Recent Developments of the Autoregressive Distributed Lag Modelling Framework*

JIN SEO CHO
School of Humanities and Social Sciences, Beijing Institute of Technology, Haidan, Beijing 100081, China
School of Economics, Yonsei University, Seodaemun-gu, Seoul 03722, Korea

jinseocho@yonsei.ac.kr

MATTHEW GREENWOOD-NIMMO
Faculty of Business and Economics, University of Melbourne, Carlton, VIC 3053, Australia
Centre for Applied Macroeconomic Analysis, Australian National University, Canberra, ACT 2600, Australia

matthew.greenwood@unimelb.edu.au

YONGCHEOL SHIN
Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, U.K.

yongcheol.shin@york.ac.uk

February 24, 2021

Abstract

We review the literature on the Autoregressive Distributed Lag (ARDL) model, from its origins in the analysis of autocorrelated trend stationary processes to its subsequent applications in the analysis of cointegrated non-stationary time series. We then survey several recent extensions of the ARDL model, including asymmetric and nonlinear generalisations of the ARDL model, the quantile ARDL model, the pooled mean group dynamic panel data model and the spatio-temporal ARDL model.

Key Words: Autoregressive Distributed Lag (ARDL) Model; Asymmetry, Nonlinearity and Threshold Effects; Quantile Regression; Panel Data; Spatial Analysis.

JEL Classifications: C22.

---

*Corresponding author: Matthew Greenwood-Nimmo. We are grateful for the helpful comments of the editor, the associate editor and two anonymous referees. Part of the work for this paper was completed when Cho was visiting the Department of Economics, Chinese University of Hong Kong, whose kind hospitality is gratefully acknowledged. Cho also acknowledges financial support from the Yonsei University Research Grant of 2020. The usual disclaimer applies.

1 Introduction

The ARDL model combines an autoregressive component (i.e. lags of a scalar dependent variable) with a distributed lag component (i.e. lags of a vector of explanatory variables). It has its origins in the analysis of autocorrelated trend stationary processes. In this context, the general practice is to model the de-trended series as a stationary distributed lag, or ARDL model (Koyck, 1954; Almon, 1965). Provided that the lag structure of the ARDL model is sufficiently rich to account for the autocorrelation structure in the data, estimation can proceed by ordinary least squares (OLS) and standard inference is applicable with respect to the long-run properties of the model. The application of the ARDL model to trend stationary data has been addressed in several excellent survey articles, so we do not repeat this discussion here; rather, the reader is referred to Griliches (1967), Wallis (1969), Nerlove (1972), Sims (1974), Maddala (1977), Thomas (1977), Zellner (1979), Hendry, Pagan, and Sargan (1984) and Wickens and Breusch (1988). For a formal treatment of the probability theory underlying many of the proposed estimators, see Dhrymes (1971).

We focus our attention on the more challenging case first considered by Pesaran and Shin (1998), in which the ARDL model is applied to the analysis of the cointegrating relation between first-order integrated, or I(1), processes. Pesaran, Shin, and Smith (2001) provide an important generalisation that makes use of a bounds-testing framework to allow for mixed orders of integration among the variables entering the ARDL model. The methodology proposed in these two papers has several appealing features. In addition to providing a means to model long-run relationships among variables of unknown integration order, the ARDL model has a straightforward and intuitive error correction interpretation, it is estimable by OLS, it can handle serial correlation through the selection of an appropriate lag order and it can provide consistent estimates of the long-run parameters, even if the explanatory variables are weakly endogenous. By virtue of its many desirable features, the methodology proposed in these two papers has proven highly influential, having given rise to thousands of empirical applications. It has also provided the basis for several notable methodological extensions, which we refer to as 'ARDL variants' and which we survey in detail.

The first group of ARDL variants that we consider share a common concern with generalising the ARDL process to accommodate various forms of asymmetry and nonlinearity. Pesaran and Shin (1998) and Pesaran et al. (2001) assume that the long-run relationship is linear. This assumption may be prove restrictive in practice, as it excludes the possibility of a nonlinear long-run relationship. Shin, Yu, and Greenwood-Nimmo (2014) were the first to consider departures from linearity in the ARDL literature. In their Nonlinear ARDL (NARDL) model, partial sum decompositions of the explanatory variables are used to accommodate asymmetric phenomena in the long-run and in the short-run simultaneously. Shin et al. (2014) propose an OLS estimation procedure, the properties of which are validated by simulation. However, the authors observe an asymptotic singularity issue that frustrates efforts to obtain the limit distribution of the OLS estimator. Cho, Greenwood-Nimmo, and Shin (2019, 2020b) show that this singularity issue arises due to the use of partial sum decompositions and propose an alternative two-step estimation framework that resolves the singularity issue and that is analytically tractable. In these two papers, Cho et al. employ a novel re-parameterisation of the NARDL model to eliminate the singularity problem and then develop asymptotic theory for their two-step estimator, in which the parameters of the long-run relationship are estimated in the first step using the fully-modified OLS (FM-OLS) estimator of Phillips and Hansen (1990) before the short-run dynamic parameters are estimated in the second step by OLS. This procedure exploits the different convergence rates of the first and second step estimators to deliver consistent and asymptotically normal estimators of both the long- and short-run parameters. This asymptotic normality allows the authors to employ Wald's (1943) testing principle to develop tests for both long- and short-run asymmetry that asymptotically converge to χ^2 distributions. This is an important innovation, as it provides a means to test the hypothesis of linearity that had previously been asserted in the linear ARDL literature.

The NARDL model has been widely adopted in the literature due to its ease of implementation and interpretation. Part of this is due to the simplifying assumption that the threshold parameter embedded in the NARDL model is known a priori. Typically, one uses a threshold value of zero in the construction of the partial sum processes, which gives rise to an elegant interpretation related to positive and negative changes in the vector of explanatory variables. This may be particularly advantageous in circumstances in which the sign of a change in an explanatory variable carries a natural interpretation, such as exchange rate appreciations and depreciations. Furthermore, from an inferential perspective, the use of a known threshold simplifies the analysis, because it ensures that the identification problem described by Davies (1977, 1987) does not arise. However, in some cases, there may be no reason to believe that setting a threshold value of zero is appropriate a priori and/or the value of the threshold parameter may be of interest in its own right. In such cases, one may wish to treat the threshold parameter as an unknown quantity to be estimated. This is the motivation for the development of the Threshold ARDL (TARDL) model, estimation and inference on which is the focus of Cho, Greenwood-Nimmo, and Shin (2020c,d).

