



Stationarity-Based Approach for Lag Length Selection in Cointegration Analysis of Lamb Wave Data

Phong B. Dao, Wieslaw J. Staszewski

► To cite this version:

Phong B. Dao, Wieslaw J. Staszewski. Stationarity-Based Approach for Lag Length Selection in Cointegration Analysis of Lamb Wave Data. Le Cam, Vincent and Mevel, Laurent and Schoefs, Franck. EWSHM - 7th European Workshop on Structural Health Monitoring, Jul 2014, Nantes, France. <hal-01020405>

HAL Id: hal-01020405

<https://hal.inria.fr/hal-01020405>

Submitted on 8 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

STATIONARITY-BASED APPROACH FOR LAG LENGTH SELECTION IN COINTEGRATION ANALYSIS OF LAMB WAVE DATA

Phong B. Dao, Wieslaw J. Staszewski

AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Poland

phongdao@agh.edu.pl, w.j.staszewski@agh.edu.pl

ABSTRACT

It is well known that when cointegration is used for the analysis of data in structural health monitoring applications, the choice of lag length has strong influence on damage detection results. The paper demonstrates how this problem could be solved. The solution utilizes the inversely proportional relationship between damage severity and stationarity of cointegration residuals. The method is validated using Lamb wave data from a structure exposed to temperature variations. The experimental results show that the proper lag length selection is essential and this can be achieved with the appropriate statistical analysis.

KEYWORDS : *structural damage detection, Lamb waves, temperature variations, stationarity, cointegration analysis, lag length selection.*

INTRODUCTION

Lamb waves are widely used for damage detection, particularly in smart structures with integrated, low-profile piezoceramic transducers. Various methods based on Lamb waves have been developed since the early 1960s, as discussed in [1–5]. However, despite considerable research effort, practical engineering applications of this technique are still limited. This is not only due to the complex wave propagation mechanism associated with Lamb waves but also due to operational and environmental conditions that can contaminate Lamb wave responses collected from real engineering structures [6]. Temperature variability (instantaneous, daily or seasonal) is one of the major problems [7] since Lamb wave features – used for damage detection – can be modified by temperature [8]. Therefore, compensation for this effect is important to develop methods that are sensitive only to damage but insensitive to operational-environmental conditions.

Various approaches were developed to deal with the undesired effect of temperature variability in data used for damage detection, as discussed in [6]. The *cointegration* approach – developed originally in the field of econometrics [9] – has been recently proposed as a new methodology for dealing with the problem of operational/environmental variability in Process Engineering [10] and Structural Health Monitoring [11–13]. The major idea used in these investigations is based on the concept of *stationarity*. Monitored variables are cointegrated to create a stationary residual whose stationarity represents intact condition. Then any departure from stationarity can indicate that monitored processes, objects or structures are no longer operating under normal condition. The work presented in [10, 11] shows that if some variables from a process under investigation are cointegrated, the stationary linear combinations of these variables during the cointegration process are purged of all common trends in the original data, leaving residuals equivalent to the long-run dynamic equilibriums of the process. This work has been extended to the concept of multiple cointegration analysis in [12], which demonstrates a non-conventional approach to cointegration for temperature effect compensation (i.e. data normalisation) and damage detection in Lamb wave based damage detection of aluminium structures. More recently this approach has been used for multiple temperature trend removal [13].

There are two major issues that one has to consider when using cointegration analysis [14]. Firstly, the number of lags to include in the model must be determined. Different criteria used for lag length selection often lead to different decisions regarding the optimal lag order that should be used in the model [14–18]. Secondly, the choice of lag length can drastically affect the results of the cointegration analysis. This is due to the fact that the cointegration procedure gives different estimates of cointegrating vectors depending on the number of lags included in the cointegration test [14–16]. Hence, the proper selection of lag length for cointegration analysis is very important.

