City Research Online

City, University of London Institutional Repository

Citation: Cook, S. & Fosten, J. (2019). Replicating rockets and feathers. Energy Economics, 82, pp. 139-151. doi: 10.1016/j.eneco.2017.12.021

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/33302/

Link to published version: https://doi.org/10.1016/j.eneco.2017.12.021

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Replicating Rockets and Feathers*

Steven Cook† Jack Fosten‡

December 7, 2017

Abstract

This paper revisits the literature of asymmetric adjustment in gasoline and diesel prices, also known as the 'rockets and feathers' hypothesis, to consider the issue of the replication of empirical research. We examine the notion of replication versus robustness proposed by Clemens (2017) and add to the literature with a further review of recent and historic work on replication in economics and other disciplines. We then focus on the rockets and feathers literature, finding that the majority of empirical work performs robustness checks rather than replication of earlier papers. We perform two contrasting replication case studies motivated by the ideas of misspecification analysis, dynamic specification, mark-up, pass-through and asymmetric adjustment. In the first case study we find that results are both replicable and robust, even when data specifications are not identical. However, in the second case study we find that the results using the original sample are overturned when reanalysing the problem using an improved model specification. Furthermore, when extending the sample with more recent data and using a more sophisticated method, asymmetry is detected in both petrol and diesel pricing; different to the findings of the original study. Particular care must be taken in future rockets and feathers replications with regard to model specification and methodology.

JEL Classification: C22, C52, Q41

Keywords: Replication, rockets and feathers, asymmetric pricing, petrol, diesel

*We are grateful to the Editor and three anonymous referees whose comments have helped to improve the content and presentation of this paper.

†School of Management, Swansea University, Bay Campus, Fabian Way, Swansea SA1 8EN. E-mail address: s.cook@swansea.ac.uk.

‡Corresponding Author: Department of Economics, University of East Anglia, Norwich, NR4 7TJ, UK. E-mail address: j.fosten@uea.ac.uk.

1 Introduction

The rockets and feathers literature has progressed at a rapid pace in the recent decades since the seminal work of Bacon (1991). Competition authorities across the globe have long been interested in the question of whether retail gasoline and diesel prices rise more quickly than they fall, relative to the movements in underlying input costs. The number of empirical applications in this field has become vast, with papers looking at the problem through various lenses: different countries, different variables, different econometric methodologies, different time spans and so on. These are so wide-ranging that different meta-analyses have been undertaken to summarize the various findings, such as Frey and Manera (2007) and more recently Perdiguero-García (2013). In this paper we wish to assess the extent to which the key findings in the rockets and feathers literature have been "replicated" by this growing body of empirical work, or whether more work is still to be done.

The issue of replication in economics cannot be considered as widespread as in the natural sciences, although it has received more attention in recent years. Indeed, many journals nowadays such as the *American Economic Review* and the *Journal of Political Economy* expect authors to submit data and computer codes with the idea of replication in mind. The *Journal of Applied Econometrics* even allows the submission of articles to a designated replication section which has a separate editor. There is also a Replication Wiki website[1] which provides a more informal outlet for replications of economic papers. However, one of the main challenges facing economists is the question of "what constitutes a replication in economics"?

The recent paper by Clemens (2017) provides a summary of the different terminologies used for replication in economics and, importantly, draws up a standard for classifying replications and differentiating them from robustness analysis. Clemens (2017) firstly splits replication into the categories of "verification" and "reproduction". These approaches both use the same model specification as the paper of interest, but verification attempts to use exactly the same sample (or as close as possible) whereas reproduction draws a different sample from the same population. Replication is then differentiated from robustness, which is split into the categories of "reanalysis" and "extension". Reanalysis usually uses the same sample as the original work but chooses a different model specification or methodology. Extension, on the other hand, uses the same model specification but changes the sample, typically by changing the country/region or time periods of analysis.

The first contribution of this paper is to review the literature, both on replication in general, and then specifically on the rockets and feathers phenomenon. We first focus on the literature on replication in economics and other disciplines, both in historical and recent research. This serves to build further upon the notions discussed in Clemens (2017) where we concur that replication should be an important feature of economic research. We then survey the rockets and feathers literature,

[1]See http://replication.uni-goettingen.de/wiki/index.php/Main_Page [Last accessed: 16/05/17]

making explicit use of the replication versus robustness framework of Clemens (2017). Since there is now a vast number of rockets and feathers papers, we do not attempt to review all of them. These have already been comprehensively surveyed in the aforementioned meta-analyses and other more recent papers such as Kristoufek and Lunackova (2015). Rather, we focus on some important papers with different countries or methodologies and question whether subsequent papers fall into the category of replication or robustness. The broad conclusion is that most recent empirical work can only be considered as robustness analysis as it does not attempt to verify and replicate the results in prior papers using the same sample and methodology.

The next contribution is to take two contrasting case studies from the rockets and feathers literature and follow the various stages of replication and robustness checks. These not only contrast in terms of country, time span, data periodicity, variable definitions and methodology used, but they also contrast in terms of their conclusions as one study detects asymmetric pricing and the other does not. In the first case study, we choose to revisit the analysis of Bachmeier and Griffin (2003), hereafter BG, published in the *Review of Economics and Statistics*. BG analyse the asymmetric adjustment of gasoline prices in the U.S. using daily price data, finding no evidence of asymmetries. This is a useful study to replicate, not only because of the importance of the U.S. gasoline market, but also due to the challenging perspective of replication where the original data is unavailable. Since the authors used a proprietary data source, we have to look elsewhere to obtain similar data over the same sample period in order to perform a replication through reproduction rather than verification.[2] However, we find that the results of BG are not only replicable but are also robust to the extension of the sample period.

In the second case study, we go into more detail and focus on the case of pre-tax petrol and diesel prices in New Zealand by replicating the paper of Liu et al. (2010), hereafter LMT, published in *Energy Economics*. This choice is motivated for several reasons. Firstly, the data used by LMT are publicly accessible and well-documented, unlike that of the first case study, thereby increasing the probability of a successful replication. Secondly, the case of New Zealand is particularly interesting as the retail petrol market is almost entirely controlled by oil companies. As mentioned in LMT, this simplifies matters by eliminating differing channels of asymmetric pricing which might otherwise be possible due to different pricing powers along the distribution chain. Thirdly, the sample used by LMT runs from April 2004 to February 2009 and only includes a small amount of sample after the global economic crisis. This gives further interest to the extension of the sample size and subsequent reanalysis.

The conclusions of this second case study, unlike those of the first, shed some important light on

[2]During the research for this project we also found that the results of the earliest papers like Bacon (1991) were not possible (or incredibly difficult) to even replicate through reproduction. In the case of Bacon (1991), the original data were taken from a print source (*Petroleum Times*) which was discontinued in the early 2000's and is now not even available in print format in the majority of libraries. Publicly available data on U.K. retail gasoline prices is subsequently only available since 2003. This is discussed further in the Literature Review, below.

the issues of replication and robustness in the rockets and feathers literature which we generalise to give advice for future studies. We firstly establish that the results of LMT can be almost perfectly verified, with asymmetric pricing found for diesel but not petrol. This is with the exception of some noise potentially caused by revisions to the data. However, we find that a reanalysis using the original sample but a different (improved) dynamic specification reverses the conclusion that diesel is asymmetrically priced. This would suggest that future studies should be careful with lag selection, and that robust model selection procedures or possibly model averaging could be employed. Finally, we perform an extension and reanalysis by increasing the sample size up until the end of 2016 using the nonlinear ARDL (NARDL) methodology of Shin et al. (2014) rather than the traditional asymmetric Error Correction Model (ECM). This further alters the conclusions in that asymmetric pricing is now detected in both petrol and diesel. In response to these differing findings for the original and extended samples, behaviour over the post-LMT period is examined. Interestingly, this analysis provides evidence of markedly different behaviour in the series as compared to the original sample. This could imply that regular checks should be made on the pricing of both diesel and petrol as the conclusions about asymmetric pricing seem to alter over time.

The rest of this paper is organised as follows. Section 2 reviews the literature on replication and the rockets and feathers phenomenon. Sections 3 and 4 provide the two case studies we consider, each with a discussion of the data, methodology and results on the replications. Finally, Section 5 concludes the paper.

2 Literature Review

We provide two separate reviews of the literature. In section 2.1 we broadly review the literature on replication in economics and other disciplines, adding to the analysis of Clemens (2017) by providing additional evidence in their useful framework of classifying replications. In section 2.2 we consider specifically the rockets and feather literature from the angle of replication.

2.1 Replication in Economics and other Disciplines

While attempting to clarify the notion of 'replication', Clemens (2017) notes the increase in attention paid to this issue in economics and social sciences. We will extend this issue via consideration of three bodies of work, namely studies providing discussion of replication further to those considered by Clemens (2017); subsequent studies considering 'replication' in economics; and 'historical' re-analyses conducted within economics which have considered in the issue of replication, albeit implicitly.

Recently, McNeeley and Warner (2015) considered the degree of replication occurring within criminology relative to social and natural sciences, giving troubling findings regarding the low levels of replication coupled with the amount of replications which conflict with the original study. They therefore question the value of single empirical studies (see also Winfree, 2010) and sought to encourage the undertaking of replication studies despite recognising the lack of incentives and rewards for individuals to undertake such activity in academia, a point also made by Frank and Saxe (2012). Finally, Makel and Plucker (2014) find that, like other disciplines, low levels of replication have occurred in educational research and that positive replications were often caused by an overlap in authorship between the original and replication studies. Clearly, these studies share a common message with Clemens (2017) that replication is to be promoted.

