Theory Section

**The DARMA(,)-X Model**

**Extending the DAR() Model**

From the stylised facts in Section XXX, we have seen that modelling conditional mean and volatility together is an important first step towards capturing the dynamics of electricity prices. Therefore, a one starting point is the Double AutoRegressive model of order – also called the DAR() – model studied in (XXX) and (XXX), which combines an autoregressive and conditional heteroskedasticity process. The DAR() model is defined as,

, with parameter space and , and standardised errors following some distribution with mean zero and unit variance . As will be demonstrated in section XXX, the DAR() model has the advantageous property of “volatility-induced stationarity”, which allows us to estimate the model in levels and thus retain more information about the time series compared to ARIMA-type models that rely on first-differing to handle unit-roots. From derivations in Appendix XXX, the conditional mean and variance of the DAR() model are given as,

, where denotes the information set at time . One crucial short-coming of the model is that volatility does not enter the conditional mean, which is an important characteristic of electricity prices as argued in Section XXX. To solve this problem, we propose a generalisation of the DAR() model that explicitly accounts for autocorrelation in the standardised residuals, essentially departing from the assumption. Under -order autocorrelation, the error term becomes,

, where is the new standardised residuals left from the autocorrelation process. The assumption of autocorrelation introduces “infinite memory” and makes it extremely difficult to derive a closed form description of whole model for . Therefore, assuming the new model, derived in Appendix XXX, can be stated as,

, with parameter space , , , and new standardised errors following some distribution with mean zero and unit variance . The restriction is necessary as the power series would otherwise not be defined. From derivations in Appendix XXX, the conditional mean and variance are given as,

We see that the volatility is now allowed to enter the conditional mean, which is due to the fact that lagged values of are contained in information set. We dub our model “DARMA(,)” as the generalisation closely resembles a Moving Average (MA) process. As argued in Section XXX, the volatility of electricity prices is strongly influenced by weather conditions and other exogenous variables. We can therefore augment the DARMA(,) model by incorporating the volatility of exogenous variables, making an ARCH()-X process,

, where for each , is a vector of parameters with cardinality corresponding to the number of exogenous variables. Now, also the volatility of weather conditions and other exogenous variables can drive the conditional mean. Conversely, we dub this model “DARMA(,)-X”. For brevity, we will from now on refer to as the standardised residuals for all the models above, as are assumed to follow the same distribution.

**Stationarity**

When the term “stationarity” is used, it is often loosely defined, but generally refers to “weak stationarity”, which assumes that both the mean and variance of is defined. However, weak stationarity is quite restrictive on the parameter space as it implies that at least , which we will show is not necessary to be able to forecast the conditional mean , is satisfied. In fact, under much looser restrictions termed “strict stationarity”, the model can be both estimated and simulated. In this section, we will provide the necessary moment conditions and outline the different stationarity requirements for models (XXX) and (XXX). Relying on results for (G)ARCH-X type models from (XXX) and (XXX), the moment conditions for model (XXX) are the same with or without the extension (XXX). This holds as long as the exogenous variables are not unit-root processes, i.e. with , but are however allowed to have long-memory as .

The results in this section hold for any distribution of the standardised residuals satisfying . Therefore, to accurately model the heavy tails and spikes of electricity prices shown in section XXX, we propose to use the scaled student-t distribution,

, with Probability Density Function (PDF) given as,

, where denotes the degrees of freedom for the student-t distribution and determines the thickness of the tails. It is worth noting that when the scaled student-t distribution converges to a standard Gaussian. Conversely, when the tails thicken.

To analyse the different stationarity conditions for general lag structure , we will use results from (FinMetrics), (Ling 2007) and (Tweedie) on the Stochastic Recurrence Equation (SRE). Let , then model (XXX) and (XXX) have the following SRE representation,

, where is a random matrix and is a random vector, and (,) is an i.i.d. process such that (,) and are independent for all . Following derivations in Appendix (XXX), and are defined as,

Model (DAR(p)):

Model (DARMA(p,1)):

, where is the -order identity matrix and is a matrix of zeros. We have that (,) and are independent in both SREs, since the only stochastic elements in (,) are the , which are by definition i.i.d.

Let denote spectral radius of matrix . From Theorem 4 in (Tweedie 1988) we have that for any positive integer then if the following conditions holds:

, where “” denotes the Kronecker product in the Kronecker power defined for . If then condition holds for all . Instead, assuming only that for , then is only guaranteed for . To analyse the condition for existence of fractional moments (strict stationarity), let denote the top Lyapunov exponent. Using (Ling 2007) then can be approximated as,

The arguments in (FinMetrics) allows us to generalise the above moment conditions to any SRE. Therefore, we can tabulate the ()-moment conditions for model (XXX) and (XXX) as,

|  |  |
| --- | --- |
| **Existence of Moments for** | |
|  | DAR() and DARMA(,) |
| **Moment Condition** | **Constraint** |
| Strict Stationarity : |  |
|  |  |
| Weak Stationarity: |  |

For all , it holds that if then also holds. Generally, implies that . It is worth emphasising that implies that , which is a sufficient condition to perform one step-ahead forecasting. However, forecasting with confidence bands is not necessarily possible as is needed for the conditional variance to exist in . Even though the mean and variance are not defined under strict stationarity, we can still estimate the model consistently and perform simulations (FinMetrics). This is useful for e.g. derivatives pricing and risk management, where it is crucial that the dynamics of the underlying process are represented accurately. Thus, we see that depending on which moment exists, we are guaranteed certain practical applications of the model.