This paper aims to address the problem of optimal lag length selection in cointegration analysis used for Lamb wave based damage detection. A new approach – based on stationarity analysis – is proposed. Cointegration residuals from undamaged data are analysed for various lag lengths. The lag length that produces the most negative statistics (or in other words the most stationary residuals) is then used for damage detection based on the cointegration analysis. The method is illustrated using Lamb wave data from a damaged metallic plate exposed to temperature variations.

The paper is organized as follows. Section 1 introduces the cointegration method and ADF unit root test. Section 2 presents a stationarity-based approach proposed for lag length selection in cointegration analysis used for structural damage detection. The Lamb wave experimental data used to illustrate the method are presented in Section 3. Damage detection results based on the optimal lag length selection are presented and discussed in Section 4. Finally, some conclusions are given.

1 COINTEGRATION ANALYSIS

A set of non-stationary time series, $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})^T$, are cointegrated if there exists (at least) a vector $\beta = (\beta_1, \beta_2, \dots, \beta_n)^T$ such that it results in a linear combination of them that is stationary, i.e.

$$\beta^T Y_t = \beta_1 y_{1t} + \beta_2 y_{2t} + \dots + \beta_n y_{nt}. \quad (1)$$

The stationary linear combination $\beta^T Y_t$ is referred to as a *cointegration residual* or a long-run equilibrium relationship between time series [17]. The vector β is called a *cointegrating vector*. It is important to note that the work presented in this paper considers the action of creating the cointegration residual ($u_t = \beta^T Y_t$) as the action of projecting the (non-stationary) time series Y_t on the cointegrating vector β . A non-stationary time series y_i in Y_t is *integrated order d* , denoted $y_i \sim I(d)$, if after differencing the series d times it becomes stationary. The number of differences required to achieve stationarity is called the order of integration.

In essence, testing for cointegration is testing for the existence of stationary linear combinations among all elements of Y_t [17]. Such tests have two important requirements. Firstly, any analysed time series must *exhibit at least a common trend*. Secondly, the analysed time series must *have the same degree of non-stationarity*, i.e. must be integrated of the same order.

When there are only two variables in Y_t , a two-step residual-based test procedure – developed in [9] – can be used. This procedure is based on regression techniques for determining if the vector β is a cointegrating vector. The first step is to form the cointegration residual $u_t = \beta^T Y_t$. The second step is to perform a unit root test on u_t to determine if it is a stationary time series. The *Augmented Dickey-Fuller (ADF) test* – described in [19] – is the most widely used unit root test in practice. The ADF test checks the null hypothesis that a time series is a non-stationary type series against the alternative hypothesis that it is a stationary type series, assuming that the dynamics in the data have an *Auto-Regressive Moving Average (ARMA)* structure [17]. The ADF test is based on estimating the following regression formula

$$y_t = \tau D_t + \phi y_{t-1} + \sum_{j=1}^p \alpha_j \Delta y_{t-j} + \varepsilon_t. \quad (2)$$

where TD_t is a deterministic linear trend. In Equation (2), the p lagged difference terms or lag length ($\sum_{j=1}^p \alpha_j \Delta y_{t-j}$) are used for approximating the ARMA structure of the errors. The value of the lag length p is set to a value, so that the error ε_t is a white noise process [17].

When Y_t includes more than two variables, a sequential procedure for determining the existence of cointegration – developed in [20] and known as the Johansen's cointegration method – is widely used. This procedure is a combination of cointegration and error correction models in a *Vector Error Correction Model (VECM)* that takes the form

$$\Delta Y_t = \Phi D_t + A B^T Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t. \quad (3)$$

where A and B are $(n \times r)$ matrices with $\text{rank}(A) = \text{rank}(B) = r$ and the first term on the right-hand side – i.e. ΦD_t – contains deterministic terms (e.g. constant only or constant plus time trend). The stationary linear combinations ($u_{rt} = B^T Y_t$) are referred to as the r cointegration residuals that are formed through projecting the (non-stationary) time series Y_t on the r cointegrating vectors. The non-deterministic part on the right hand side of Equation (3), i.e. $\Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1}$, denotes the $p-1$ lagged difference terms (or the lag length p) used for approximating the VECM.