The growing interest in replication in economics, noted by Clemens (2017), is reflected in the two sessions devoted to this topic at the 2017 annual meeting of the American Economic Association and the publication of a series of papers on this topic in the subsequent '*Papers and Proceedings*'. The resulting collection of papers covered a range of topics including examination of the replication rate in economics (Berry et al., 2017); alternative initiatives to promote the replication rate in economics including the role of journals, the provision of data and code, and the adoption of ideas in other disciplines (Chang et al., 2017; Coffman et al., 2017; Höffler, 2017); meta analyses not being an alternative to replication (Anderson and Kichkha, 2017); the factors underlying the replication of studies, with Google scholar citation counts found a prominent issue (Sukhtankar, 2017).

A final replication-related element of research to consider is that associated with the increased interest in the history of econometrics witnessed in the 1980s. This was reflected in the subsequent works of De Marchi and Gilbert (1989), Gilbert (1986, 1989, 1991), Morgan (1990) and Thomas (1989, 1992). Associated with this research are empirical studies such as Spanos (1989) and Cook (2000) which, although not considered by Clemens (2017), give differing illustrations of the importance of clarifying the notion of replication. More precisely, Spanos (1989) revisited the work of Davis (1952) by using a revised version of the same data but new econometric consumption function models, finding that a misguided 'stylised history' of the consumption function had evolved as a result of the simplistic nature of the econometric models considered. In contrast, Cook (2000) reconsidered the early empirical debates on the cyclical nature of the marginal propensity to consume to support the use of the simple model of Woytinsky (1946) using the author's original data. These two studies alone illustrate alternative perspectives employed in the empirical literature when revisiting previous research.[3]

2.2 Rockets and Feathers Literature

The papers in the rockets and feathers literature vary considerably by time span, which country is used and which methodology is employed. We will focus on a few different dimensions in this review, starting the seminal work of Bacon (1991) and then moving on to survey the literature on the U.S.

[3]A further example of the need for replication is the work of Wulwick (1996) who finds that, while the inflation-unemployment findings of Phillips (1958) are replicated, the subsequent results of Lipsey (1960) could not be reproduced with the cause of the 'failure to replicate' being uncertain.

and New Zealand, the countries of our replications, and some newer literature. In all cases, special focus will be in determining whether later studies provided replications (verification or reproduction) or robustness (extension or reanalysis) or both.

The analysis of Bacon (1991) uses a nonlinear partial adjustment model for bi-weekly U.K. retail gasoline prices with respect to ex-refinery petroleum product prices over the period June 1982 to January 1990. These results are very difficult (if not impossible) to verify in the present day as the data came from the print publication, *Petroleum Times*, which was discontinued in 2001 and is only partially available in non-digitized form from some subscription services.[4] More recent data on retail gasoline prices are available but only start in 2003. Studies such as Manning (1991) and Reilly and Witt (1998), written around a similar time, used completely different (monthly) data and employ different methods using different variables over different time periods to the Bacon (1991) paper. Therefore these studies can only be considered as reanalysis of Bacon (1991) and not replication.[5] Recent U.K. studies such as Bermingham and O'Brien (2011) also use a different monthly sample period (1994-2009) and therefore cannot meet the condition for replication of any of the above studies.

For the United States, perhaps the most important early paper was that of Borenstein et al. (1997), which used bi-weekly retail gasoline and crude oil prices from March 1986 to December 1992 and found evidence of asymmetric pricing in gasoline. Subsequently, the paper by Bachmeier and Griffin (2003), upon which our second case study is based, performed an updated analysis to "rigorously test the robustness" of Borenstein et al. (1997) with no direct claim of replication. They note that, although they try to match the variable definitions and time period as closely as possible, they do not exactly use the same sample. However they go on to compare their own methodology with that of Borenstein et al. (1997) both with weekly and daily data, and arrive at the conclusion that asymmetry is found only in the weekly but not daily model. The claim in Bachmeier and Griffin (2003) of being a robustness analysis therefore matches the Clemens (2017) framework as being both an extension and reanalysis, but not a replication. Since these papers, there have been a great many further studies for the U.S. However, as can be seen in the summary Table 1 of Perdiguero-García (2013), these studies use different samples involving a mixture of time spans and sampling frequency.

Turning attention to New Zealand, we first note that there are relatively few rockets and feathers studies; something explicitly mentioned by Kristoufek and Lunackova (2015). The paper of Delpachitra (2002) uses an ECM-type approach with no asymmetries using weekly data on crude oil and domestic prices from 1989 to July 1994. The paper of LMT, our second case study, appears to be one of the

[4]We are grateful to Sian Blake, the Social Sciences Librarian at the University of East Anglia, who traced back the history of this publication. The *Petroleum Times* ran from 1919 but switched title and ISSN several times to *International Petroleum Times* and *Petroleum Times Energy Report* before finally being discontinued in 2001. Certain specialist newspaper subscription services contain some very incomplete coverage of this publication, but not sufficient for verification of Bacon (1991).

[5]In fact, Perdiguero-García (2013) found that the sampling frequency (daily/weekly/monthly) in particular had a significant effect in determining the finding of symmetric versus asymmetric pricing.

first to analyse pricing asymmetries in New Zealand, with the exception of a report prepared for the MED by Twomey and West (2008). A further paper by Alom and Ritson (2012) uses the same data source as LMT, extended to the period April 1994 to June 2011. Their study does not use the same data span as LMT but they do provide analysis using the same methodology, and so their study can be classed as an extension but not replication. They find asymmetric pricing in petrol but not diesel prices. This result is different to that of LMT, however without performing a replication followed by extension, it is difficult to tell whether the different result is due to the original results not being replicable, or due to the extension of the sample period. Our full case study on replication, extension and reanalysis will shed light on whether this conclusion is valid, and from which channel it results.

While papers on New Zealand are scarce, there have been a larger number of works about neighbouring Australia. Interestingly, all of the recent papers of Valadkhani (2013a,b), Valadkhani et al. (2015) and Chua et al. (2017) perform studies on more localised geographical units, in contrast to the above literature which focusses on prices at a national level. This is presumably to due to the greater land mass and geographical dispersion of Australia in comparison with other countries. For example, Valadkhani et al. (2015) look at seven seaport cities whereas Chua et al. (2017) look at twenty-eight retail gas stations in Queensland. While none of these studies are replications, in these cases there is potentially scope for performing a replication by reproduction, as suggested by Clemens (2017).

Finally, there are large number of recent studies using newer methodologies to address the rockets and feather phenomenon.[6] By definition, none of these studies can be classified as a replication, but must be a robustness analysis of the original rockets and feathers phenomenon. Examples of recently proposed methodologies include semiparametric approaches (Polemis and Tsionas, 2016), asymmetric seemingly unrelated regression (Blair et al., 2017) and quantile ARDL models (Lahiani et al., 2017).

3 Replication Case Study I

In this first case study, we aim to perform a replication and extension of the results in BG, who find no evidence of the rockets and feathers phenomenon using U.S. data. They use gasoline and crude oil prices based on the Houston, Texas regional bulk price reported by the data source Platt's. These are daily data over the period February 1985 to November 1998. We first present the data we use to replicate their results before describing the error correction models they use to analyse asymmetric pricing.

[6]By contrast, the meta-analysis of Perdiguero-García (2013) finds that over 70% of the rockets and feathers studies considered use ECM-type methods.

3.1 Data

In forming a dataset for this replication study, as mentioned above, BG use the Houston, Texas regional bulk prices for gasoline and crude oil from Platt's. Since these data are subscription-only, we are interested in seeing whether replication of this study is possible through publicly available sources. We therefore obtain daily data from June 1986 from the U.S. Energy Information Administration (EIA)[7] for the Spot Cushing crude oil price and the New York Harbor Conventional Gasoline Regular Spot price. We firstly note that since these are not identical variables to those used in BG, this may be considered more as a reproduction rather than a verification.[8] We are also missing almost a year or data at the beginning of the sample. This must be taken into consideration when comparing our findings to those of BG.

Figure 1 displays a graph of crude oil prices versus gasoline prices. At this stage we note that the raw crude oil price data comes in units of U.S. dollars per barrel, whereas the gasoline price is in dollars per gallon. While Figure 1 charts these in their native units, for the replication to be successful relative to BG, we convert gasoline prices into dollars per barrel by multiplying by the scale factor 42 in subsequent analysis.

The chart shows that since the BG sample, whose cut-off point is indicated by the vertical line, gasoline and oil prices have become somewhat more volatile and include the oil price crash in 2008 where prices had peaked at around \$145 per barrel and fell to \$32 per barrel. The daily data also display some unusual features of gasoline prices relative to the underlying crude oil price. For example, in September 2005 when hurricane Katrina hit the U.S., panic-buying of gasoline led the price to soar by almost a dollar per litre in a single day in some locations. We will consider these idiosyncracies in our analysis, but they do not have a major bearing on the main features of the results.