Cho et al. (2020c) consider a setting with a single explanatory variable that is decomposed into partial sum processes around an unknown momentum-type threshold. In this environment, the authors demonstrate that one must address a multifold identification problem in order to conduct inference on the number of regimes. The multifold identification problem arises because the null hypothesis of a single regime is composed of three sub-null hypotheses, such that the alternative hypothesis is defined by the negation of the union of these sub-null hypotheses. Testing the joint null hypothesis is complicated by the fact that each sub-null hypothesis represents an alternative of the other sub-null hypotheses. Drawing on the existing literature on the multifold identification problem, Cho et al. (2020c) develop a quasi likelihood ratio (QLR) test for the existence of a distinct threshold level and show that the null limit distribution of the test statistic can be represented by separable functionals of two Gaussian stochastic processes provided that the TARDL specification contains an intercept term.

Cho et al. (2020d) pursue a different approach, with the objective of determining the number of regimes in a TARDL model with a single explanatory variable via a model selection procedure. To this end, the authors consider six candidate information criteria, namely the Akaike information criterion (AIC), the Schwarz information criterion (SIC) and the Hannan-Quinn information criterion (HQIC), as well as modified versions of each criterion defined following Pitarakis (2006), which we denote pAIC, pSIC and pHQIC. By means of an extensive simulation study, Cho et al. (2020d) show that, in general, the standard SIC outperforms the other information criteria if the independent variable is not driven by a time drift. However, if this is not the case, then the pSIC outperforms the other information criteria in small samples, while SIC and pSIC jointly outperform the remainder in large samples.

As it is typically implemented, the NARDL model focuses on sign asymmetry, while the TARDL model allows for greater flexibility and admits the possibility of size asymmetry as well as sign asymmetry. Neither of these models addresses the issue of locational asymmetry, whereby the error correction relationship between a scalar dependent variable, y_t, and a vector of explanatory variables, $\mathbf{x}_t$, is allowed to vary throughout the conditional distribution of $y_t|\mathbf{x}_t$. Cho, Kim, and Shin (2015) were the first to address this issue by applying quantile regression following Koenker and Bassett (1978) and Xiao (2009) to estimate the parameters of an ARDL model at different locations in the conditional distribution. This contrasts with the classical estimators employed by Pesaran and Shin (1998) and Pesaran et al. (2001), which characterise the relationship at the conditional mean only. Cho et al. (2015) establish that the limiting distribution of the quantile regression estimators of the long-run and short-run parameters is mixed-normal, which implies that asymptotic critical values for Wald statistics testing various hypotheses on the quantile coefficients can be retrieved from χ^2 distributions. In a subsequent paper, Cho, Greenwood-Nimmo, Kim, and Shin (2020a) apply quantile regression to the estimation of a NARDL model, thereby providing a unified framework for the estimation of an ARDL process that encompasses both size and locational asymmetry.

In addition to the asymmetric and nonlinear ARDL variants described above, the ARDL specification has also been applied to large T panels – that is, panels with many time series observations per group. Pesaran, Shin, and Smith (1999) consider a setting in which a panel ARDL specification is estimated under the assumption that the long-run parameters are homogeneous across groups, while the short-run dynamic parameters are heterogeneous. This setting has intuitive appeal, as there are often good reasons to believe that long-run equilibrium relationships should be approximately common across groups but, in general, the same cannot be said for dynamic parameters. The so-called Pooled Mean Group (PMG) estimator employs an innovative hybrid estimation approach, whereby the homogeneous long-run parameters are estimated by maximum likelihood with the data pooled over groups, while the heterogeneous short-run parameters are estimated on a group-specific basis and their group-wide distribution is summarised by taking averages across groups. In the case of stationary data, Pesaran et al. demonstrate that the PMG estimator is consistent and asymptotically normal while, under non-stationarity, it is consistent and converges to a mixed-normal distribution.

The final model that we consider is another panel data variant of the ARDL process, this time designed to exploit the growing availability of spatial time series data. The Spatio-Temporal ARDL (STARDL) model proposed by Shin and Thornton (2019) is a system of ARDL equations, each of which is augmented with variables defined as spatially-weighted averages. Shin and Thornton develop a quasi-maximum likelihood estimator, as well as an alternative estimator based on the control function framework. In both cases, the authors demonstrate that their estimators are consistent and asymptotically normal. The STARDL framework can be thought of as an omnibus model that nests several popular spatial models, including the spatial Durbin model analysed by Lee and Yu (2010) and Elhorst (2014) and the heterogeneous spatial autoregressive panel data model of Aquaro, Bailey, and Pesaran (2021). Shin and Thornton also note that the network structure of the STARDL model is amenable to the application of many of the popular tools of network analysis, including centrality statistics and clustering algorithms. By analogy to the dynamic multiplier effects that are widely used in the analysis of ARDL models, the authors derive diffusion multipliers from the spatial system that can be used to explore the properties of the spatial dynamic network.

This paper proceeds as follows. In Section 2, we briefly review the ARDL model of Pesaran and Shin

(1998) and Pesaran et al. (2001) to provide relevant background for the six ARDL variants on which our survey focuses. In Section 3, we introduce the classic NARDL model of Shin et al. (2014) based on a known threshold, before discussing the case of unknown thresholds in the context of the TARDL model. In Section 4, we outline the quantile ARDL and quantile NARDL models, while Sections 5 and 6 are devoted to the PMG estimator of Pesaran et al. (1999) and the STARDL model of Shin and Thornton (2019), respectively. In Section 7, we offer concluding remarks and we briefly discuss some promising avenues for continuing development of the ARDL framework.

2 The ARDL Model with Non-Stationary Data

Pesaran and Shin (1998) study the following process:

$$y_t = \gamma_* + \sum_{j=1}^{p} \phi_{j*} y_{t-j} + \sum_{j=0}^{q} \boldsymbol{\theta}'_{j*} \boldsymbol{x}_{t-j} + \varepsilon_t, \tag{1}$$

which is obtained by applying the distributed-lag form to the integrated time series $(y_t, \boldsymbol{x}'_t)' \in \mathbb{R}^{1+k}$. Pesaran and Shin refer to (1) as an ARDL(p, q) process. This model represents an extension of the prior ARDL literature focusing on trend stationary data (e.g. Koyck, 1954; Almon, 1965) with the intention of capturing not just the dynamic relationship but also the long-run relationship between y_t and $\boldsymbol{x}_t$. The authors go on to demonstrate that (1) can be expressed in the following alternate form:

$$y_t = \gamma_* + \boldsymbol{x}'_t \boldsymbol{\gamma}_* + \sum_{j=1}^{p} \phi_{j*} y_{t-j} + \sum_{j=0}^{q-1} \Delta \boldsymbol{x}'_{t-j} \boldsymbol{\delta}_{j*} + \varepsilon_t, \tag{2}$$

where $\boldsymbol{\gamma}_* := \sum_{j=0}^{q} \boldsymbol{\theta}_{j*}$ and $\boldsymbol{\delta}_{j*} := -\sum_{i=j+1}^{q} \boldsymbol{\theta}_{i*}$. The long-run relationship between y_t and $\boldsymbol{x}_t$ embedded in (2) can be represented as $y_t = \mu_* + \boldsymbol{x}'_t \boldsymbol{\beta}_* + u_t$, where:

$$\mu_* := \gamma_* \left(1 - \sum_{i=1}^{p} \phi_{i*}\right)^{-1}, \quad \boldsymbol{\beta}_* := \boldsymbol{\gamma}_* \left(1 - \sum_{i=1}^{p} \phi_{i*}\right)^{-1},$$

and u_t is a stationary process defined by $\{\Delta \boldsymbol{x}_t, \varepsilon_t, \Delta \boldsymbol{x}_{t-1}, \varepsilon_{t-1}, \ldots\}$.