The Johansen's cointegration method is used in this study for cointegration analysis. The method is a quite complex sequential procedure and therefore it is not presented in this paper. For more detailed description of the entire procedure, potential readers are referred to [20].

2 ALGORITHM FOR LAG LENGTH SELECTION BASED ON STATIONARITY ANALYSIS

The selection of lag length p that should be included in the test regression models in Equation (2) and the VECM in Equation (3) is one of the most important practical issues for the implementation of the ADF test and cointegration analysis. However, this choice is not a trivial task. If p is too small then the remaining serial correlation in the errors will bias the test. If p is too large then the power of the test will suffer [17, 21]. The lag length can be determined by using model selection criteria. The general approach is to fit models with values of (e.g. $p = 0, \dots, p_{\max}$) and then to select the value of p that minimizes some model selection criteria [17]. Several lag selection criteria have been proposed in the econometric and statistic literature for this purpose. The three most widely used information criteria are: the Akaike Information Criterion (AIC), the Schwarz-Bayesian Criterion (SBC) and the Hannan-Quinn Criterion (HQC). However, the choice of these information criteria for determining the number of lags is generally arbitrary in practice and sometimes these criteria are inconsistent in choosing the lag order [14–17]. Therefore, there are many arguments in the econometric and statistical literature with respect to the selection of lag length in cointegration analysis. However, these arguments give very little practical guidance that could be used in engineering applications. This is mainly due to the fact that the size of engineering data is usually much larger, if compared with the data used in the field of econometrics and statistics.

This section presents a new approach that can be used for lag length selection in damage detection studies based on cointegration analysis. The method utilises the concept of stationarity.

Previous applications of cointegration for damage detection show that the ADF test is firstly carried out to measure the degree of stationarity or non-stationarity (i.e. the order of integration) of the analysed data. In principle, the more negative the ADF t-statistic value obtained, the more stationary the data are, as illustrated in [12, 13]. Usually data representing undamaged condition of monitored structures are stationary time series. The assumption is that this stationarity can be potentially changed by damage. In addition, different severities of damage can lead to different stationary characteristics. Therefore, analysis of stationarity can be used for optimal lag selection in damage detection investigations. The algorithm proposed for lag length selection in cointegration

analysis used for damage detection is illustrated in Figure 1. The entire procedure can be described using four major steps:

- **Step 1: Determine the p_{\min} and p_{\max} values.** It is clear that $p_{\min} = 1$ is the minimum value of lag length that could be used in cointegration analysis. The maximum lag length value p_{\max} can be calculated using the following equation [22].

$$p_{\max} = \left\lceil 12 \left(\frac{N}{100} \right)^{1/4} \right\rceil. \quad (4)$$

where the square brackets denote the integer part of the result, and N is the number of data samples. Equation (4) guarantees that p_{\max} grows with the number of data samples used.

- **Step 2: Cointegration analysis.** After the p_{\min} and p_{\max} values are established, N sets of Lamb wave data representing undamaged condition are cointegrated using the Johansen's cointegration procedure. This results in $N - 1$ linearly independent cointegrating vectors. These vectors are then used to produce $N - 1$ cointegration residuals by performing the so-called "*undamaged data on undamaged data*" projection. This projection means that data representing undamaged condition are projected on cointegrating vectors obtained from data representing undamaged condition. The entire analysis is performed for all lag length values $p = 1, 2, 3, \dots, p_{\max}$, leading to a $(N - 1) \times p_{\max}$ matrix of cointegration residuals.
- **Step 3: ADF test.** ADF t-statistics are calculated for all cointegration residuals (i.e. $(N - 1) \times p_{\max}$ matrix of cointegration residuals) and lag lengths. As a result, $N - 1$ ADF t-statistics are obtained for each value of lag length.
- **Step 4: Averaged ADF t-statistics calculation.** An averaged value of ADF t-statistics is calculated for each lag length $p = 1, 2, 3, \dots, p_{\max}$. The most negative value from all averaged ADF t-statistics indicates the value of lag length that produces the most stationary residuals obtained for the undamaged data. The assumption is that the selected lag length is the optimal value, leading to the best results when cointegration analysis is used for damage detection.