3.2 Symmetric and Asymmetric Error Correction Model

The main part of the BG analysis is in comparing symmetric versus asymmetric ECMs using daily data which had not been considered in the preceding literature. The ECMs are based on the underlying long-run regression:

$$y_t = \alpha_0 + \alpha_1 x_t + \varepsilon_t \tag{1}$$

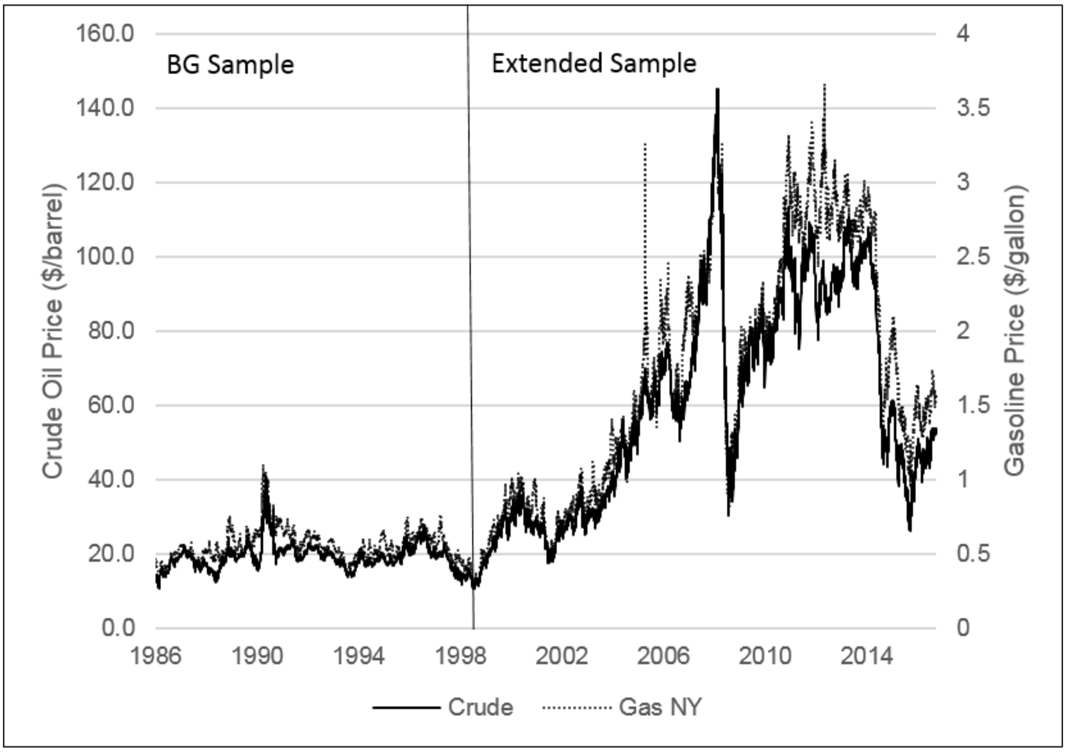
where y_t is the gasoline price and x_t is the crude oil price.

As in BG, we first consider the symmetric ECM approach of Engle and Granger (1987), hereafter EG. The two-step EG approach proceeds as follows. Firstly, the long-run regression in Equation (1) is estimated by OLS, and the residuals are obtained, denoted $\widehat{\varepsilon}_t = y_t - \widehat{\alpha}_0 - \widehat{\alpha}_1 x_t$. These residuals

[7] See: https://www.eia.gov [Last accessed 20/02/17]

[8] In the case of time series data, is is not clear what is meant by "resampling from the same population" as defined for reproduction by Clemens (2017). In the sense that we use gasoline prices over roughly the same time period from different locations, this could be considered a reproduction of the BG analysis.

Figure 1: U.S. Gasoline and Crude Oil Price



Notes: Series are New York Harbor Conventional Gasoline Regular Spot Price (FOB) and the Cushing, OK WTI Spot Price FOB as discussed in the text. The vertical line displays the end of the BG sample.

are then lagged and added to a stationary error correction model. BG consider the following one-lag ECM specification:

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 \widehat{\varepsilon}_{t-1} + \phi_1 \Delta y_{t-1} + \psi_1 \Delta x_{t-1} + v_t \tag{2}$$

They then consider an asymmetric version of the ECM in Equation (2):

$$\Delta y_t = \beta_1^+ \Delta x_t^+ + \beta_1^- \Delta x_t^- + \beta_2^+ \widehat{\varepsilon}_{t-1}^+ + \beta_2^- \widehat{\varepsilon}_{t-1}^- + \phi_1^+ \Delta y_{t-1}^+ + \phi_1^- \Delta y_{t-1}^- + \psi_1^+ \Delta x_{t-1}^+ + \psi_1^- \Delta x_{t-1}^- + v_t \tag{3}$$

where $\Delta x_t^+ = \Delta x_t$ if $\Delta x_t > 0$ and 0 otherwise, and so on for the remaining variables. This specification allows for asymmetric responses, not just in the error correction term $\widehat{\varepsilon}_{t-1}$, but also in the other ECM variables. This differs from the asymmetric ECM we consider later in the paper where only the error correction term $\widehat{\varepsilon}_{t-1}$ enters the model in an asymmetric fashion. Nevertheless, the main coefficients of interest are β_2^+ and β_2^- and specifically whether we can reject the symmetry null hypothesis $H_0 : \beta_2^+ = \beta_2^-$.

Table 1 presents the results of the first table in BG along with the replication results when using the same data span. In attempting to verify the standard errors presented in BG, we also use the Newey and West (1987) HAC standard errors with a lag truncation of $4(T/100)^{2/9}$. This gives us a lag truncation of 8.

Table 1: Replication of Bachmeier & Griffin Error Correction Models

	Bachmeier & Griffin				Replication			
	Symmetric ECM		Asymmetric ECM		Symmetric ECM		Asymmetric ECM	
	Coef	Std Error	Coef	Std Error	Coef	Std Error	Coef	Std Error
β_1	0.775	0.038	-	-	0.778	0.040	-	-
β_1^+	-	-	0.748	0.049	-	-	0.776	0.043
β_1^-	-	-	0.799	0.055	-	-	0.785	0.060
ψ_1	-0.056	0.026	-	-	-0.026	0.031	-	-
ψ_1^+	-	-	0.002	0.048	-	-	0.055	0.061
ψ_1^-	-	-	-0.101	0.042	-	-	-0.094	0.047
ϕ_1	0.145	0.026	-	-	0.095	0.029	-	-
ϕ_1^+	-	-	0.139	0.046	-	-	0.044	0.050
ϕ_1^-	-	-	0.154	0.036	-	-	0.144	0.042
β_2	-0.021	0.004	-	-	-0.021	0.005	-	-
β_2^+	-	-	-0.017	0.007	-	-	-0.018	0.007
β_2^-	-	-	-0.025	0.008	-	-	-0.025	0.007
			Long-run Regression					
α_0	2.970	0.177	2.970	0.177	2.890	0.205	2.890	0.205
α_1	1.063	0.009	1.063	0.009	1.069	0.010	1.069	0.010
							sym	0.563

Notes: The standard errors are Newey and West (1987) HAC standard errors with lag length $4(T/100)^{2/9}$, as discussed in the text. The figure *sym* present the p-value of a Wald test of the null hypothesis of symmetric error correction $H_0 : \beta_2^+ = \beta_2^-$.

Despite being based on as sample which differs by a substantial number of observations to the original data of BG, the results from the replication in the right panel of Table 1 are remarkably similar. The most striking result is that the coefficients and standard errors on the error correction terms, both in the symmetric (β_2) and asymmetric equations (β_2^+ and β_2^-), are almost perfectly replicated with changes only at the third decimal place. We also present a p-value testing the null hypothesis of symmetric adjustment which shows that there is no evidence of asymmetric pricing. This drives the rockets and feathers phenomenon, therefore it is interesting to note that the results of Table 1 of Bachmeier and Griffin (2003) are almost perfectly verified both quantitatively and qualitatively in the sense of Clemens (2017), even with a slight variation on the variable definitions and data period.

Turning to the other results, in the long-run regression we see that the coefficient α_1, known as the pass-through coefficient, is very similar in both sets of results. For α_0, the mark-up coefficient, we see slight differences which may be due to the differing means of the price of gasoline in our sample relative to that of BG. A more complete discussion of the mark-up coefficient is given later in the second case study. The differences in coefficients which appear to be 'large' come from the coefficient ψ_1 in the symmetric equation and for ψ_1^+ and ϕ_1^+ in the asymmetric equation, but this is only because these variables are not significantly different from zero and a close replication on the magnitude of the coefficients cannot be expected unless the data and variables are identical.

3.3 Extended Sample

We now re-run the results of Table 1 to include the extended sample period running up to February 2017. Table 2 presents two sets of results. In the left panel we use the entire sample from June 1986 to February 2017 to perform the analysis. In the right panel, we take only the data after the period considered by BG, namely November 1998 to February 2017. The motivation for using this sub-sample can be seen by looking at Figure 1 where it is apparent that all variables display different behaviour after the BG sample finished. Although our variable definitions are not identical to that of BG, we consider these to be extension in the terminology of Clemens (2017).

The results in Table 2 show that the results of BG are rather robust to changing the sample period. This is perhaps unexpected given the changes in gasoline and crude oil price behaviour since 1998. The coefficient β_2 for the symmetric error correction model is almost identical upon changing the sample; an interesting finding. The asymmetric ECM coefficients β_2^+ and β_2^- change in magnitude slightly and, in fact, in Table 2 we have that $|\beta_2^+| > |\beta_2^-|$ whereas in 1 it was the case that $|\beta_2^+| < |\beta_2^-|$. However, the p-values of the test of symmetry suggest otherwise as we, again, find that there is no evidence to suggest that adjustment is asymmetric. These findings are also robust to the addition of a hurricane Katrina dummy to the model.[9]

The results of Tables 1 and 2 together show that, although we do not use an identical variable specification to that of BG, their results can be almost perfectly replicated when using almost the sample period, and are rather robust when we extend the sample from the 2000s onwards.

[9]The results are not presented here for the sake of brevity, but are available from the authors on request.

Table 2: Extension of Bachmeier & Griffin Sample Period

	Full Sample (1986-2017)				Post-BG Sample (1998-2017)			
	Symmetric ECM		Asymmetric ECM		Symmetric ECM		Asymmetric ECM	
	Coef	Std Error	Coef	Std Error	Coef	Std Error	Coef	Std Error
β_1	0.804	0.053	-	-	0.806	0.059	-	-
β_1^+	-	-	0.740	0.107	-	-	0.732	0.117
β_1^-	-	-	0.875	0.038	-	-	0.889	0.043
ψ_1	-0.009	0.049	-	-	-0.010	0.052	-	-
ψ_1^+	-	-	0.147	0.138	-	-	0.152	0.146
ψ_1^-	-	-	-0.160	0.066	-	-	-0.168	0.071
ϕ_1	0.024	0.056	-	-	0.021	0.059	-	-
ϕ_1^+	-	-	-0.053	0.140	-	-	-0.058	0.148
ϕ_1^-	-	-	0.103	0.057	-	-	0.101	0.060
β_2	-0.021	0.005	-	-	-0.021	0.006	-	-
β_2^+	-	-	-0.020	0.014	-	-	-0.020	0.015
β_2^-	-	-	-0.015	0.011	-	-	-0.016	0.013
	Long-run Regression							
α_0	1.719	0.124	1.719	0.124	3.132	0.265	3.132	0.265
α_1	1.159	0.002	1.159	0.002	1.142	0.004	1.142	0.004
			sym	0.846			*sym*	0.904

Notes: See notes in Table 1.