When the moment conditions reduce to,

From results in (Ling 2004) and appendix XXX, the moment conditions for can be explicitly derived and are stated in Table XXX,

|  |  |  |  |
| --- | --- | --- | --- |
| **Implication** | **Moment Condition** | **Constraint** | |
|  | | DAR(1) | DARMA(1,1) |
| Fractional moments (strict stationarity) | for |  |  |
| Absolute first-order moments |  |  |  |
| Second-order moments (weak stationarity) |  |  |  |

Intuitively, we find that if , the DARMA(1,1) moment conditions reduce to those of the DAR(1).

**Estimation**

The models XXX, XXX and XXX with the augmented volatility process XXX, can be estimated using the Maximum Likelihood Estimator (MLE),

Under the assumption that the standardised residuals follow a student-t distribution, the conditional distribution of electricity prices is also student-t distributed,

, with corresponding log-likelihood function,

With where , , and . The parameter denotes the degree of freedom of the student-t distribution and determines the thickness of the tails. As shown in Table XXX, the set of model parameters changes of quite a lot based on the model, with the (XXX) having parameters more than (XXX) with (XXX).

|  |  |
| --- | --- |
| **Model** | **Number of Parameters** |
| DAR() |  |
| DARMA |  |
| DARMA-X |  |

Based on (Ling 2007), (Jiang 2019) and (FinMetrics) we can state the following results for the MLE:

1. The MLE is consistent such that under the assumption that and is strictly stationary and ergodic, i.e. for
2. Additionally, the MLE is asymptotically Gaussian distributed such that with some positive semidefinite covariance matrix if and

We can derive the fourth order moment of the scaled student-t distribution as,

For then it holds that . However, as the focus of this paper is on forecasting rather than inference, we only require that the MLE is consistent, which in term imposes much looser restrictions on . In fact, we only require that , which directly follows from the assumption of unit variance .

Due to the complexity and non-linearity of the models, the MLE has no closed form solution, and we therefore need to use numerical methods to optimise the objective function. We will use the Sequantial Least Squares Programming (SLSQP) optimisation algorithm, as it is ideal for mathematical problems for which the objective function and the constraints are twice continously differentable ([link](https://qiskit-community.github.io/qiskit-algorithms/stubs/qiskit_algorithms.optimizers.SLSQP.html?utm_source=chatgpt.com)).

**Appendix (Proofs)**

Some useful rules for absolute value:

* Rule 1) Multiplicativeness:
* Rule 2) Subadditivity 1:
* Rule 3) Subadditivity 2:
* Rule 4) Equivalence:

**DAR(1) Moments**

**DAR(1): Absolute First Order Moment**

Let be the drift function such that is bounded. Then,

Using that Rule 3) and Rule 4),

Using Rule 2) and Rule 1) as well as the fact that is contained in the information set and is independent of ,

Since by assumption then,

Using the definition this can be written as,

Using Rule 2) and that ,

Since is a constant we have that,

, where the constant is defined as . To mimic a first order autoregressive process, this can be reformulated as,

, with such that by definition. Therefore, it is clear that if the drift criterion holds and the process is bounded such that .

**DARMA(1,1) Moments**

**DARMA(1,1): Fractional Moments**

The DARMA(1,1) can be written as,

Using Rule 2),

Now using the SRE representation we get,

Using the top Lyapunov coefficient and that is a scalar,

Using the logarithmic rule , this becomes a sum. Furthermore, the expectation of a sum is equal to the sum of expectations. Since all are i.i.d. then,

Since the expression converges in distribution to,

Thus, if , then the process is strictly stationary and has finite fractional moment such that for some .

**DARMA(1,1): Absolute First Order Moment**

Since the DARMA(1,1) model is not Markov chain we define . Let be the drift function such that is bounded, with denoting the Manhattan norm. Then,

First expectation:

Using Rule 2),

Since and is assumed to be i.i.d., therefore is independent of ,

Using Rule 2) and Rule 1) with the assumption , and inserting

By assumption, which holds for over all , thus . Likewise, for all . Therefore, using standard variance-rules it follows that,

Since for all then,

Therefore, the distribution of becomes,

And, expected absolute value given as,

Inserting back into the first term yields,

For then we have the following asymptotic convergence,

This holds under previously specified assumptions for . Thus, for large the expression converges to,

Second expectation:

(using exactly the same arguments as in the first expectation we get)

Combining both terms yield,

Using the definition this can be written as,

Using Rule 2) and that ,

Since is a constant we have that,

, where the constant is defined as . To mimic a first order autoregressive process, this can be reformulated as,

, with such that by definition. Therefore, it is clear that if the drift criterion holds and the process is bounded such that and therefore .

**DARMA(1,1): Second Order Moment**

Since the DARMA(1,1) model is not Markov chain we define . Let be the drift function such that is bounded, with denoting the Euclidian norm. Then,

First expectation:

All the cross-terms cancel out since ,

Using that is i.i.d., we now need to derive . With the distributive property this can be decomposed as,

Then the taking the expectation yields,

Since and for then from the assumption of independence. If then sum converges to an infinite geometric series,

Inserting back yields,

Second expectation:

(using exactly the same arguments as in the first expectation we get)

Combining the two yields,

Using the definition this can be written as,

, where the constant is defined as . To mimic a first order autoregressive process, this can be reformulated as,

, with such that by definition. Therefore, it is clear that if the drift criterion holds and the process is bounded such that and therefore .