). Solar radiation should therefore be considered by the gas industry both for peak-day and daily gas consumption forecasting.

The NLS method was finally used to investigate the practicality of estimating weather and day-type adjusted HDDs ($\text{HDD}_{\text{WDAS}}$) for individual SMEs in Chapter 7. This was assessed by comparing the resulting individual non-linear least squares consumer
models to simpler ordinary least squares models. It was found that although non-linear least squares based models can deliver individual consumer models with comparable accuracy to ordinary least squares based models, the practicality of the applied non-linear least squares models is questionable as they rely on real-time smart-metering data, suffer from non-convergence and have processing times orders of magnitude greater than the ordinary least squares based models. The applied ordinary least squares based models were simpler, slightly more accurate and, importantly, may not be reliant on potentially expensive real-time smart metering data when used for portfolio gas consumption estimates.

8.3 Recommendations

The benchmarking tool developed in this research was limited by MLR models of low, medium and high categories for each of the intercept, slope and base temperature parameter distributions. This limitation was applied for simple classifications of each parameter distribution and to reduce the size of the logistic regression models. It is recommended that the number of categories for each parameter distribution is increased for larger consumer samples. This could allow the energy efficiency of buildings at lower regions of each parameter distribution to be compared.

It is also recommended that the forecast-AWDDs used in the Irish NDM market’s FAR procedures (see Equation 3.7) can indirectly benefit from the new NDD\textsubscript{CA} parameter, if this is applied as part of the top-down NDM market forecasting process used to calculate forecast-AWDDs. Should the non-linear least squares models in Chapter 7 be re-evaluated in the future it is recommended that the additional heat required to pre-heat buildings following a normally unheated day or weekend is addressed in the model if
appropriate. This additional effect could not be successfully modelled for the SME sample due to limited instances of such occurrences.

Additional applications of the HDD parameters and the NLS methods developed in this research are possible. For example, the HDD\textsubscript{WA} parameter developed in Section 4.3.2 could be used to identify buildings with limited solar gains, excessive air-infiltration heat loss and inefficient continuous heating time control settings using its $\gamma_1$, $\gamma_2$ and $a_1$ parameter estimates, respectively. While the NDD\textsubscript{CA} parameter could be applied to estimate rolling intercept and slope parameters for several years of NDM market consumption, in order to infer the impact of housing energy efficiency programs or improvements made to building fabric U-value standards over time.

8.4 Afterword

It is worth highlighting the advantages of alternative building energy modelling approaches to the HDD methods applied in this research. Examples of these include the PRinceton Scorekeeping Method (PRISM) developed in the US [75], and the Standard Assessment Procedure (SAP) [76], used to calculate Building Energy Ratings (BERs) in the UK. Both of these methods, their advantages and some recent applications are summarised in the following sections along with future research opportunities.

8.4.1 Princeton Scorekeeping Method

PRISM is a software tool based on a HDD regression model of energy data from monthly utility bills. It applies an iterative procedure based on Newton’s method to estimate the building’s reference temperature along with ordinary least squares to estimate the model’s base-level and heating-slope parameters [75] – or, for consistency
with this thesis, the building’s base temperature, intercept and slope parameters. Although the method is based on monthly data, standard errors are estimated for each of the model’s parameters, including the non-linear base temperature parameter. Such standard errors are easily computed using daily data and the non-linear least squares methods applied in this research – for example, see Table 5.6. However with monthly data, these errors are only easily computed for the intercept and slope parameters using ordinary least squares methods. While estimates of the standard error for the base temperature parameter were not found in other methods based on monthly data, this standard error is estimated by PRISM software. This is an important advantage of PRISM over other monthly methods which do not provide this estimate, as the resulting standard error can be used to quantify possible variation in a building’s base temperature.

PRISM extends on HDD regression modelling through the use of a Normalised Annual Consumption (NAC) index. This is a weather-adjusted index of consumption that is calculated using intercept and slope parameters and seasonal HDDs based on the building’s base temperature and local seasonal temperatures. It is an estimate of the expected consumption for the modelled building for a typical year. Typically, NACs are estimated for a building before and after an energy saving intervention in order to determine if any (weather-adjusted) savings have been made. Such savings are given by the difference in these before and after NAC indices, and as standard errors are also provided by PRISM for these indices, the significance of any resulting savings can be determined.