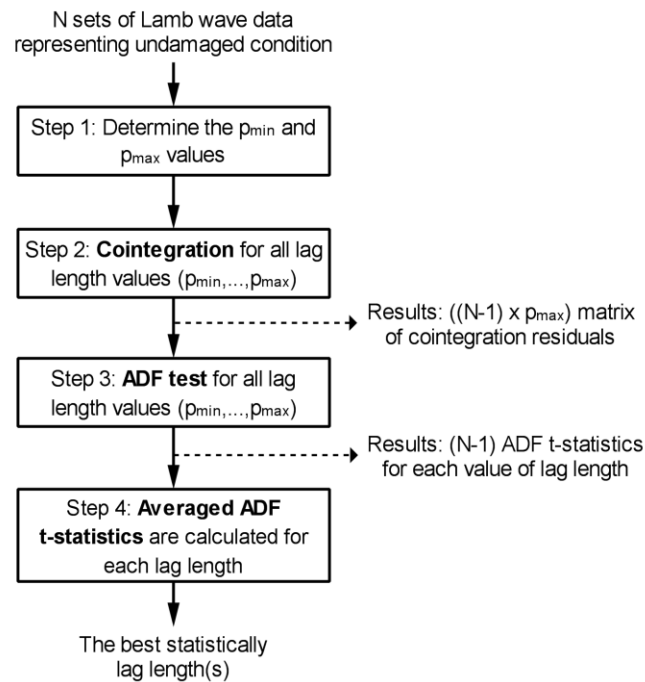


Figure 1: Stationarity-based lag length selection procedure

3 LAMB WAVE DATA CONTAMINATED BY TEMPERATURE VARIATIONS

Lamb wave experimental data [23] was used in this paper to illustrate the lag length selection method for damage detection based on cointegration. The data were gathered from an aluminium plate (200 x 150 x 2 mm). The plate was instrumented with two low-profile, surface-bonded piezoceramic *Sonox P155* transducers (diameter 10 mm and thickness 1 mm) that were used for Lamb wave generation and sensing. A five-cycle 75 kHz cosine burst signal of maximum peak-to-peak amplitude equal to 10 V was enveloped using a half-cosine wave and then used for excitation. The excitation signal was generated using the *TTi TGA 1230* arbitrary waveform generator. Lamb wave responses were acquired using a digital 4-channel *LeCroy LT264 Waverunner* oscilloscope. The plate was placed in a 100 liter *LTE Scientific* oven to obtain data for various temperatures. The temperature on the surface of the plate was monitored using a thermal probe.

Firstly, the experimental tests were performed using the intact (or undamaged) plate that was firstly heated up (from 35°C to 70°C) and then cooled down (from 70°C to 35°C) with a step change of 5°C. The heating and cooling cycles were performed twice to address the problem of repeatability and check for possible hysteresis between cycles. Then, a hole was drilled the middle of the plate and the entire experimental work was repeated. The analysis presented in this paper utilised Lamb wave response data for four different damage conditions (i.e. the undamaged plate and the damaged plates with 1, 3 and 5 mm holes) and four different temperatures (i.e. 35, 45, 60 and 70°C). Altogether twenty (i.e. $N = 20$) Lamb wave responses were used for single combined damage-temperature conditions. Each response measurement consisted of 5000 data points acquired using the sampling rate of 10 MHz. Strong influence of temperature on Lamb wave responses (amplitude and phase) was observed, as reported previously in [23].