The ARDL process can be represented as an error correction process in the tradition of the London

School of Economics (e.g. Sargan, 1964; Hendry and Mizon, 1978, among others) as follows:

$$\Delta y_t = \rho_* y_{t-1} + \boldsymbol{\theta}_*' \boldsymbol{x}_{t-1} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}' \Delta \boldsymbol{x}_{t-j} + \varepsilon_t, \tag{3}$$

where $\{\varepsilon_t; \mathcal{F}_t^0\}$ is a martingale difference sequence and $\mathcal{F}_t^0$ is the smallest σ-algebra driven by $\{y_{t-1}, \boldsymbol{x}_t, y_{t-2}, \boldsymbol{x}_{t-1}, \ldots\}$. Note that the error correction process (3) can be derived from the ARDL process (1) by letting:

$$\rho_* := \sum_{j=1}^{p} \phi_{j*} - 1, \quad \boldsymbol{\theta}_* := \sum_{j=0}^{q} \boldsymbol{\theta}_{j*}, \quad \boldsymbol{\pi}_{0*} := \boldsymbol{\theta}_{0*}, \quad \varphi_{\ell*} := -\sum_{i=\ell+1}^{p} \phi_{i*}, \quad \text{and} \quad \boldsymbol{\pi}_{j*} := -\sum_{i=j+1}^{q} \boldsymbol{\theta}_{i*}$$

for $\ell = 1, 2, \ldots, p-1$ and $j = 1, 2, \ldots q-1$. Furthermore, if y_t is cointegrated with $\boldsymbol{x}_t$ such that $u_{t-1} := y_{t-1} - \boldsymbol{\beta}_*' \boldsymbol{x}_{t-1}$ is stationary, then (3) can be re-written as:

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*} \Delta y_{t-j} + \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}' \Delta \boldsymbol{x}_{t-j} + \varepsilon_t, \tag{4}$$

where $\boldsymbol{\beta}_* := -(\boldsymbol{\theta}_*/\rho_*)$, which provides motivation for the separate estimation of the long-run and short-run parameters. Engle and Granger (1987) estimate the long-run parameter by OLS, while Phillips and Hansen (1990) propose to estimate the same parameter using the fully-modified OLS (FM-OLS) estimator, which overcomes the asymptotic bias in the OLS estimator and which asymptotically follows a mixed normal distribution. Many other procedures for the consistent and efficient estimation of the long-run coefficients have been proposed in the literature (e.g. Stock and Watson, 1993; Johansen, 1988, among others). Pesaran and Shin (1998) propose an elegant procedure, in which one first estimates the coefficients of (2) by OLS and then estimates $\boldsymbol{\beta}_*$ as $\widehat{\boldsymbol{\beta}}_T := \widehat{\boldsymbol{\gamma}}_T \left(1 - \sum_{i=1}^{p} \widehat{\phi}_{T,i}\right)^{-1}$, where $\widehat{\boldsymbol{\gamma}}_T$ and $\widehat{\phi}_{T,i}$ are the OLS estimators of $\boldsymbol{\gamma}_*$ and ϕ_{i*}, respectively. The authors show that $\widehat{\boldsymbol{\beta}}_T$ converges to $\boldsymbol{\beta}_*$ at the rate of T and asymptotically follows a mixed normal distribution, while the short-run parameter estimators converge to the unknown parameters in (2) at the rate of $\sqrt{T}$ and are asymptotically normal. For this reason, Pesaran and Shin (1998) are able to apply Wald's (1943) testing principle to conduct inference on the unknown long-run and short-run parameters, with simulation evidence indicating that the behaviour of the Wald test statistic is well-approximated by asymptotic results, even in relatively small samples. These results are obtained under relatively mild conditions on the times series. In particular, in addition to the conditions described above, Pesaran and Shin (1998) assume that the variables in $\boldsymbol{x}_t$ are not cointegrated among themselves and that the roots of $\phi(L) = 0$ are strictly greater than unity, where $\phi(L) := (1 - \sum_{j=1}^{p} \phi_{j*} L^j)$. Without these conditions, the desired long-run relationship is not established. Furthermore, Pesaran and Shin (1998) suggest that the lag orders of the ARDL process can be selected using the Schwarz (1978) information criterion, although other approaches including general-to-specific lag selection are also commonly used in practice.

In an influential contribution to the ARDL literature, Pesaran et al. (2001) develop a bounds test for the null hypothesis of no long-run (cointegrating) levels relationship using the error correction representation of the ARDL process. The authors derive the null limit distributions of the F-statistic testing the null hypothesis $\rho_* = 0$ and $\boldsymbol{\theta}_* = \mathbf{0}$ in (3) as functionals of a Wiener process under various assumptions on the model coefficients and the order of integration of the variables. Pesaran et al. (2001) tabulate critical value bounds from these asymptotic distributions for use in applied research in which one is unsure of the true order of integration of the variables entering the model.

The ARDL framework developed by Pesaran and Shin (1998) and Pesaran et al. (2001) has been widely adopted in applied research. At the time of writing, Google Scholar records approximately 6,000 and 14,000 citations to these two papers, respectively. Applications of the ARDL model can be found in many areas of the social sciences, with a particular prevalence in finance, macroeconomics and energy economics. There are many reasons for the popularity of the ARDL model. Perhaps most importantly, the functional form of the ARDL(p, q) process has intuitive appeal, as it allows for partial adjustment toward an economically meaningful long-run equilibrium relationship between y_t and $\boldsymbol{x}_t$. In addition, the ARDL model naturally accounts for the serial correlation structure that exists among the first differences of y_t and $\boldsymbol{x}_t$ and can provide consistent estimates of the long-run parameters even in the presence of weak endogeneity.

In addition to the many empirical applications of the ARDL model, a considerable body of work has been dedicated to the development of variants of the ARDL model that allow for departures from linearity and alternative data structures, including panel data and spatial panel data. The remainder of this paper is dedicated to reviewing several of these developments.