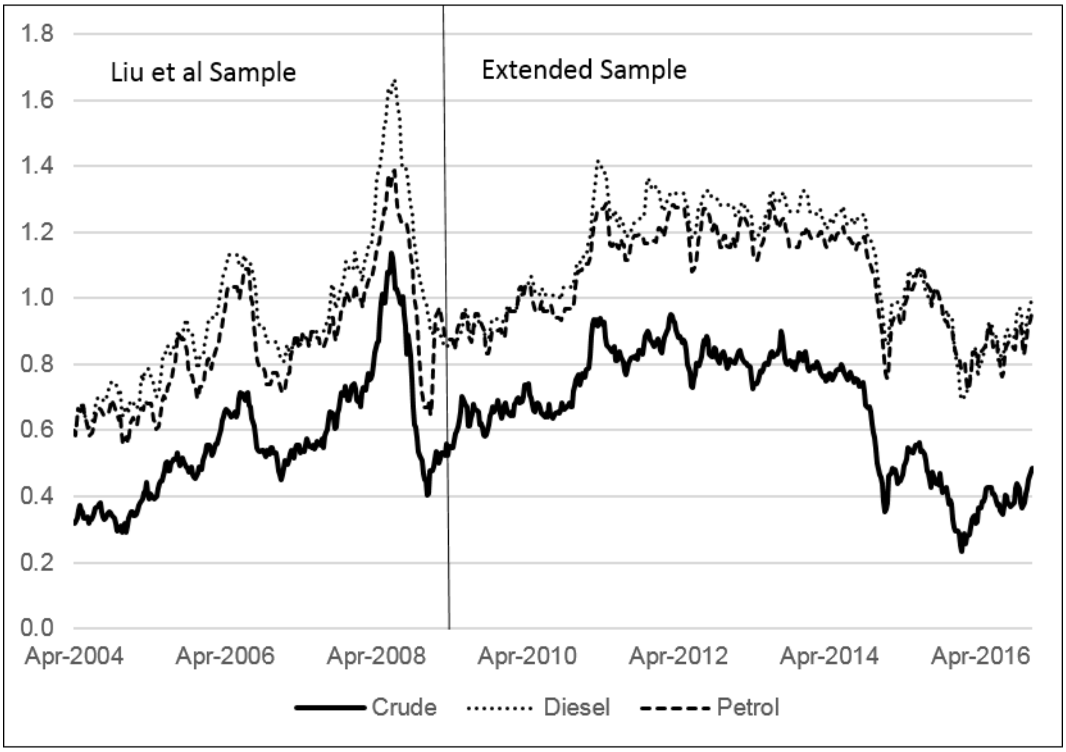
4 Replication Case Study II

In this section we perform a second, more detailed, replication study based on the paper by LMT. They use series for New Zealand petrol and diesel prices excluding taxes (following Equation (1), we generically label petrol or diesel prices as y_t) along with the Dubai crude oil price (x_t), over a period from April 2004 to February 2009. In this paper, we first present the updated dataset in order to observe the differences in the data post-2009. We then proceed to replicate LMT's analysis, which follows the procedures of unit root testing and cointegration and symmetric and asymmetric error correction methods. In the process of implementing a variety of tests and models, LMT consider a number of issues in relation to the properties of petrol and diesel prices with respect to the underlying crude oil price. The notions of mark-up, pass-through and asymmetric adjustment are all raised. These issues motivate the present case study. In each case, the initial findings will be replicated with the original methods applied to the original sample. However, additional improved and more informative methods will also be utilised in order to perform a reanalysis. We will also make use of the extended sample period with the extension relating to additional observations available since the original study.

4.1 Data

We use the same data as LMT which is taken from the Ministry for Economic Development in New Zealand.[10] We obtain weekly data for the periods from April 2004 until February 2017. However, since the final observations are only "provisional" data, we exclude them from our sample and perform analysis only up until the last week of 2016. Figure 2 presents a chart of the data over the full period, with a vertical line displaying where the LMT sample ends.

Figure 2: Diesel, Petrol and Dubai Crude Oil Prices



Notes: Pre-tax diesel and petrol prices, and Dubai crude oil price are all in New Zealand Dollars per litre. The vertical line displays the end of the LMT sample.

The chart shows that all three variables were trending upwards over almost the entire period considered by LMT. There was also a large spike in prices towards the end of their sample, which they mention as leading to "public outcry over petrol pricing by oil companies and demand for more

[10]See: www.med.govt.nz [Last Accessed: 21/02/2017]

regulation by the government" (p.928). On the contrary, when extending the dataset past 2009, we find that pre-tax petrol and diesel prices as well as the Dubai crude oil price appear more stationary. Since the different sample appears to have somewhat different properties, this may affect the results with respect to the extension of the time period. This makes it important to follow the stages of replication and extension also with respect to examining the unit root properties of the series.

4.2 Unit root testing

Unit root tests are performed by LMT at the outset of their analysis, using both the Augmented Dickey Fuller test (ADF, Dickey and Fuller, 1979, 1981) and the Phillips-Perron test (PP, Phillips and Perron, 1988). Given the popularity of the ADF test, and in the interests of brevity, we only report the results for this test. LMT do not state the deterministic components used in the unit root testing equations, nor do they specify the means for optimising or selecting the number of lags with which they augment the test. We therefore present results for the ADF test using alternative deterministics and lag lengths.

In Table 3 the results presented by LMT are provided alongside results obtained for intercept and trend models using the lag lengths specified by LMT (these test statistics are denoted as τ_μ and τ_τ respectively) and under lag optimisation using the modified AIC (the resulting test statistics denoted as τ_μ^{maic} and τ_τ^{maic} respectively). From inspection of the data over the the original sample, consideration of the trend specification appears to be warranted. We note that this analysis relates to Clemens (2017) notions of verification (τ_μ or τ_τ) and re-analysis (τ_μ^{maic} and τ_τ^{maic}) as the former uses the same sample and methodology whereas the latter uses a different methodology.

Table 3: Unit Root Tests

	Liu *et al.*	τ_μ	τ_τ	τ_μ^{maic}	τ_τ^{maic}
Crude Oil	0.427	0.425	0.923	0.430	0.780
Diesel	0.348	0.385	0.616	0.385	0.616
Petrol	0.179	0.179	0.175	0.167	0.175

Notes: The numbers are p-values for the different statistics τ_μ, τ_τ, τ_μ^{maic} and τ_τ^{maic} for testing the null hypothesis of a unit root against the one-sided alternative of stationarity.

To summarise, the results in Table 3 show that qualitatively nothing has changed: the series were considered to be $I(1)$ in the original study and they continue to be $I(1)$ after extending the analysis to include alternative deterministic terms and lag lengths. The original results are not re-produced exactly, but we note that the τ_μ column is indeed very close to the original results. The small differences may potentially be due to the final observations of the sample being subject to revision, which could be different in our vintage of the dataset.

4.3 Long-Run Regressions

To explore the long-run relationships between diesel and petrol prices as a function of the crude oil price, LMT use a simple linear regression of y_t, the diesel/petrol price, and x_t, the Dubai crude oil price (all in natural logarithms). This regression was described earlier in Equation (1). Estimating this equation over the original sample by OLS results in estimates of the static long run parameters α_0 and α_1. These are reported below in Table 4.

Table 4: Long-Run Regressions

	α_0	t_{α_0}	α_1	t_{α_1}	$\overline{R}^2$
			Diesel		
Liu *et al.*	1.34	22.18	0.72	53.06	0.92
Replication	1.38	23.81	0.71	54.29	0.92
			Petrol		
Liu *et al.*	1.41	26.92	0.68	57.67	0.93
Replication	1.45	27.22	0.67	55.65	0.92

Notes: The term t_{α_0} denotes the t-ratio for the coefficient α_0 and likewise for α_1.

The findings of our analysis show a rather close replication in support of the results of LMT, although the results are not identical. Again, the above issue of data revision is the likely cause of these discrepancies. However, in the process of discussing the results from estimation of the long-run model in Equation (1), two further issues are raised by LMT, namely mark-up and pass-through, which were mentioned briefly in the previous case study. To elaborate further, mark-up refers to the non-crude oil price element of diesel and petrol pricing and is given by the constant of the regression, α_0, while pass-through relates to the degree of changes in crude oil prices which impact on diesel and petrol prices and is hence given by α_1. Consideration of these notions raises a number of further issues unconsidered by LMT.

With regard to pass-through, it is well recognised that in the presence of asymmetric adjustment, the estimator of α_1 in the Engle-Granger equation suffers a non-negligible downward bias, see Holly et al. (2003). With regard to mark-up, the scale of the data considered will impact upon the measurement of this quantity. In the original analysis, diesel and petrol are measured in NZ cents per litre, while crude is measured in NZ Dollars per barrel. Therefore the scales of the diesel/petrol and crude series differ. This can be illustrated by the simple equations below which indicate how scaling either in the form of cents/dollars or litres/barrels will impact upon the estimate of the intercept or mark-up coefficient. In a regression of $ln\left(Y_t\right)$ on $ln\left(X_t\right)$, rescaling Y_t and X_t by a and b respectively[11] has the following effect:

$$ln\left(\frac{Y_t}{a}\right) = \alpha + \beta\ ln\left(\frac{X_t}{b}\right) + \nu_t \tag{4}$$

$$\implies ln\left(Y_t\right) = \left[\alpha + ln\left(a\right) - \beta\ ln\left(b\right)\right] + \beta\ ln\left(X_t\right) + \nu_t \tag{5}$$

Hence, following the scaling the regression has a new constant term $\alpha^* = \left[\alpha + ln\left(a\right) - \beta ln\left(b\right)\right]$.