4 RESULTS AND DISCUSSION

Lamb wave experimental data – described in Section 3 – were used to illustrate the algorithm for optimal lag length selection in cointegration analysis applied for structural damage detection. Following the description given in Section 2, the maximum lag length value was computed using Equation (4) as $p_{\max} = 31$. The minimum value was selected arbitrarily as $p_{\min} = 1$.

The cointegration analysis was first used for Lamb wave data representing the undamaged condition. This analysis – performed for all lag lengths investigated, i.e. $p = 1, 2, 3, \dots, 31$ – resulted in $N - 1 = 19$ cointegration residuals for each value of lag length. The “*undamaged data on undamaged data*” projection was used in the analysis. Figures 2a and 2b show examples of the cointegration residuals calculated for $p = 5$ and $p = 30$, respectively. Then the ADF test was performed to obtain t-statistics for all cointegration residuals. The results – given in Figure 3 – display the variability of t-statistics for all values of lag length investigated. The values of t-statistics were then averaged for each lag length. The results – given in Figure 4 – show that the minimum averaged value of t-statistics was obtained for $p = 1$. However, this value of lag length is too small and therefore not considered in any further analysis, as explained in Section 2. The remaining averaged t-statistic values exhibit a clear “deep” for $p = 4, 5, 6$ with the local minimum achieved for $p = 6$. These three lag lengths are considered as the best values for the cointegration analysis. The assumption is that if one of these values of lag length is used, damage detection procedure will give much better results, if compared with other lag lengths.

In order to confirm the proposed approach, two of the best statistically lag lengths found (i.e. $p = 4$ and $p = 6$) and two arbitrarily chosen lag lengths (i.e. $p = 17$ and $p = 27$) were selected for damage detection analysis. This time Lamb wave data for the damaged plates (i.e. the ones with 1, 3 and 5 mm holes) that was exposed to different temperature conditions were used. The damage detection analysis involved the “*damaged data on damaged data*” projection. This projection

means that data representing damage conditions are projected on the cointegrating vectors obtained from data representing damage conditions.

After the cointegration analysis was used, the ADF test was applied to cointegration residuals and ADF t-statistics were calculated. These statistics were used for damage detection to separate data representing damaged and undamaged conditions.