It has been highlighted that the NAC index is the most important feature of PRISM [75]. This is because it has been shown that it is less sensitive to variations in base
temperature than the intercept and slope parameters of the HDD regression model; and because, even if the (base-level, heating-slope and reference temperature) parameters of the NAC index have been poorly determined, the standard error of the resulting NAC estimate is usually only 2-4% [75]. It is therefore recommended that future work based on the Benchmarking Tool in Chapter 5 should explore the benefit of the NAC index, its potential for identifying inefficient domestic gas consumers (or dwellings) and its application in estimating (weather-adjusted) energy savings.

Although the issue of base temperature variation was addressed in this research, by accounting for the effect of solar gains within the HDDWA, HDDWDA and NDDCA variables of Chapter 4, it was assumed that the mean indoor temperature and internal heat gain components of base temperature are relatively constant in comparison. However, it has since been found that the mean indoor temperature of buildings decreases at a rate of approximately 0.25°C for every 1°C reduction in daily mean outdoor temperatures (during the heating season) [77]. It is therefore recommended that future work based on this research should explore the effect of such indoor ‘temperature droop’.

For example, differential base temperatures for each month of the heating season (e.g. ΔTB,DEC) could be trialled in the HDD and NDDCA variables of Chapter 5 and 6, respectively, using a dummy variable approach similar to that applied in Equation 6.1. If proven effective, this could result in more accurate estimates of the coefficients of these variables. This is particularly important with respect to the slope coefficient of the HDD variable, should it be related to the modelled building’s heat loss coefficient (see Equation 4.11), and subsequently used to estimate the benefit of improvements in building fabric insulation or heating system efficiency.


8.4.2 Standard Assessment Procedure

SAP is used to calculate the BERs used to quantify the energy performance and CO$_2$ emissions of planned dwellings in the UK. It is based on building design drawings and specifications, rather than fuel consumption data in the case of HDD regression models. It is used to estimate the expected annual energy consumption of a dwelling based on its heat loss coefficient and seasonal temperatures for the region. SAP is similar to the Dwellings Energy Assessment Procedure (DEAP) used in Ireland to calculate BERs. Both of these BER modelling methods are based on EN ISO 13790 [78].

The main difference between these BER approaches and HDD regression modelling is that they explicitly calculate mean indoor temperatures and utilised internal and solar heat gains for a building for each month of the year. Whereas in HDD regression modelling, internal and solar heat gains are modelled as equivalent temperature adjustments to a building’s mean indoor temperature, in order to define its base temperature (see Section 4.3.1). However, as a building’s base temperature is assumed to be a constant value parameter in HDD modelling, each of its heat gain and mean indoor temperature components are also assumed to be constant for the modelled period, even though they may vary across the year.

This issue of base temperature variation was discussed in the previous PRISM section. It was proposed that indoor ‘temperature droop’ could be addressed by differential base temperatures for each month of the heating season, and it was highlighted that variable solar gains have already been addressed in this research within the HDD$_{WA}$, HDD$_{WDA}$ and NDD$_{CA}$ variables. However, additional future research opportunities in this regard, may be identified by summarising the methods used in SAP to calculate internal and
solar heat gains and mean indoor temperature – or the main components of base temperature in HDD regression modelling.

SAP estimates the internal heat gain from occupants, lighting, appliances, water heating, pumps and fans for each month of the year using engineering formulae. For example, lighting gains are estimated based on the building’s floor area, number of occupants, proportion of low-energy lighting outlets, and a sinusoidal (cosine) function that defines the season or time-of-year [76]. Given that the number of occupants in a building contributes both to metabolic and lighting heat gains it is recommended that future research investigates if this occupancy factor can be used to improve the MLR model for base temperature in Table 5.5.

Solar heat gains are estimated in SAP for each month of the year based on the area, orientation, transmittance and framing material of each window, and seasonal solar radiation [76]. The extent, in which these solar (and internal) heat gains are utilised within the building to offset heating system fuel consumption, is estimated in SAP by a gain utilisation factor. This factor is given by a function of the ratio of heat gains to heat losses and a parameter that depends on the time constant (h) of the modelled building [76] – this time constant is used to quantify a building’s thermal inertia and is given by the ratio of its internal heat capacity to its heat loss coefficient. Gain utilisation factors are generally better for buildings with heavy thermal mass or long time constants. This is because the internal heat capacity of such buildings allows heat gains to be utilised more effectively, by absorbing more heat gains that can be released when needed to offset fuel consumption [55].
It is recommended that future research explores the relationship between the gain utilisation factor, time constant and solar heat gain formulae applied in SAP and the \( (\gamma_1) \) solar gain parameter applied in the HDD\(_{WA} \) variable. Once this relationship is established, the \( (\gamma_1) \) solar gain parameter could be used to identify dwellings with limited solar gains, as previously recommended in Section 8.3. Such research could be used to extend the Benchmarking Tool in Chapter 5 so that dwellings suited towards a glazing upgrade may be identified.