3 Nonlinearity and Threshold Effects

In this section, we consider two generalisations of the ARDL model that accommodate departures from linearity in a manner that is pertinent to many integrated economic time series and that nest the linear

ARDL model as a special case, thereby providing a means to test the hypothesis of linearity. The nonlinear autoregressive distributed lag (NARDL) and threshold autoregressive distributed lag (TARDL) models employ partial sum decompositions of the explanatory variables to accommodate asymmetry and nonlinearity, respectively. Consider the model:

$$y_t = \gamma_* + \sum_{j=1}^{p} \phi_{j*} y_{t-j} + \sum_{j=0}^{q} (\boldsymbol{\theta}_{j*}^{+\prime} \boldsymbol{x}_{t-j}^{+} + \boldsymbol{\theta}_{j*}^{-\prime} \boldsymbol{x}_{t-j}^{-}) + \varepsilon_t, \tag{5}$$

where $\boldsymbol{x}_t \in \mathbb{R}^k$, $\boldsymbol{x}_t^{+} := \sum_{j=1}^{t} \Delta \boldsymbol{x}_j^{+}$ and $\boldsymbol{x}_t^{-} := \sum_{j=1}^{t} \Delta \boldsymbol{x}_j^{-}$ with:

$$\Delta x_{tj}^{+} := \begin{cases} 0, & \text{if } \tau_{j*} \geq \Delta x_{tj}; \\ \Delta x_{tj}, & \text{otherwise,} \end{cases} \quad \text{and} \quad \Delta x_{tj}^{-} := \begin{cases} 0, & \text{if } \tau_{j*} < \Delta x_{tj}; \\ \Delta x_{tj}, & \text{otherwise.} \end{cases}$$

In this case, $\{\Delta \boldsymbol{x}_t\}$ is a strictly stationary process and we define $\boldsymbol{\tau}_* := [\tau_{1*}, \ldots, \tau_{j*}, \ldots, \tau_{k*}]'$. Note that (5) extends the ARDL specification in (1) to accommodate a nonlinear relationship between y_t and $\boldsymbol{x}_t$, where differences in the values of $\boldsymbol{\theta}_{j*}^{+}$ and $\boldsymbol{\theta}_{j*}^{-}$ capture the asymmetric effect of $\boldsymbol{x}_{t-j}^{+}$ and $\boldsymbol{x}_{t-j}^{-}$ on y_t. The motivation behind the development of the NARDL process is the desire to capture the asymmetric stochastic trend of an integrated series and its long-run relationship with other integrated series. The literature on asymmetric exchange rate pass-through into import prices offers a good example of such an asymmetric long-run relationship (e.g. Brun-Aguerre, Fuertes, and Greenwood-Nimmo, 2017).

If the threshold level, $\boldsymbol{\tau}_*$, is known (e.g. 0 or the mean of Δx_t) then, although (5) is linear in the unknown parameters, it can accommodate asymmetry in the long-run and the short-run relationships between y_t and $\boldsymbol{x}_t$. This is the NARDL(p, q) model proposed by Shin et al. (2014). If the threshold parameter, $\boldsymbol{\tau}_*$, is unknown, then one must estimate it in order to evaluate the nonlinear relationship between y_t and $\boldsymbol{x}_t$. In this case, the estimation problem is nonlinear due to the necessity to estimate the threshold parameter. This is the TARDL(p, q) model advanced by Cho et al. (2020c,d). The motivation for the development of the TARDL model lies in the realisation that asymmetric phenomena of the type modelled by the NARDL process may arise around unknown threshold levels that cannot be deduced a priori, which makes it necessary to consistently estimate the threshold level, $\boldsymbol{\tau}_*$. By way of illustration, returning to the example of exchange rate pass-through into import prices, one can easily conceive of a situation in which import prices adjust differentially according not just to the sign but also to the magnitude of exchange rate changes.

The NARDL/TARDL model (5) can be transformed into an error correction model. For some ρ_*, $\boldsymbol{\theta}_*^{+}$,

$\boldsymbol{\theta}_*^{-1}$, φ_{j*} $(j = 1, 2, \ldots, p-1)$, $\boldsymbol{\pi}_{j*}^{+}$, and $\boldsymbol{\pi}_{j*}^{-}$ $(j = 0, 1, \ldots, q-1)$, we can re-write (5) as follows:

$$\Delta y_t = \rho_* y_{t-1} + \boldsymbol{\theta}_*^{+\prime}\boldsymbol{x}_{t-1}^{+} + \boldsymbol{\theta}_*^{-\prime}\boldsymbol{x}_{t-1}^{-} + \gamma_* + \sum_{j=1}^{p-1}\varphi_{j*}\Delta y_{t-j} + \sum_{j=0}^{q-1}\left(\boldsymbol{\pi}_{j*}^{+\prime}\Delta\boldsymbol{x}_{t-j}^{+} + \boldsymbol{\pi}_{j*}^{-\prime}\Delta\boldsymbol{x}_{t-j}^{-}\right) + \varepsilon_t, \quad (6)$$

where $\{\varepsilon_t, \mathcal{F}_t\}$ is a martingale difference sequence, $\mathcal{F}_t$ is the smallest σ-algebra driven by $\{y_{t-1}, \boldsymbol{x}_t^{+}, \boldsymbol{x}_t^{-}, y_{t-2}, \boldsymbol{x}_{t-1}^{+}, \boldsymbol{x}_{t-1}^{-}, \ldots\}$, $\rho_* := \sum_{j=1}^{p}\phi_{j*} - 1$, $\boldsymbol{\theta}_*^{+} := \sum_{j=0}^{q}\boldsymbol{\theta}_{j*}^{+}$, $\boldsymbol{\theta}_*^{-} := \sum_{j=0}^{q}\boldsymbol{\theta}_{j*}^{-}$, $\boldsymbol{\pi}_{0*}^{+} := \boldsymbol{\theta}_{0*}^{+}$ and $\boldsymbol{\pi}_{0*}^{-} := \boldsymbol{\theta}_{0*}^{-}$ while $\varphi_{\ell*} := -\sum_{i=\ell+1}^{p}\phi_{i*}$, $\boldsymbol{\pi}_{j*}^{+} := -\sum_{i=j+1}^{q}\boldsymbol{\theta}_{i*}^{+}$ and $\boldsymbol{\pi}_{j*}^{-} := -\sum_{i=j+1}^{q}\boldsymbol{\theta}_{i*}^{-}$ for $\ell = 1, 2, \ldots, p-1$ and $j = 1, 2, \ldots q-1$.