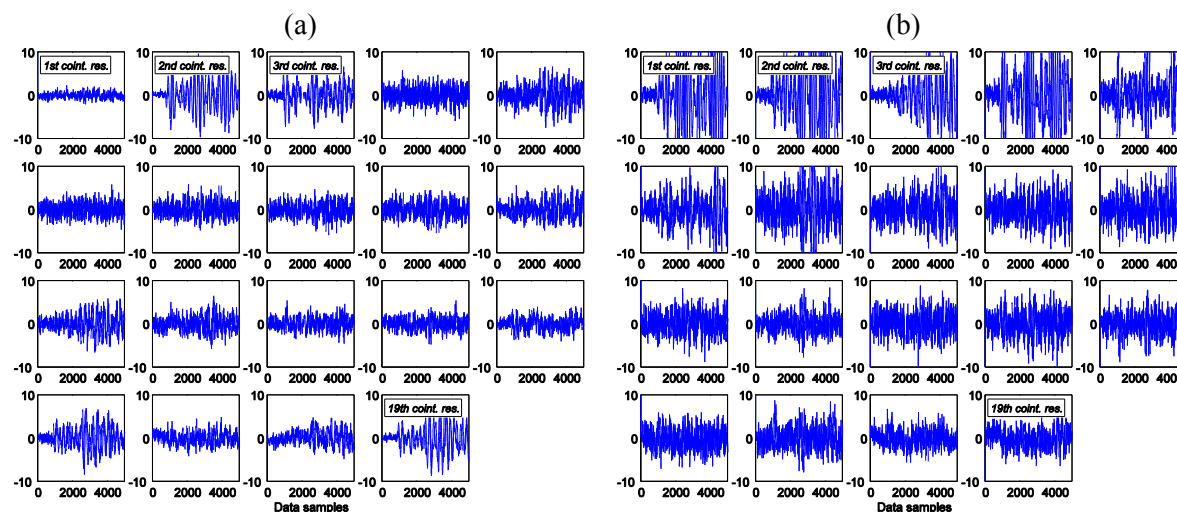


Figure 2: Examples of cointegration residuals calculated from Lamb wave responses representing the undamaged condition. The “undamaged data on undamaged data” projection was used in the analysis:

(a) lag length $p = 5$; (b) .lag length $p = 30$.

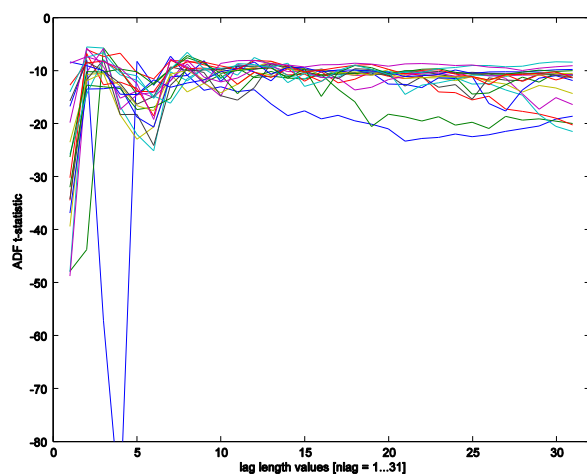


Figure 3: ADF test results for Lamb wave data representing the undamaged condition. The t-statistics were calculated for all residuals and lag lengths.

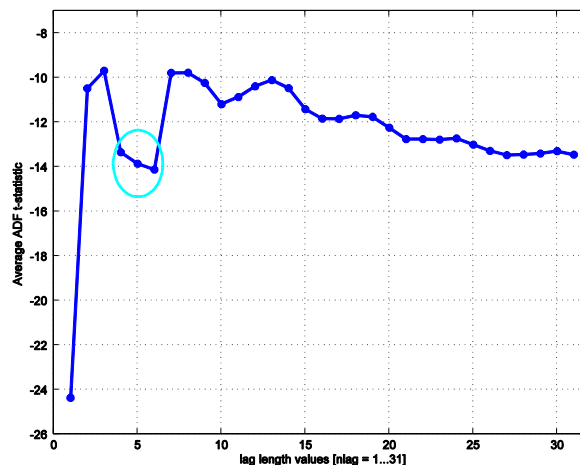


Figure 4: The same as Figure 3 but the averaged ADF t-statistics are given. For a given value of lag length averaging was performed over cointegration residuals.

Figure 5 presents selected damage detection results calculated for the Lamb wave data representing the undamaged plate and the 3 mm hole damaged plate at 35°C. The results show that the average ADF t-statistics for the damaged plate with the 3 mm hole are very well separated from the relevant t-statistics calculated for the undamaged plate for all nineteen cointegration residuals, when the lag lengths are equal to $p = 4$ and $p = 6$. Thus the investigated seeded damage can be detected. In contrast, when the lag lengths $p = 17$ and $p = 27$ are used, the average ADF t-statistics

– for both undamaged and damaged plates – overlap for the majority of cointegration residuals. Damage detection is questionable this time and possible only for a handful of cointegration residuals that are difficult to select in practice when the specimen is undamaged.