SAP estimates the mean indoor temperature for a dwelling for each month of year based on the proportion and temperature of its living space. The mean temperature of the living space is estimated by adjustments to its set-point temperature that account for the reduction in temperature when the heating system is off. These adjustments are based on the time constant and heat loss coefficient of the modelled building, seasonal temperatures, utilised heat gains and heating schedules for weekdays and weekends [76]. The mean temperature for the rest of the dwelling is calculated in a similar manner but is based on lower comfort temperatures. The mean indoor temperature for the dwelling is simply the weighted average of the mean temperature estimates for the living space and the rest of the building; plus an adjustment to account for the effectiveness of the heating control system [76]. Given that the proportion of living space contributes to a building’s mean indoor temperature, it is recommended that future research investigates if this factor, or a suitable proxy variable such as number of bedrooms, can also be used to improve the MLR model for base temperature in Table 5.5.

One of the main outputs of SAP is to estimate the space heating requirement of the modelled dwelling for a normal year and for each month of the heating season. These
space heating requirements are given by monthly heat loss estimates (based on the building’s heat loss coefficient, seasonal temperatures and monthly mean indoor temperature estimates), less corresponding utilised heat gain estimates (based on the building’s gain utilisation factor, and monthly internal and solar heat gain estimates). However, it has been recently found that a model based on SAP tends to overestimate the annual gas consumption (or space heating requirement) of older dwellings in the UK housing stock [79].

This was found by comparing estimates of annual gas consumption for three-bedroom dwellings from a model based on SAP to another model based on PRISM methods and monthly smart-metering data. It was found that the annual consumption of dwellings built prior to 1919 tends to be overestimated by the model based on SAP. Consequently, it was suggested that the assumptions made by such models with regard to heating schedules and building thermal performance needs to be re-examined [79]. This finding suggests that engineering models such as SAP (and DEAP) may need to be adjusted to reflect the results from HDD regression models such as PRISM. It is therefore recommended that future research investigates this for DEAP using the Irish domestic smart-metering dataset applied in this thesis.
REFERENCES
REFERENCES


Joint Office of Gas Transporters, Uniform Network Code – Transportation
Principal Document Section H – Demand Estimation and Demand Forecasting,


Project ELVIRA, Natural Gas Consumption Forecasting, Available:


LIST OF PUBLICATIONS
List of Publications

Chapter 5:

Ronan Oliver, Aidan Duffy, Ian Kilgallon; “Statistical models to infer gas end-use efficiency in individual dwellings using smart metered data”; Sustainable Cities and Society, Volume 23, May 2016, Pages 1-10, ISSN 2210-6707, http://dx.doi.org/10.1016/j.scs.2016.01.009
LIST OF PRESENTATIONS
LIST OF PRESENTATIONS


(2) “Residential Gas Consumption Forecasting”, Poster Presentation, DIT’s Annual Research Symposium, November 2012.


Figures A.1 to A.9 present the in-sample modelling accuracy of weekday NDM market gas consumption models based on the incrementally adjusted HDD variables in Table 6.4.
Figure A.1: Linear relationship between gas consumption for the HDD parameter in Table 6.4.
Figure A.2: Linear relationship between gas consumption for the HDD($\gamma_1$ & $\alpha_1=0$, wind-speed and outdoor temperature) parameter in Table 6.4.
Figure A.3: Linear relationship between gas consumption for the HDD($\gamma_2$ & $\alpha_1=0$, solar radiation and outdoor temp.) parameter in Table 6.4.
Figure A.4: Linear relationship between gas consumption for the HDD(\(\gamma_1 & \gamma_2=0\), effective outdoor temperature) parameter in Table 6.4.
Figure A.5: Linear relationship between gas consumption for the HDD($\gamma_1=0$, wind-speed and effective outdoor temp.) parameter in Table 6.4.
Figure A.6: Linear relationship between gas consumption for the HDD ($\gamma_2=0$, solar radiation and effective outdoor temp.) parameter in Table 6.4.
Figure A.7: Linear relationship between gas consumption for the HDD$_{WA}$ parameter in Table 6.4.

$R^2 = 0.9734$

MAPE = 11.3

MAPE($C_{WD} \geq 70.6$) = 4.14

70.6 = upper 5% threshold

$C_{NDM, WD}$ (GWh)

°C.day
Figure A.8: Linear relationship between gas consumption for the NDD_{WA} parameter in Table 6.4.
Figure A.9: Linear relationship between gas consumption for the NDD\textsubscript{CA} parameter in Table 6.4.