If y_t is cointegrated with $(\boldsymbol{x}_t^{+\prime}, \boldsymbol{x}_t^{-\prime})'$ such that $u_{t-1} := y_{t-1} - \boldsymbol{\beta}_*^{+\prime}\boldsymbol{x}_{t-1}^{+} - \boldsymbol{\beta}_*^{-\prime}\boldsymbol{x}_{t-1}^{-}$ is a cointegration error, then (6) can be re-written as follows:

$$\Delta y_t = \rho_* u_{t-1} + \gamma_* + \sum_{j=1}^{p-1}\varphi_{j*}\Delta y_{t-j} + \sum_{j=0}^{q-1}\left(\boldsymbol{\pi}_{j*}^{+\prime}\Delta\boldsymbol{x}_{t-j}^{+} + \boldsymbol{\pi}_{j*}^{-\prime}\Delta\boldsymbol{x}_{t-j}^{-}\right) + \varepsilon_t, \quad (7)$$

where $u_{t-1} := y_{t-1} - \boldsymbol{\beta}_*^{+\prime}\boldsymbol{x}_{t-1}^{+} - \boldsymbol{\beta}_*^{-\prime}\boldsymbol{x}_{t-1}^{-}$ is a stationary process that is possibly correlated with $\Delta\boldsymbol{x}_t$, $\boldsymbol{\beta}_*^{+} := -(\boldsymbol{\theta}_*^{+}/\rho_*)$ and $\boldsymbol{\beta}_*^{-} := -(\boldsymbol{\theta}_*^{-}/\rho_*)$ are the nonlinear long-run parameters and the remaining parameters capture the (nonlinear) short-run dynamics. The error correction form (7) has the interesting implication that, even if the population mean of $\Delta\boldsymbol{x}_t$ is zero, y_t can be an integrated process with a time drift, because both $\Delta\boldsymbol{x}_t^{+}$ and $\Delta\boldsymbol{x}_t^{-}$ cannot have zero population mean. This is a different form of cointegrating relationship from the standard case, in which a cointegrating relationship between integrated processes with and without a time drift is driven by the presence of a non-zero intercept in (7).

The NARDL framework has been adopted widely, because it provides a simple method for the analysis of asymmetries of the type that may arise in many areas of the social sciences. Furthermore, Shin et al. (2014) show that one may construct cumulative dynamic multipliers from the estimated parameters of the NARDL model, which provides an easily interpreted visualisation of the traverse to an equilibrium position following a shock. Unlike structural impulse response analysis, the cumulative dynamic multipliers do not rely on controversial procedures for the identification of structural shocks. Due in part to these desirable attributes, Google Scholar reports more than 1,400 citations to Shin et al. (2014) at the time of writing, with applications in diverse fields such as criminology (e.g. Box, Gratzer, and Lin, 2018), energy economics (e.g. Hammoudeh, Lahiani, Nguyen, and Sousa, 2015), financial economics (e.g. He and Zhou, 2018), monetary economics (e.g. Claus and Nguyen, 2019) and tourism (e.g. Süssmuth and Woitek, 2013), among others. The majority of existing NARDL applications follow Shin et al. (2014) and employ a single known threshold value of zero, but several studies including Fedoseeva (2013) and Pal and Mitra (2015) have used multiple thresholds. As described above, the nascent interest in the use of non-zero thresholds in this literature represents an important motivating factor behind the development of the TARDL model by Cho et al. (2020c,d).

3.1 Estimation and Inference for the NARDL Model

Shin et al. (2014) propose to estimate the NARDL model in a single step by OLS. However, they note that efforts to derive asymptotic theory for the single-step OLS estimator are frustrated by an asymptotic singularity problem that arises from the presence of partial sum decompositions of the explanatory variables. Consequently, Shin et al. provide simulation evidence regarding the performance of the single-step estimator but the corresponding theory has yet to be developed. Cho et al. (2019) study the asymptotic singularity issue in detail and develop theory for a new two-step estimation framework. Let:

$$\boldsymbol{z}_t := \left[\begin{array}{ccc:cccccccc} y_{t-1} & \boldsymbol{x}_{t-1}^{+\prime} & \boldsymbol{x}_{t-1}^{-\prime} & 1 & \Delta\boldsymbol{y}_{t-1}' & \Delta\boldsymbol{x}_t^{+\prime} & \dots & \Delta\boldsymbol{x}_{t-q+1}^{+\prime} & \Delta\boldsymbol{x}_t^{-\prime} & \dots & \Delta\boldsymbol{x}_{t-q+1}^{-\prime} \end{array}\right]',$$

where $\Delta\boldsymbol{y}_{t-1} := [\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p+1}]'$. Under the assumption that $\mathbb{E}[\Delta\boldsymbol{x}_t] = \mathbf{0}$, Cho et al. (2019) show that $\mathbf{D}_T^{-1}(\sum_{t=1}^T \boldsymbol{z}_t\boldsymbol{z}_t')\mathbf{D}_T^{-1}$ converges to a singular matrix, where $\mathbf{D}_T := \text{diag}(T^{3/2}\mathbf{I}_{2+2k}, T^{1/2}\mathbf{I}_{p+2qk})$. This is the singularity to which Shin et al. (2014) allude and which provides motivation for the development of a two-step estimation strategy by Cho et al. (2019). Consider the long-run relationship embedded in the NARDL model, which can be written as $y_t = \varsigma_* + \boldsymbol{x}_t^{+\prime}\boldsymbol{\beta}_*^+ + \boldsymbol{x}_t^{-\prime}\boldsymbol{\beta}_*^- + u_t$. Cho et al. (2019) show that another asymptotic singularity issue arises here, as the matrix inverse required to obtain the OLS estimator of $(\varsigma_*, \boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$ is asymptotically singular, because $\bar{\mathbf{D}}_T^{-1}\left(\sum_{t=1}^T \boldsymbol{v}_t\boldsymbol{v}_t'\right)\bar{\mathbf{D}}_T^{-1}$ converges to a singular matrix, where $\boldsymbol{v}_t := (1, \boldsymbol{x}_t^{+\prime}, \boldsymbol{x}_t^{-\prime})'$ and $\bar{\mathbf{D}}_T := \text{diag}(T^{1/2}, T^{3/2}\mathbf{I}_{2k})$. To overcome this problem, the authors re-parameterise the long-run relationship as $y_t = \varsigma_* + \boldsymbol{x}_t^{+\prime}\boldsymbol{\lambda}_* + \boldsymbol{x}_t'\boldsymbol{\eta}_* + u_t$, where $\boldsymbol{x}_t \equiv \boldsymbol{x}_t^+ + \boldsymbol{x}_t^-$, $\boldsymbol{\lambda}_* = \boldsymbol{\beta}_*^-$ and $\boldsymbol{\eta}_* + \boldsymbol{\lambda}_* = \boldsymbol{\beta}_*^+$. Using this re-parameterisation, Cho et al. (2019) assume k to be unity (i.e., $k = 1$) and show that one can first estimate $(\varsigma_*, \boldsymbol{\lambda}_*', \boldsymbol{\eta}_*')'$ by OLS to obtain $(\widehat{\varsigma}_T, \widehat{\boldsymbol{\lambda}}_T', \widehat{\boldsymbol{\eta}}_T')'$ and one can then estimate $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$ as $\widehat{\boldsymbol{\beta}}_T^- := \widehat{\boldsymbol{\lambda}}_T$ and $\widehat{\boldsymbol{\beta}}_T^+ := \widehat{\boldsymbol{\eta}}_T + \widehat{\boldsymbol{\lambda}}_T$.