The issue of bias in the estimation of the equilibrium coefficients (α_0, α_1) in the static representation in Equation (1) can be remedied using NLS along the lines of Holly et al. (2003). This was not considered by LMT, and could provide more appropriate estimates of these parameters. The estimation of symmetric and asymmetric error correction models (denoted as ECM and AECM respectively) under non-linear least squares (NLS), and the insights it provides in relation to mark-up and pass-through, are discussed below following consideration of the ECM and AECM specifications utilised by LMT.

4.4 Symmetric Error Correction Models

In their paper, LMT first look at the case of symmetric adjustment using the standard EG-ECM. In specifying their ECM, they do not allow for lagged changes in x_t and y_t to enter the equation as in Equation (2). However, the use of lagged changes is rather more common in the literature for improving dynamic specification and, amongst other things, mitigating serial correlation. Therefore, in this section we will re-analyse this problem by allowing further lags to enter the equation as follows:

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 \widehat{\varepsilon}_{t-1} + \sum_{i=1}^{p} \phi_j \Delta y_{t-i} + \sum_{j=1}^{q} \psi_i \Delta x_{t-j} + v_t \tag{6}$$

where similarly to above, y_t is the petrol/diesel price and x_t is the crude oil price, and where the lag lengths p and q on the terms $\{\Delta y_{t-i}\}_{i=1}^{p}, \{\Delta x_{t-j}\}_{j=1}^{q}$ are to be chosen. The LMT specification therefore corresponds to the case where $p = 0$ and $q = 0$. Note that we did not present further lag specifications in the BG replication as they explicitly allowed lags to enter the ECM.[12]

[11] For example, if the petrol and diesel price series are measured in NZc per litre and crude in NZD per barrel, transformation of the series to a consistent scale of NZD per litre results in $a = 100$ and $b = 159$.

[12] We did some further experimentation with the lags considered in the BG results in Tables 1 and 2 however we found the results to be very insensitive to reasonable lag length differences. The results are therefore omitted but are available upon request to the interested reader. We thank an anonymous referee for suggesting that we make this additional check in the BG replication.

We will also compare the EG approach to a more robust approach of Holly et al. (2003) which instead inserts the lagged error term ε_{t-1} into the ECM instead of the residuals $\widehat{\varepsilon}_{t-1}$ and jointly estimates all of the parameters in the following expression:

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 \left(y_{t-1} - \alpha_0 - \alpha_1 x_{t-1} \right) + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \sum_{j=1}^{q} \psi_i \Delta x_{t-j} + v_t \tag{7}$$

using nonlinear least squares (NLS).[13]

For now, we focus only on the two-step method in Equation (6). Since LMT do not include any lags of Δy_t and Δx_t in their ECM specifications, we use our data to estimate both static and dynamic specifications of Equation (6), and then compare the results obtained to those of LMT. In the terminology of Clemens (2017), we are attempting to replicate the results of LMT while also allowing a reanalysis using the same data. We will additionally perform some misspecification tests to examine the validity of the models considered.

The results of LMT are presented in Table 5 along with the results of re-estimation of the two-step ECM method in Equations (1) and (6) with and without lagged differenced regressors included in the ECM. The absence of lagged differenced terms is denoted by $(p, q) = (0, 0)$ while estimation with these terms included is denoted by non-zero values for (p, q). In the latter instances the actual values are determined via minimisation of the AIC, as is standard practice in the literature. From inspection for Table 5, a close correspondence between the results of LMT and the static models with $(p, q) = (0, 0)$ is apparent. However, the Lagrange Multiplier (LM) serial correlation tests indicate that the static ECMs for both petrol and diesel suffer from serial correlation and are therefore misspecified which would make resulting analysis invalid. Re-specifying the dynamics of the ECMs using the AIC results in $ADL(6, 3)$ and $ADL(3, 3)$ specifications for diesel and petrol respectively. While the LM tests show this overcomes the problem of serial correlation, heteroskedasticity is still apparent in the petrol equation. Consequently, t-ratios are re-calculated using White (1980) standard errors.

Overall, the results indicate that the static models considered in LMT can be improved upon, and that consideration of misspecification analysis and dynamic specification leads to the generation of better fitting models. These models, however, do have less significant error correction coefficients when robust inference is performed. This does not alter or reverse the conclusions of LMT in any qualitative way.

[13] Note that there is an element of uncertainty in LMT, p.928, where they define the residual series as $\widehat{\varepsilon}_t = y_t - \alpha_0 - \alpha_1 x_t$ in terms of the population parameters α_0 and α_1 instead of their estimates $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$. However it is clear that they are, in fact, running a two-step EG method and do not estimate an Equation like (7) by NLS.

Table 5: Symmetric ECM (Engle-Granger two-step)

	p, q	β_1	t_{β_1}	β_2	t_{β_2}	$\overline{R}^2$	lm_5	lm_{10}	het
					Diesel				
Liu *et al.*	$*$	0.15	4.47	-0.16	-8.07	0.28	$*$	$*$	$*$
Replication	0, 0	0.17	4.78	-0.16	-7.62	0.27	0.00	0.00	0.31
Reanalysis	6, 3	0.01	3.15	-0.07	-3.22	0.47	0.75	0.96	0.12
			(2.49)		(-2.44)				
					Petrol				
Liu *et al.*	$*$	0.22	5.05	-0.19	-6.42	0.24	$*$	$*$	$*$
Replication	0, 0	0.26	5.76	-0.16	-5.39	0.21	0.00	0.00	0.00
Reanalysis	3, 3	0.15	3.80	-0.13	-4.35	0.49	0.56	0.12	0.00
			(2.72)		(-3.77)				

Notes: The above tabulated values are obtained from the second stage of the two-step estimation of Equations (1) and (6). Diagnostic tests: lm_i denotes p-values for the LM test of i^{th} order residual autocorrelation; *het* denotes p-values for White's test of heteroskedasticity including cross-products. Figures in parentheses are t-ratios obtained using White standard errors. $*$ denotes information unavailable from LMT.

4.5 Asymmetric Error Correction Models

The way LMT consider asymmetric adjustment of petrol and diesel prices is to extend the symmetric EG-ECM to allow for asymmetries in a similar way to the first case study.[14] The first step is still to estimate the long-run Equation (1) and save the residuals. However, in the next step, the asymmetry in the error correction mechanism is developed using a 0-1 Heaviside indicator function λ_t. The resulting two-step and non-linear AECMs extending Equations (6) and (7) are given as follows:

$$\Delta y_t = \delta_1 \Delta x_t + \delta_2 \widehat{\lambda}_t \widehat{\varepsilon}_{t-1} + \delta_3 \left(1 - \widehat{\lambda}_t\right) \widehat{\varepsilon}_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \sum_{j=1}^{q} \psi_i \Delta x_{t-j} + v_t \tag{8}$$

[14]This is done in many papers in the rockets and feathers literature. Fosten (2012), on the other hand, uses a threshold vector error correction model as proposed by Hansen and Seo (2002).

$$\Delta y_t = \delta_1 \Delta x_t + \delta_2 \lambda_t (y_{t-1} - \alpha_0 - \alpha_1 x_{t-1}) + \delta_3 (1 - \lambda_t) (y_{t-1} - \alpha_1 - \alpha_2 x_{t-1}) + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \sum_{j=1}^{q} \psi_i \Delta x_{t-j} + v_t \tag{9}$$

where $\widehat{\lambda}_t = 1$ when $\widehat{\varepsilon}_{t-1} > 0$ and zero otherwise in Equation (8), and $\lambda_t = 1$ when $(y_{t-1} - \alpha_0 - \alpha_1 x_{t-1})$ > 0 and zero otherwise in Equation (9).

Focussing on the two-step AECM in Equation (8), the results of LMT are presented in Table 6 below. The results for the AECMs reflect those for the analogous symmetric specifications in Table 5 in that the static models with $(p, q) = (0, 0)$ produce results very close to those of LMT. This demonstrates that their results can be verified to a high degree of accuracy. However, these static models again are found to be misspecified and hence dynamic models are estimated to perform a robustness check through reanalysis, and also correcting for heteroskedasticity using White standard errors. This yields two prominent features. First, the finding of significant asymmetry for the diesel equation found in the LMT specification is reversed with the symmetry test for the dynamic, corrected model which has a p-value of 22% compared to the value of 5.4% in LMT. Second, the estimated asymmetry coefficients follow the findings of LMT with regard to their relative sizes, but are larger (less negative) in value and less significant.

4.6 Reanalysing Mark-up

As mentioned above, there are two problems with LMT's assertion that the estimated equilibrium intercept, α_0, of the static Engle-Granger cointegrating regression can be considered as a measure of mark-up. Firstly, the scales or measurements of the series considered will impact upon the estimate obtained, as noted in Equations (4) and (5). Secondly, research such as Holly et al. (2003) notes the bias associated with the coefficient estimates of the Engle-Granger equation.

In response to these issues, we reanalyse the mark-up measures by comparing those calculated using the Engle-Granger equation to those obtained by estimating Equations (7) and (9); the symmetric ECM estimated by NLS and the asymmetric AECM estimated by NLS. These are undertaken using the original scales of the series (NZc per litre for diesel and petrol price and NZD per barrel for crude oil price) and then using a common scale for all series (NZD per litre for each of diesel, petrol and crude oil price). The results are presented in Table 7. We firstly note that the mark-ups of LMT are well replicated when we use the same scalings and method.