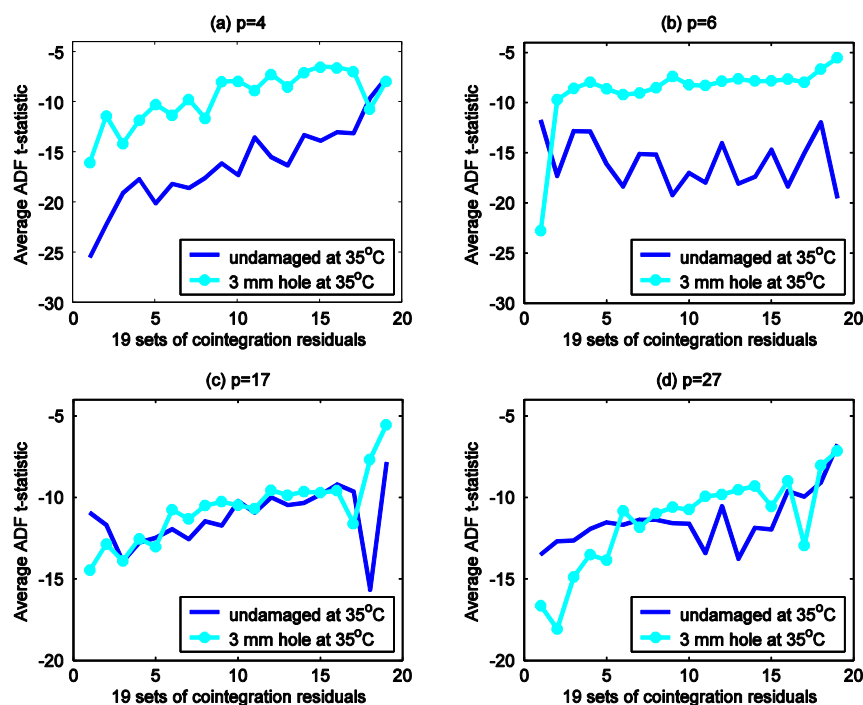


Figure 5: Damage detection based on Lamb wave data – average ADF t-statistics calculated for the cointegration residuals representing the undamaged plate and the plate with the 3 mm hole exposed to 35°C. The analysis was performed for the lag lengths: (a) $p = 4$; (b) $p = 6$; (c); $p = 17$; (d) $p = 27$.

CONCLUSION

The problem of optimal selection of lag length in cointegration analysis – used for structural damage detection – has been addressed. A new approach – based on stationarity analysis – has been proposed. The method investigates various lag lengths for data representing undamaged condition. The lag length that produces the most negative statistics (or in other words, the most stationary residuals) is then used for damage detection based on the cointegration analysis. The method has been illustrated using Lamb wave data from damaged metallic plates exposed to temperature variations.

The results show that that lag lengths that produce the most stationary cointegration residuals for the data representing undamaged condition, give better damage detection results than arbitrarily selected lag lengths. Damage detection was successful when the value of lag length was selected following the proposed methodology. In contrast, damage detection was not possible when other lag length values were used. The work presented is a feasibility study. Therefore, further research work is required to confirm all findings.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by funding from WELCOME research project no. 2010-3/2 sponsored by the Foundation for Polish Science (Innovative Economy, National Cohesion Programme, EU).

The experimental Lamb wave tests were performed in the Department of Mechanical Engineering at Sheffield University, UK. The technical assistance of Dr Boon Lee in these tests is acknowledged.