With the long-run parameters estimated in this way, the singularity problem is resolved. If we let $\boldsymbol{q}_t := (1, \boldsymbol{x}_t^{+\prime}, \boldsymbol{x}_t')'$ and $\widetilde{\mathbf{D}}_T := \text{diag}(T^{1/2}, T^{3/2}\mathbf{I}_k, T\mathbf{I}_k)$, then $\widetilde{\mathbf{D}}_T^{-1}\left(\sum_{t=1}^T \boldsymbol{q}_t\boldsymbol{q}_t'\right)\widetilde{\mathbf{D}}_T^{-1}$ weakly converges to a random matrix associated with a Brownian motion. Consequently, it is possible to show that the long-run parameter estimator is T-consistent and has an asymptotic distribution characterised by a Brownian motion. Due to the super-consistency of the long-run parameter estimator, it is possible to estimate the short-run parameters of (7) by OLS by replacing u_{t-1} in (7) with $\widehat{u}_{t-1} := y_{t-1} - \boldsymbol{x}_{t-1}^{+}\widehat{\boldsymbol{\beta}}_T^{+\prime} - \boldsymbol{x}_{t-1}^{-}\widehat{\boldsymbol{\beta}}_T^{-\prime}$. Because all of the variables in (7) are stationary and $\{\varepsilon_t\}$ is a martingale difference array, the short-run parameter estimator is $\sqrt{T}$-consistent and asymptotically normal.

The procedure detailed above provides an operational framework for the estimation of the NARDL model but it suffers from the drawback that the asymptotic distribution of the long-run parameter estimator is non-normal and depends on nuisance parameters, which complicates inference on the long-run parameters. To address this issue, Cho et al. (2019) note that one may simply estimate $(\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$ using the FM-OLS estimator of Phillips and Hansen (1990) rather than by OLS. If $\widetilde{\boldsymbol{\Sigma}}_T$ and $\widetilde{\boldsymbol{\Pi}}_T$ are consistent estimators for the following asymptotic covariance matrices:

$$\boldsymbol{\Sigma} := \begin{bmatrix} \Sigma^{(1,1)} & \Sigma^{(1,2)} \\ \Sigma^{(2,1)} & \sigma^{(2,2)} \end{bmatrix} := \text{acov}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix} \Delta\boldsymbol{x}_t \\ u_t \end{bmatrix}, \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix} \Delta\boldsymbol{x}_t \\ u_t \end{bmatrix}\right] \quad \text{and}$$

$$\boldsymbol{\Pi}_T := \begin{bmatrix} \Pi^{(1,1)} & \Pi^{(1,2)} \\ \Pi^{(2,1)} & \pi^{(2,2)} \end{bmatrix} := \text{acov}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix} \Delta\boldsymbol{x}_t \\ u_t \end{bmatrix}, \begin{bmatrix} \Delta\boldsymbol{x}_0 \\ u_0 \end{bmatrix}\right],$$

then, the FM-OLS estimator is given by:

$$\widetilde{\boldsymbol{\varrho}}_T := (\widetilde{\varsigma}_T, \widetilde{\boldsymbol{\lambda}}_T', \widetilde{\boldsymbol{\eta}}_T')' := \left(\sum_{t=1}^{T}\boldsymbol{q}_t\boldsymbol{q}_t'\right)^{-1}\left(\sum_{t=1}^{T}\boldsymbol{q}_t\widetilde{y}_t - T\mathbf{S}'\widetilde{\boldsymbol{\Lambda}}_T\right),$$

where $\widetilde{y}_t := y_t - \Delta\boldsymbol{x}_t'\left(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)}\right)^{-1}\widetilde{\boldsymbol{\Sigma}}_T^{(1,2)}$, $\widetilde{\boldsymbol{\Lambda}}_T := \widetilde{\boldsymbol{\Pi}}_T^{(1,2)} - \widetilde{\boldsymbol{\Pi}}_T^{(1,1)}\left(\widetilde{\boldsymbol{\Sigma}}_T^{(1,1)}\right)^{-1}\widetilde{\boldsymbol{\Sigma}}_T^{(1,2)}$, and $\mathbf{S} := [\mathbf{0}_{1\times 2}, 1]$. The long-run parameter estimator is then obtained as $\widetilde{\boldsymbol{\beta}}_T^{-} := \widetilde{\boldsymbol{\lambda}}_T$ and $\widetilde{\boldsymbol{\beta}}_T^{+} := \widetilde{\boldsymbol{\eta}}_T + \widetilde{\boldsymbol{\lambda}}_T$. Cho et al. (2019) prove that the FM-OLS estimator of the long-run parameters is T-consistent and asymptotically follows a mixed-normal distribution. Therefore, if one replaces u_{t-1} in (7) with $\widetilde{u}_{t-1} := y_t - \boldsymbol{x}_t^{+}\widetilde{\boldsymbol{\beta}}_T^{+\prime} - \boldsymbol{x}_t^{-}\widetilde{\boldsymbol{\beta}}_T^{-}$, then the short-run parameters of (7) can be estimated consistently by OLS.

Cho et al. (2019) demonstrate that the Wald test statistic applied to both the long-run and the short-run parameters asymptotically follows a χ^2 distribution under the null hypothesis. The authors examine two null hypotheses: first, that for some $\mathbf{R}_\ell \in \mathbb{R}^{r\times 1}$ and $\mathbf{r} \in \mathbb{R}^r$, $H_0' : \mathbf{R}_\ell(\boldsymbol{\beta}_*^{+} - \boldsymbol{\beta}_*^{-}) = \mathbf{r}$ and, second, that for some $\mathbf{R} \in \mathbb{R}^{r\times 2}$, $H_0'' : \mathbf{R}\boldsymbol{\beta}_* = \mathbf{r}$, where $\boldsymbol{\beta}_* := (\boldsymbol{\beta}_*^{+\prime}, \boldsymbol{\beta}_*^{-\prime})'$. The first null, H_0', is given to test the full or partial equality between $\boldsymbol{\beta}_*^{+}$ and $\boldsymbol{\beta}_*^{-}$, while the second null, H_0'', is considered to test a general family of hypotheses on the long-run parameters. Cho et al. (2019) show that the Wald test statistics follow chi-squared distributions under their respective null hypotheses and that they exhibit consistent power against the relevant $o(T^3)$ and $o(T^2)$ alternatives. These results are established under mild regularity conditions. In particular, the regular stationary, mixing and moment conditions apply to $\{(\Delta\boldsymbol{x}_t', u_t)'\}$, such that the functional central limit theorem may be invoked (e.g. Phillips and Hansen, 1990) along with the positive definite global covariance matrix condition. In addition, the authors assume that $E[\Delta\boldsymbol{x}_t] = \mathbf{0}$, such that cointegration is defined among an integrated series with a time drift, y_t, and an integrated series without drift, $\boldsymbol{x}_t$. Therefore, before implementing the methodology proposed by Cho et al. (2019), the empirical researcher may wish to test whether $\Delta\boldsymbol{x}_t$ has a population mean equal to zero.