However, when we reexamine the methodology and data scaling, a different picture emerges with noticeable impacts on the measurement of mark-up. Using the original NZc per litre and NZD per barrel scales, this is illustrated by the 8.97% decrease in the estimated value of mark-up for petrol

Table 6: Asymmetric ECM (Engle-Granger two-step)

p, q	δ_1	t_{δ_1}	δ_2	t_{δ_2}	δ_3	t_{δ_3}	sym	$\overline{R}^2$	lm_5	lm_{10}	het
					Diesel						
					Liu *et al.*						
$*$	0.15	4.50	−0.13	−5.48	−0.21	−6.35	0.05	0.29	$*$	$*$	$*$
					Replication and Reanalysis						
0, 0	0.17	4.78	−0.13	−5.12	−0.20	−5.93	0.13	0.27	0.00	0.00	0.27
6, 3	0.10	3.14	−0.05	−1.98	−0.10	−3.19	0.16	0.47	0.71	0.90	0.08
		(2.43)		(−1.33)		(−3.50)	0.22				
					Petrol						
					Liu et al.						
$*$	0.22	5.07	−0.17	−3.92	−0.21	−5.19	0.43	0.23	$*$	$*$	$*$
					Replication and Reanalysis						
0, 0	0.26	5.79	−0.14	−3.19	−0.18	−4.43	0.44	0.21	0.00	0.00	0.00
3, 4	0.15	3.93	−0.10	−2.42	−0.14	−3.85	0.38	0.50	0.76	0.17	0.00
		(2.83)		(−2.01)		(−3.43)	0.44				

Notes: See notes in Table 5. Additionally, *sym* denotes the p-values of an F-test for symmetry H_0: $\delta_2 = \delta_3$.

under the AECM-NLS model relative to that obtained under application of the static Engle-Granger regression. Importantly, as noted above, this latter model is the more robust specification. Further to this, an increase in the estimated value of mark-up of 6.74% is observed for diesel in analogous circumstances. The decrease for diesel and increase for petrol illustrates the variation in estimated mark-up when broadening the analysis to consider more robust, or improved, specifications. The arguments for the initial scaling hold when employing NZD per litre scaling for the series with a decrease of 14.29% noted for diesel. However, the variation in mark-up is obviously most noticeable when moving between scales with the use of the AECM-NLS model under the second scaling (NZD per litre for all series) leading to decreases in mark-up of 77.4% and 84.1% for diesel and petrol respectively relative to application of the Engle-Granger approach under the NZc per litre and NZD per barrel measures.

Table 7: Measures of Mark-up

	Diesel	Petrol
Liu *et al.*	1.34	1.41
Petrol, diesel: NZc/litre; Crude: NZD/barrel		
EG	1.38	1.45
ECM-NLS	1.37	1.48
AECM-NLS	1.47	1.32
Petrol, diesel, crude: NZD/litre		
EG	0.36	0.24
NLS-ECM	0.37	0.23
NLS-AECM	0.31	0.23

Notes: All significant at 1% level.

4.7 Reanalysing Pass-through effects

We next re-analyse the pass-through coefficients found in LMT, as these may be similarly biased in the presence of asymmetries, as discussed in the previous section. Again, the NLS ECM and NLS AECM are employed to address this issue. The results obtained are displayed in Table 8. In all cases the null of complete pass-through is rejected at the 1% level. Upon reanalysis of the pass-through coefficients of LMT re-estimated using the NLS AECM method, we find some small differences which may be due to biases in the original methodology. The diesel pass-through coefficient decrease relative to the original study, whereas the petrol pass-through coefficient increases.

Table 8: Measures of Pass-through

	Diesel	Petrol
Liu *et al.*	0.72	0.68
EG	0.71	0.67
NLS-ECM	0.71	0.66
NLS-AECM	0.68	0.69

4.8 Reanalysing Asymmetric Adjustment

We finally compare the estimates of the asymmetric adjustment coefficients when using the NLS method of Holly et al. (2003), as this can be useful in remedying the low power of AECMs (see also Cook et al., 1999). Table 9 below considers the estimated asymmetry coefficients and associated test for asymmetry presented by LMT and from estimation of two-step and NLS AECMs. The AECMs are considered in both their static and dynamic forms as in Table 6 above. This is due to the dynamic misspecification identified in the static model.

The results of Table 9 again confirm the above finding that the significant asymmetry (p-value = 5.4%) for diesel found in the paper of LMT appears to be invalid due to misspecification. The fact that this occurs both in the two-step AECM and also in that estimated by NLS lends a robustness to this finding, over and above that in LMT. Once the equations are corrected for dynamic misspecification, the asymmetry is not apparent at conventional significance levels for diesel with a p-value of 11.8%. This indicates that further work into the rockets and feathers phenomenon should be careful to consider lag length selection as this is found to have an impact on the decision to reject symmetry or otherwise.

Table 9: Asymmetric adjustment

	Diesel					Petrol				
	δ_2	t_{δ_2}	δ_3	t_{δ_3}	sym	δ_2	t_{δ_2}	δ_3	t_{δ_3}	sym
Liu et al.										
	−0.13	−5.48	−0.21	−6.35	0.05	−0.17	−3.92	−0.21	−5.19	0.43
AECM Two-step										
Static	−0.13	−5.12	−0.20	−5.93	0.13	−0.14	−3.19	−0.18	−4.43	0.44
Dynamic	−0.05	−1.33	−0.10	−3.50	0.22	−0.10	−2.01	−0.14	−3.43	0.44
AECM NLS										
Static	−0.11	−4.37	−0.26	−3.70	0.08	−0.08	−2.27	−0.28	−6.12	0.09
Dynamic	−0.04	−1.96	−0.21	−1.99	0.12	−0.08	−1.80	−0.20	−2.62	0.20

Notes: See notes in Tables 5 and 6.

4.9 Extended Sample and Reanalysis Using NARDL Model

In the previous analysis, we kept the same sample as LMT and performed replication by verification and then robustness check by reanalysing the same data using different methodologies. Since we now have a longer sample available, in this section we will also consider the notion of extension as in Clemens (2017). The weekly data we use runs from April 2004 to December 2016. As mentioned before, we also have preliminary data for some weeks in 2017, but prefer not to contaminate the sample with these data which are likely to be revised.

We start by applying unit root tests by extending the sample used to produce the results in Table 3. The results for τ_{μ}^{maic} and τ_{τ}^{maic} are displayed in Table 10.

Table 10: Unit root tests

	2004-2016	
	τ_{μ}^{maic}	τ_{τ}^{maic}
Crude Oil	0.178	0.499
Diesel	0.030	0.189
Petrol	0.016	0.059

Notes: See notes for Table 3.

The results in Table 10 cast some doubt over the original findings of LMT that all series are unit root $I(1)$ processes. In the full sample case, judging from the chart in Figure 2, it is less clear that a deterministic trend is required in the test. Furthermore, the p-value corresponding to τ_{μ}^{maic} for the intercept only ADF test, reject the null hypothesis of a unit root for both diesel and petrol prices (although not for crude oil).

This change in inference concerning the order of integration is an interesting issue which warrants some discussion.[15] Following Shiller and Perron (1985) it has long been recognised that the sample span, or calendar period, employed in empirical analysis rather than the number of observations *per se* has a significant effect upon the power of unit root tests. This issue was subsequently demonstrated by Pierse and Snell (1995). Considering the present analysis, it is therefore possible that an increase in power as a result of extending the sample to 2016 has permitted the rejection of the unit root null hypothesis for the diesel and petrol price series. Alternatively, it may be that diesel and petrol prices have experienced genuine changes in persistence. Recognition of the possibility of such switching in the order of integration of time series processes has resulted in the development of alternative tests to

[15]We are grateful to an anonymous referee for raising this issue.

establish its presence (see, inter alia, Harvey et al., 2006; Leybourne et al., 2007) and received empirical support in the examination of a number of series such as U.S. inflation (Noriega and Ramos-Francia, 2009; Fuhrer, 2010), historical unemployment (Fosten and Ghoshray, 2011) and real interest rates (Soon et al., 2017). While not the focus of the current analysis, a closer examination of the integrated nature and persistence of diesel and petrol prices is an interesting avenue for future research.

With this in mind, since the extension of the sample period introduces $I(0)$ variables, we no longer us the Engle-Granger two-step approach. Instead, we consider the NARDL model of Shin et al. (2014) to explore asymmetric adjustment as it is applicable to both $I(0)$ and $I(1)$ processes. In the terminology of Clemens (2017), this analysis constitutes robustness both in terms of extension and reanalysis, as we are changing the sample and the methodology. The general NARDL model is given as:

$$\Delta y_t = \lambda + \rho y_{t-1} + \delta^+ x_{t-1}^+ + \delta^- x_{t-1}^- + \sum_{i=1}^{q-1} \gamma_i \Delta y_{t-1} + \sum_{i=0}^{p-1} \pi_i^+ \Delta x_{t-1}^+ + \sum_{i=0}^{p-1} \pi_i^- \Delta x_{t-1}^- + \nu_t \tag{10}$$

where:

$$x_t^+ = \sum_{k=1}^{t} \Delta x_k^+ = \sum_{k=1}^{t} max\left(\Delta x_k, 0\right) \tag{11}$$

$$x_t^- = \sum_{k=1}^{t} \Delta x_k^- = \sum_{k=1}^{t} min\left(\Delta x_k, 0\right) \tag{12}$$

Application of the NARDL model of (10) is of interest not just as it allows us to work with series that might be either $I(1)$ or $I(0)$, but also allows examination of potential asymmetry in alternative forms. Using the long-run parameters, $\theta^+ = \delta^+/\rho$ and $\theta^- = \delta^-/\rho$, asymmetry over the longer term can be considered via examination of the hypothesis of their equality. Similarly, asymmetry over the short-run can be explored via the (equality of the) coefficients π_i^+ and π_i^-. In addition, considering the equality of sums of these coefficients, the impact multipliers $\left(\pi_0^+, \pi_0^-\right)$ can be utilised to explore an asymmetric effect of a more immediate nature. Finally, pass-through of an asymmetric nature can be considered via examination of whether (θ^+, θ^-) are equal to 1.