REFERENCES

- [1] S. S. Kessler, S. M. Spearing, C. Soutis. Damage Detection in Composite Materials Using Lamb Wave Methods. *Smart Materials and Structures*, 11 (2002) 269-278.
- [2] W. J. Staszewski. Structural Health Monitoring Using Guided Ultrasonic Waves, in: J. Holnicki-Szulc and C.A. Mota Soares (Eds.). *Advances in Smart Technologies in Structural Engineering*, Springer, Berlin, 2004, pp. 117-162.
- [3] Z. Su, L. Ye, Y. Lu. Guided Lamb Waves for Identification of Damage in Composite Structures: A Review. *Journal of Sound and Vibration*, 295 (2006) 753-780.
- [4] A. J. Croxford, P. D. Wilcox, B. W. Drinkwater, G. Konstantinidis. Strategies for Guided-Wave Structural Health Monitoring. *Proc. R. Soc. A.*, 463 (2007) 2961-2981.
- [5] A. Raghavan, C. E. S. Carlos. Review of Guided-Wave Structural Health Monitoring. *The Shock and Vibration Digest*, 39 (2007) 91-114.
- [6] H. Sohn. Effects of Environmental and Operational Variability on Structural Health Monitoring. *Phil. Trans. R. Soc. A.*, 365 (2007) 539-560.
- [7] F. Lanza di Scalea, S. Salamone. Temperature Effects in Ultrasonic Lamb Wave Structural Health Monitoring Systems. *Journal of the Acoustical Society of America*, 124 (2008) 161-174.
- [8] A. Marzania, S. Salamone. Numerical Prediction and Experimental Verification of Temperature Effect on Plate Waves Generated and Received by Piezoceramic Sensors. *Mechanical Systems and Signal Processing*, 30 (2012) 204-217.
- [9] R. F. Engle, C. W. J. Granger. Cointegration and Error-Correction: Representation, Estimation, and Testing. *Econometrica*, 55 (1987) 251-276.
- [10] Q. Chen, U. Kruger, A. Y. T. Leung. Cointegration Testing Method for Monitoring Non-stationary Processes. *Ind. Eng. Chem. Res.*, 48 (2009) 3533-3543.
- [11] E. J. Cross, K. Worden, Q. Chen. Cointegration: A Novel Approach for The Removal of Environmental Trends in Structural Health Monitoring Data. *Proc. R. Soc. A*, 467 (2011) 2712-2732.
- [12] P. B. Dao, W. J. Staszewski. Cointegration Approach for Temperature Effect Compensation in Lamb Wave Based Damage Detection. *Smart Materials and Structures*, 22 (2013) 095002.
- [13] P. B. Dao, W. J. Staszewski. Data Normalisation for Lamb Wave-Based Damage Detection Using Cointegration: A Case Study with Single- and Multiple-Temperature Trends. *Journal of Intelligent Material Systems and Structures*, 25 (2014) 845-857.
- [14] J. Emerson. Cointegration Analysis and The Choice of Lag Length. *Applied Economics Letters*, 14 (2007) 881-885.
- [15] A. Hatemi-J, R. S. Hacker. Can the LR test be helpful in choosing the optimal lag order in the VAR model when information criteria suggest different lag orders? *Applied Economics*, 41 (2009) 1121-1125.
- [16] V. K. S. Liew. Which lag length selection criteria should we employ?. *Economics Bulletin*, 3 (2004) 1-9.
- [17] E. Zivot, J. Wang. *Modeling Financial Time Series with S-PLUS*, second ed., Springer, New York, 2006.
- [18] M. Golder. *Time Series Models*. Lecture Note on Methods IV: Advanced Quantitative Analysis, September 2012, available from: <https://files.nyu.edu/mrg217/public/timeseries.pdf>.
- [19] D. Dickey, W. Fuller. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49 (1981) 1057-1072.
- [20] S. Johansen. Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12 (1988) 231-254.
- [21] J. Gonzalo. A Comparison of Five Alternative Methods of Estimating Long Run Equilibrium Relationships. *Journal of Econometrics*, 60 (1994) 203-33.
- [22] W. Schwert. Test for Unit Roots: A Monte Carlo Investigation. *Journal of Business and Economic Statistics*, 7 (1989) 147-159.
- [23] B. C. Lee, G. Manson, W. J. Staszewski. Environmental Effects on Lamb Wave Responses from Piezoceramic Sensors. *Materials Science Forum: Modern Practice in Stress and Vibration Analysis*, Trans Tech Publications, Switzerland, 440-441 (2003) 195-202.