In a subsequent paper, Cho et al. (2020b) replace the condition $E[\Delta\boldsymbol{x}_t] = \mathbf{0}$ with $\mathbb{E}[\Delta\boldsymbol{x}_t] \neq \mathbf{0}$ and show that the same asymptotic singularity issue surveyed above arises in this case. Cho et al. (2020b) develop an extension of the two-step estimation framework proposed by Cho et al. (2019) that can accommodate the non-zero population mean of $\Delta\boldsymbol{x}_t$ and that is also effective when $E[\Delta\boldsymbol{x}_t] = \mathbf{0}$ with $k > 1$. If we let $\boldsymbol{\mu}_*^{+} := E[\Delta\boldsymbol{x}_t^{+}]$ and $\boldsymbol{\mu}_*^{-} := E[\Delta\boldsymbol{x}_t^{-}]$, then we can write:

$$\boldsymbol{x}_t^{+} = \boldsymbol{\gamma}_*^{+} + \boldsymbol{\mu}_*^{+} t + \sum_{j=1}^{t} \boldsymbol{s}_j^{+} \quad \text{and} \quad \boldsymbol{x}_t^{+} = \boldsymbol{\gamma}_*^{-} + \boldsymbol{\mu}_*^{-} t + \sum_{j=1}^{t} \boldsymbol{s}_j^{-}, \tag{8}$$

where $\boldsymbol{s}_t^{+} := \Delta\boldsymbol{x}_t^{+} - \boldsymbol{\mu}_*^{+}$ and $\boldsymbol{s}_t^{-} := \Delta\boldsymbol{x}_t^{-} - \boldsymbol{\mu}_*^{-}$. The long-run relationship embedded in the NARDL model can be written as:

$$y_t = \boldsymbol{\beta}_*^{+\prime}\left(\boldsymbol{\mu}_*^{+} t + \boldsymbol{w}_t^{+}\right) + \boldsymbol{\beta}_*^{-\prime}\left(\boldsymbol{\mu}_*^{-} t + \boldsymbol{w}_t^{-}\right) + u_t = \boldsymbol{\beta}_*^{+\prime}\boldsymbol{w}_t^{+} + \boldsymbol{\beta}_*^{-\prime}\boldsymbol{w}_t^{-} + \xi_* + \delta_* t + v_t, \tag{9}$$

where $\boldsymbol{w}_t^{+} := \sum_{j=1}^{t} \boldsymbol{s}_j^{+}$ and $\boldsymbol{w}_t^{-} := \sum_{j=1}^{t} \boldsymbol{s}_j^{-}$, implying that the long-run parameters can be consistently estimated by regressing y_t on $(\boldsymbol{w}_t^{+\prime}, \boldsymbol{w}_t^{-\prime}, 1, t)'$. As $\boldsymbol{w}_t := (\boldsymbol{w}_t^{+\prime}, \boldsymbol{w}_t^{-\prime})'$ is not observable, Cho et al. (2020b) propose to first estimate (8) by regressing $\boldsymbol{x}_t^{+}$ and $\boldsymbol{x}_t^{-}$ on $(1, t)$ to obtain regression residuals, $\widehat{\boldsymbol{w}}_t^{+}$ and $\widehat{\boldsymbol{w}}_t^{-}$, that approximate $\boldsymbol{w}_t^{+}$ and $\boldsymbol{w}_t^{-}$. Next, one can estimate the long-run parameters from the regression of y_t on $(\widehat{\boldsymbol{w}}_t^{+\prime}, \widehat{\boldsymbol{w}}_t^{-\prime}, 1, t)'$. The long-run parameter estimator obtained in this way is shown to be T-consistent. Meanwhile, the short-run parameters can be estimated consistently by OLS in a subsequent step.

Despite the simple structure of this estimator, it is not straightforward to derive its asymptotic distribution, which complicates inference on the long-run parameters; specifically, the null limit distribution of the Wald statistic testing restrictions on the long-run parameters does not follow the chi-squared distribution. Cho et al. (2020b) show that this issue can be resolved by using FM-OLS to estimate the long-run parameters. The resulting long-run parameter estimator is asymptotically mixed-normal, so the associated Wald test statistic asymptotically follows a mixed chi-squared distribution under the null hypothesis. These results are obtained by imposing the regular stationary, mixing and moment conditions on $\{(\boldsymbol{s}_{t-1}^{+\prime}, \boldsymbol{s}_{t-1}^{-\prime}, u_{t-1}, \varepsilon_t)'\}$, together with the positive definite global covariance matrix condition and a martingale difference sequence, $\{\varepsilon_t, \mathcal{F}_t\}$, so as to apply the functional central limit theorem. These conditions are standard in the literature. The most restrictive assumption is that the level of threshold is known a priori and it is to this issue that we now turn.

3.2 Estimation and Inference for the TARDL Model

The NARDL model is relatively simple to estimate due to the fact that the threshold parameter in (5), $\boldsymbol{\tau}_*$, is known a priori. By contrast, if the threshold is unknown, then it must be estimated. Cho et al. (2020c) examine the existence of the threshold using an inferential procedure, while Cho et al. (2020d) address the same issue from a model selection perspective.

Cho et al. (2020c) assume a single regressor (i.e., $k = 1$) and note that the TARDL model (5) reduces to the linear ARDL model if $\theta_{j*}^{+} = \theta_{j*}^{-}$ for all $j = 0, 1, \ldots, q$. If so, the TARDL model has a set of redundant parameters, which may hamper standard inferential procedures. Consequently, the authors first test for the existence of a distinct threshold level using a quasi-likelihood ratio (QLR) test statistic. Let the ARDL and TARDL processes in (3) and (6) be the null and alternative models, respectively. The QLR test statistic is defined by:

$$\mathcal{QLR}_n := T\left(1 - \frac{\widehat{\sigma}_T^2}{\widehat{\sigma}_{T,0}^2}\right),$$

where $\widehat{\sigma}_T^2$ and $\widehat{\sigma}_{T,0}^2$ are the error variance estimators obtained under the alternative and null models by optimising the sums of squared residuals with respect to the unknown parameter:

$$\widehat{\sigma}_{T,0}^2 := \inf \frac{1}{T}\sum_t \left[\Delta y_t - \rho y_{t-1} - \theta x_{t-1} - \gamma - \sum_{j=1}^{p-1} \varphi_j \Delta y_{t-j} - \sum_{j=0}^{q-1} \pi_j \Delta x_{t-j}\right]^2, \quad \text{and}$$

$$\widehat{\sigma}_T^2 := \inf \frac{1}{T}\sum_t \left[\Delta y_t - \rho y_{t-1} - \theta^+ x_{t-1}^+ - \theta^- x_{t-1}^- - \gamma - \sum_{j=1}^{p-1} \varphi_j \Delta y_{t-j} - \sum_{j=0}^{q-1} \pi_j^+ \Delta x_{t-j}^+ + \pi_j^- \Delta x_{t-j}^-\right]^2.$$

While $\widehat{\sigma}_{T,0}^2$ is estimated by OLS, $\widehat{\sigma}_T^2$ is estimated by nonlinear least squares, because τ_* must also be estimated. Cho et al. (2020c) obtain consistent estimates of the linear coefficients as well as τ_* by first estimating the linear coefficients by OLS for each τ and then optimising them with respect to τ.