Before considering the results presented in Table 11, it should be noted that these findings involve the use of White's (1980) heteroskedasticity robust standard errors as a result of detection of heteroskedasticity.[16] Considering the results in Table 11 for the original sample, evidence of long-run

[16] All estimated models were subjected to diagnostic testing in the form of LM testing for serial correlation (using 5

Table 11: NARDL results

	Original sample (April 2004-Feb 2009)				Extended sample (Feb 2009-Dec 2016)			
	Diesel		Petrol		Diesel		Petrol	
ρ	-0.081	[0.002]	-0.117	[0.001]	-0.032	[0.003]	-0.068	[0.000]
θ^+	0.747	[0.000]	0.562	[0.000]	0.668	[0.000]	0.556	[0.000]
θ^-	0.764	[0.000]	0.516	[0.000]	0.649	[0.000]	0.521	[0.000]
H(θ^+)	[0.001]		[0.000]		[0.000]		[0.000]	
H(θ^-)	[0.014]		[0.000]		[0.000]		[0.000]	
π_0^+	0.088	[0.268]	-0.004	[0.972]	0.193	[0.000]	0.180	[0.007]
π_0^-	0.127	[0.054]	0.308	[0.000]	0.153	[0.000]	0.210	[0.000]
$\sum_{i=1}^{p-1} \pi_i^+$	0.223	[0.000]	0.532	[0.001]	0.369	[0.000]	0.375	[0.000]
$\sum_{i=1}^{p-1} \pi_i^-$	0.152	[0.019]	0.272	[0.033]	0.289	[0.000]	0.318	[0.001]
$\theta^+ = \theta^-$	0.543		0.199		0.041		0.000	
$\pi_0^+ = \pi_0^-$	0.747		0.083		0.585		0.768	
$\sum_{i=1}^{p-1} \pi_i^+ = \sum_{i=1}^{p-1} \pi_i^-$	0.448		0.208		0.358		0.680	
PSS	3.706		5.236		3.222		5.305	
BDM	-3.203		-3.463		-2.983		-3.874	

Notes: The above table contains results obtained from estimation of the NARDL model of Equation (10). ρ, θ^+, θ^-, π_0^+, π_0^-, $\Sigma_{i=1}^{p-1}\pi_i^+$, $\Sigma_{i=1}^{p-1}\pi_i^-$ denote estimated coefficients with p-values of tests of their significance provided in square brackets. H(θ^+) and H(θ^-) denote p-values of tests for long-run pass-through, i.e. $\theta^+ = 1$; $\theta^- = 1$. $\theta^+ = \theta^-$, $\pi_0^+ = \pi_0^-$ and $\Sigma_{i=1}^{p-1}\pi_i^+ = \Sigma_{i=1}^{p-1}\pi_i^-$ denote hypothesis tests of long-run, impact and short-run asymmetry. PSS and BDM denote calculated test statistics of Pesaran et al. (2001) and Banerjee et al. (1998) respectively. The 10%, 5% and 1% critical values for the PSS test are $\{4.04, 4.78\}$, $\{4.94, 5.73\}$ and $\{6.84, 7.84\}$, while the corresponding values for the BDM test are -2.89, -3.19 and -3.78.

relationships between crude prices and both petrol and diesel are apparent from the PSS and BDM tests. However, complete pass-through is rejected for both petrol and crude, thus supporting the

and 10 lags) and White's test of heteroskedasticity. While the null of no serial correlation was not rejected, the null for White's test was rejected for all models. Further details on diagnostic testing are available from the authors upon request.

earlier findings using alternative methods, and little evidence of asymmetry is detected. With regard to this latter issue, the one exception to this concerns 'impact' asymmetry for petrol where the null of symmetry is rejected at the 10% level. Turning to the results for the extended sample, the prominent finding concerns the detection of significant long-run asymmetry for both petrol and diesel. This finding contrasts to the results of LMT where asymmetry was detected for diesel alone. It also contrasts with Alom and Ritson (2012) who find asymmetry only for petrol, although their study does not attempt to first replicate the LMT results and has a shorter time span than our own data. Further to this, the finding of asymmetry is both more significant for petrol and differs from the findings of LMT in that greater adjustment is detected for positive, rather than negative changes.

The results in Table 11 show a clear difference in findings over the original (LMT) and extended sample periods. Prompted by this, Table 12 presents results obtained from the application of the NARDL model to the post-LMT (Feb 2009 - Dec 2016) sample, the period distinguishing the original and extended samples.[17] Considering long-run asymmetry, this is detected for both diesel and petrol prices in the post-LMT sample which is in keeping with its absence in the results for the original sample but appearance in the extended sample. That is, the long-run asymmetry in the extended sample is driven by its presence in the post-LMT period. Turning to short-run asymmetry, the findings for diesel prices are marginally significant (p-value = 5.2%) while no evidence of short-run asymmetry is detected in petrol prices. This is the mirror-image of the findings for the original sample where (marginally) significant results were obtained for petrol, rather than diesel, prices. Again the marginal significance for different series in the two sub-samples provides an explanation for the absence of short-run asymmetry over the extended sample. Finally, the rejection of complete pass-through for the post-LMT sample matches the findings for the original and extended samples. However, while the findings for pass-through are consistent across all samples, the additional results for the post-LMT period provide further support for the above arguments concerning the sample-dependence of findings.

5 Concluding remarks

This paper has revisited the rockets and feathers literature in the context of replication in empirical economics. We firstly provide a review of the literature on replication in economics and other disciplines. This adds to the recent work of Clemens (2017) who proposes a framework for classifying replication in the sphere of economics. We consider both recent and historical research and find, in a similar way to Clemens (2017), that many papers in economics have been unsuccessful in performing

[17]We are grateful to an anonymous referee for the suggestion that behaviour over this sample should be considered.

Table 12: NARDL results for post-LMT sample (Feb 2009 - Dec 2016)

	Diesel		Petrol	
ρ	-0.086	[0.000]	-0.068	[0.000]
θ^{+}	0.823	[0.000]	0.710	[0.000]
θ^{-}	0.750	[0.000]	0.642	[0.000]
$\mathrm{H}(\theta^{+})$	[0.001]		[0.000]	
$\mathrm{H}(\theta^{-})$	[0.000]		[0.000]	
π_0^{+}	0.279	[0.000]	0.247	[0.000]
π_0^{-}	0.160	[0.000]	0.195	[0.000]
$\sum_{i=1}^{p-1} \pi_i^{+}$	0.324	[0.004]	0.248	[0.019]
$\sum_{i=1}^{p-1} \pi_i^{-}$	0.149	[0.276]	0.180	[0.122]
$\theta^{+} = \theta^{-}$	0.000		0.000	
$\pi_0^{+} = \pi_0^{-}$	0.052		0.505	
$\sum_{i=1}^{p-1} \pi_i^{+} = \sum_{i=1}^{p-1} \pi_i^{-}$	0.390		0.692	
PSS	6.958		5.368	
BDM	-4.248		-3.547	

Notes: See notes in Table 11.

replication. We then turn the focus on reviewing the rockets and feathers literature with replication in mind. This reveals that the growing body of empirical work in this field tends not to perform a replication using the definition of Clemens (2017). Papers mostly use different samples of data by choosing different time spans or different countries to previous studies. This means they can only be considered a robustness check rather than replication.

In the main part of the paper, we take two contrasting replication case studies, which highlight different challenges faces by economists wishing to replicate empirical research. The first case study focusses on Bachmeier and Griffin (2003) (BG) which uses daily gasoline and crude oil price data in the U.S. We find that the data for the original paper was not easily available; something which links to the discussion in Winfree (2010) of the difficulties faced when undertaking replication. In spite of this, we are able to replicate the BG results closely.

The second case study looks at asymmetric pricing in diesel and petrol prices in New Zealand, following the paper of Liu et al. (2010) (LMT). While we are able to replicate their results using the same data and methodology, we find that simple reanalysis using different dynamic lag specifications can reverse certain conclusions about symmetry/asymmetry. We also find that extending the time period of analysis can change conclusions about the $I(1)/I(0)$ properties of the data, making it unclear whether the original methods are even valid upon extending the dataset. A reanalysis using the longer time span and the more recent NARDL methodology finds that the conclusions of LMT are reversed, and that asymmetric pricing is found in the petrol market. This sample-dependence of the results was explored further via consideration of behaviour over the post-LMT sample, with differing results obtained relative to the original sample. The fact that the original results may not be robust to variations in the sample period is a feature of replication studies discussed in Makel and Plucker (2014) and McNeeley and Warner (2015).

The contrasting results and conclusions of our two case studies indicate the merit and importance of undertaking replications in economics. On the one hand, our ability to show replicability and robustness of the BG paper shows that some results are able to stand the test of time. However our second case study shows that, even though valuable studies like LMT are interesting to economists in their own right, we must be careful in placing too much weight on the results of individual empirical studies. As a guide for future replications of rockets and feathers papers, researchers should be careful to assess the robustness of results, not only to the use of more recent and appropriate econometric methodologies but also in considering model lag specifications such as in the error correction models frequently employed in this literature. They should also separately consider the changes in the results which come from using data in the “post” sample after the end-date of the replicated study, as well as the results when pooling all of the data in the pre and post sample. This can explain features such as changes in the persistence of different series which may determine replication or non-replication; something we experienced in our second replication.

There is clearly a lot of future work to be done in this area. More widespread replications of the growing body of work on the rockets and feathers phenomenon is to be encouraged, noting that smaller findings and results could be submitted to the Replication Wiki website mentioned above. More generally, promoting the increased use of replication in economics is a challenge left to future researchers and students.

References

Alom, F. and N. Ritson (2012). Asymmetric adjustment of diesel or petrol retail prices to crude oil price movements: New Zealand evidence. *OPEC Energy Review 36*(2), 230–245.

Anderson, R. G. and A. Kichkha (2017). Replication, Meta-Analysis, and Research Synthesis in Economics. *American Economic Review 107*(5), 56–59.

Bachmeier, L. J. and J. M. Griffin (2003). New evidence on asymmetric gasoline price responses. *Review of Economics and Statistics 85*(3), 772–776.

Bacon, R. W. (1991). Rockets and feathers: the asymmetric speed of adjustment of UK retail gasoline prices to cost changes. *Energy Economics 13*(3), 211–218.

Banerjee, A., J. Dolado, and R. Mestre (1998). Error-correction mechanism tests for cointegration in a single-equation framework. *Journal of Time Series Analysis 19*(3), 267–283.

Bermingham, C. and D. O'Brien (2011). Testing for asymmetric pricing behaviour in Irish and UK petrol and diesel markets. *The Energy Journal 32*(3), 1–26.

Berry, J., L. C. Coffman, D. Hanley, R. Gihleb, and A. J. Wilson (2017). Assessing the Rate of Replication in Economics. *American Economic Review 107*(5), 27–31.

Blair, B. F., R. C. Campbell, and P. A. Mixon (2017). Price Pass-Through in US Gasoline Markets. *Energy Economics 65*, 42–49.

Borenstein, S., A. C. Cameron, and R. Gilbert (1997). Do gasoline prices respond asymmetrically to crude oil price changes? *The Quarterly Journal of Economics 112*(1), 305–339.

Chang, A. C., P. Li, and Others (2017). A Preanalysis Plan to Replicate Sixty Economics Research Papers That Worked Half of the Time. *American Economic Review 107*(5), 60–64.

Chua, C. L., C. De Silva, and S. Suardi (2017). Do petrol prices increase faster than they fall in market disequilibria? *Energy Economics 61*, 135–146.

Clemens, M. A. (2017). The meaning of failed replications: A review and proposal. *Journal of Economic Surveys 31*(1), 326–342.

Coffman, L. C., M. Niederle, and A. J. Wilson (2017). A Proposal to Organize and Promote Replications. *American Economic Review 107*(5), 41–45.

Cook, S. (2000). A neglected controversy in the modelling of consumers' expenditure. *Cambridge Journal of Economics 24*(2), 177–191.

Cook, S., S. Holly, and P. Turner (1999). The power of tests for non-linearity: the case of Granger-Lee asymmetry. *Economics Letters 62*(2), 155–159.

Davis, T. E. (1952). The consumption function as a tool for prediction. *The Review of Economics and Statistics 34*(3), 270–277.

De Marchi, N. and C. Gilbert (1989). *History and Methodology of Econometrics.* Oxford Economic Papers Special Issue. Oxford: Clarendon Press.

Delpachitra, S. B. (2002). Price rigidity in the downstream petroleum industry in New Zealand: where does it happen? *Energy Economics 24*(6), 597–613.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*(366), 427–431.

Dickey, D. A. and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica 49*(4), 1057–1072.

Engle, R. F. and C. W. J. Granger (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica 55*(2), 251–276.

Fosten, J. (2012). Rising household diesel consumption in the United States: A cause for concern? Evidence on asymmetric pricing. *Energy Economics 34*(5), 1514–1522.

Fosten, J. and A. Ghoshray (2011). Dynamic persistence in the unemployment rate of oecd countries. *Economic Modelling 28*(3), 948–954.

Frank, M. C. and R. Saxe (2012). Teaching replication. *Perspectives on Psychological Science 7*(6), 600–604.

Frey, G. and M. Manera (2007). Econometric models of asymmetric price transmission. *Journal of Economic Surveys 21*(2), 349–415.

Fuhrer, J. (2010). Inflation persistence. In B. Friedman and M. Woodford (Eds.), *Handbook of Monetary Economics*, Volume 1, pp. 423–486. Amsterdam: Elsevier B.V.

Gilbert, C. L. (1986). Professor Hendry's Econometric Methodology. *Oxford Bulletin of Economics and Statistics 48*(3), 283–307.

Gilbert, C. L. (1989). LSE and the British approach to time series econometrics. *Oxford Economic Papers 41*(1), 108–128.

Gilbert, C. L. (1991). Richard Stone, demand theory and the emergence of modern econometrics. *The Economic Journal 101*(405), 288–302.

Hansen, B. E. and B. Seo (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics 110*(2), 293–318.

Harvey, D. I., S. J. Leybourne, and A. R. Taylor (2006). Modified tests for a change in persistence. *Journal of Econometrics 134*(2), 441–469.

Höffler, J. H. (2017). Replication and economics journal policies. *American Economic Review 107*(5), 52–55.

Holly, S., P. Turner, and M. Weeks (2003). Asymmetric adjustment and bias in estimation of an equilibrium relationship from a cointegrating regression. *Computational Economics 21*(3), 195–202.

Kristoufek, L. and P. Lunackova (2015). Rockets and feathers meet Joseph: Reinvestigating the oilgasoline asymmetry on the international markets. *Energy Economics 49*, 1–8.

Lahiani, A., A. Miloudi, R. Benkraiem, and M. Shahbaz (2017). Another look on the relationships between oil prices and energy prices. *Energy Policy 102*, 318–331.

Leybourne, S., R. Taylor, and T.-H. Kim (2007). Cusum of squares-based tests for a change in persistence. *Journal of Time Series Analysis 28*(3), 408–433.

Lipsey, R. G. (1960). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1862-1957: a further analysis. *Economica 27*(105), 1–31.

Liu, M.-H., D. Margaritis, and A. Tourani-Rad (2010). Is there an asymmetry in the response of diesel and petrol prices to crude oil price changes? Evidence from New Zealand. *Energy Economics 32*(4), 926–932.

Makel, M. C. and J. A. Plucker (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher 43*(6), 304–316.

Manning, D. N. (1991). Petrol prices, oil price rises and oil price falls: some evidence for the UK since 1972. *Applied Economics 23*(9), 1535–1541.

McNeeley, S. and J. J. Warner (2015). Replication in criminology: A necessary practice. *European Journal of Criminology 12*(5), 581–597.

Morgan, M. S. (1990). *The History of Econometric Ideas*. Historical Perspectives on Modern Economics. Cambridge: Cambridge University Press.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–708.

Noriega, A. E. and M. Ramos-Francia (2009). The dynamics of persistence in us inflation. *Economics Letters 105*(2), 168–172.

Perdiguero-García, J. (2013). Symmetric or asymmetric oil prices? A meta-analysis approach. *Energy Policy 57*, 389–397.

Pesaran, M. H., Y. Shin, and R. J. Smith (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics 16*(3), 289–326.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica 25*(100), 283–299.

Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*(2), 335–346.

Pierse, R. G. and A. J. Snell (1995). Temporal aggregation and the power of tests for a unit root. *Journal of Econometrics 65*(2), 333–345.

Polemis, M. L. and M. G. Tsionas (2016). An alternative semiparametric approach to the modelling of asymmetric gasoline price adjustment. *Energy Economics 56*, 384–388.

Reilly, B. and R. Witt (1998). Petrol price asymmetries revisited. *Energy Economics 20*(3), 297–308.

Shiller, R. J. and P. Perron (1985). Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters 18*(4), 381–386.

Shin, Y., B. Yu, and M. Greenwood-Nimmo (2014). Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. In *Festschrift in Honor of Peter Schmidt*, pp. 281–314. Springer.

Soon, S.-V., A. Z. Baharumshah, and N. S. M. Shariff (2017). The persistence in real interest rates: Does it solve the intertemporal consumption behavior puzzle? *Journal of International Financial Markets, Institutions and Money 50*, 36–51.

Spanos, A. (1989). Early empirical findings on the consumption function, stylized facts or fiction: a retrospective view. *Oxford Economic Papers 41*(1), 150–169.

Sukhtankar, S. (2017). Replications in Development Economics. *American Economic Review 107*(5), 32–36.

Thomas, J. J. (1989). The early econometric history of the consumption function. *Oxford Economic Papers 41*(1), 131–149.

Thomas, J. J. (1992). Income Distribution and the Estimation of the Consumption Function: A Historical Analysis of the Early Arguments. *History of Political Economy 24*(1), 153–181.

Twomey, I. and S. West (2008). 2007 ACCC report into Australian petrol prices Review of applicability to the New Zealand petrol market. *Hale & Twomey Ltd. - Document Prepared for the Ministry of Economic Development.*

Valadkhani, A. (2013a). Do petrol prices rise faster than they fall when the market shows significant disequilibria? *Energy Economics 39*, 66–80.

Valadkhani, A. (2013b). Modelling the terminal gate prices of unleaded petrol in Australia. *Economic Modelling 33*, 233–243.

Valadkhani, A., R. Smyth, and F. Vahid (2015). Asymmetric pricing of diesel at its source. *Energy Economics 52*, 183–194.

White, H. (1980). A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica 48*(4), 817–838.

Winfree, J. A. (2010). Issues With Replicating Results in Sports Economics. *Journal of Sports Economics 11*(1), 48–59.

Woytinsky, W. S. (1946). Relationship between consumers' expenditures, savings, and disposable income. *The Review of Economics and Statistics 28*(1), 1–12.

Wulwick, N. J. (1996). Two econometric replications: The historic Phillips and Lipsey-Phillips curves. *History of Political Economy 28*(3), 391–439.