Cho et al. (2020c) examine the QLR test statistic because such statistics have been applied in the literature to handle multifold identification problems.[1] To understand the multifold identification problem in the TARDL model, note that the identification problem described by Davies (1977, 1987) can arise in three different ways. First, if $\theta_*^+ = \theta_*^-$ and $\boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-$ in (6), where $\boldsymbol{\pi}_*^+ := (\pi_{0*}^+, \pi_{1*}^+, \ldots, \pi_{q-1*}^+)'$ and $\boldsymbol{\pi}_*^- := (\pi_{0*}^-, \pi_{1*}^-, \ldots, \pi_{q-1*}^-)'$, then τ_* is not identified. Second, if $\tau_* = \Phi^{-1}(0)$, where $\Phi(\cdot)$ denotes the cumulative distribution function of Δx_t, then it follows that $\Delta x_t^+ \equiv \Delta x_t$ and $\Delta x_t^- \equiv 0$, which means that the coefficients associated with the negative regime, such as θ_*^- and $\boldsymbol{\pi}_*^-$, are not identified. Finally, if $\tau_* = \Phi^{-1}(1)$, it follows that $\Delta x_t^- \equiv \Delta x_t$ and $\Delta x_t^+ \equiv 0$, such that the coefficients associated with the positive regime are not identified. Consequently, the null hypothesis of a single regime is composed of the following three sub-null hypotheses:

$$H_{01} : \theta_*^+ = \theta_*^- \quad \text{and} \quad \boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-; \quad H_{02} : \tau_* = \Phi^{-1}(0); \quad \text{and} \quad H_{03} : \tau_* = \Phi^{-1}(1),$$

and the negation of the union of the sub-null hypotheses becomes the alternative hypothesis.

Due to this multifold identification problem, testing the joint null hypothesis by standard methods is a challenging task, because each sub-null hypothesis represents an alternative to another sub-null hypothesis, with the result that testing the union of the sub-null hypotheses cannot proceed on the basis of testing the hypothetical locations of the null parameter space. Consequently, in keeping with the established precedent in the literature, Cho et al. (2020c) tackle the multifold identification problem using the QLR test statistic.

Cho et al. (2020c) show that the null limit distribution of the QLR test statistic can be represented as a

[1]The multifold identification problem has been shown to arise in many popular models. For example, Cho and White (2007, 2010) examine a regime-switching process and note that a multifold identification problem arises when testing for one regime versus two regimes. The multifold identification problem is also observed when testing a linear model against an artificial neural network model or a smooth transition autoregressive model and in a consistent moment specification testing context (e.g. Baek, Cho, and Phillips, 2015; Cheong, Cho, and White, 2011; Cho and Ishida, 2012; Cho, Ishida, and White, 2011, 2014; Cho and Phillips, 2018; Seong, Cho, and Teräsvirta, 2019). All of these studies show that test statistics constructed using the likelihood ratio testing principle can overcome the multifold identification problem by providing the null limit distributions of the test statistics, which can be represented as functionals of Gaussian stochastic processes.

functional of two Gaussian stochastic processes, in contrast to results in the prior literature. As the variables entering the TARDL model are cointegrated $I(1)$ processes, an additional Gaussian process is introduced to derive the null limit distribution. Specifically, Cho et al. (2020c) first show that the null approximation of the QLR test statistic under H_{01} dominates the other null approximations obtained under H_{02} and H_{03}, such that the null limit distribution of the QLR test statistic is given by the limit of the null approximation of the QLR test statistic under H_{01}. Next, they note that the functional of the two Gaussian stochastic processes is a sum of two separable functionals of the two Gaussian processes, such that the functional of the first Gaussian process is obtained while testing $\boldsymbol{\pi}_*^+ = \boldsymbol{\pi}_*^-$, whereas the functional of the second Gaussian process is obtained when testing $\theta_*^+ = \theta_*^-$. Furthermore, the authors discover that the first Gaussian process is indexed by the parameter τ due to the Davies (1977, 1987) identification problem. However, the second Gaussian process is formed by a partial sum process introduced by Phillips (1991, theorem 1) when testing a hypothesis on cointegration coefficients. Based on these findings, Cho et al. (2020c) perform a simulation study to show that the limit distribution of each functional can be separately approximated by combining a bootstrap method with a χ^2 distribution. Specifically, the first functional can be approximated by the weighted bootstrap procedure of Hansen (1996), while a χ^2 distribution approximates the second functional, as demonstrated by Phillips (1991).

Cho et al. (2020c) draw attention to several aspects of testing with TARDL models. First, as the null limit distribution of the QLR test statistic is characterised mainly by a functional of a Gaussian process indexed by τ, the authors restrict the dimension of τ be equal to unity (i.e., $k = 1$). This is a pragmatic decision motivated by the difficulty of obtaining the null limit distribution of τ in higher-dimensional cases. Second, the authors show that the specification of the TARDL process also determines the form of the functional of the two Gaussian processes. For example, in the case where the intercept is excluded from the TARDL model (6) because it is known to be zero a priori, the null limit distribution is characterised by *non-separable* functionals of the two Gaussian processes. Consequently, it is recommended that one should estimate the intercept even if its value is known ex ante in order to facilitate the application of the QLR test statistic.

To derive the null distribution of the QLR test statistic, Cho et al. (2020c) impose the regular stationary, mixing, and moment conditions to $\{(\Delta y_t, \Delta \boldsymbol{x}_t')'\}$, along with the globally positive definite covariance matrix condition and a martingale difference sequence, $\{\varepsilon_t, \mathcal{F}_t\}$. In addition, they impose the strong assumption that $(\Delta y_t, \Delta \boldsymbol{x}_t')'$ is continuously distributed with known marginal distribution for $\Delta \boldsymbol{x}_t$. The assumption that the marginal distribution of $\Delta \boldsymbol{x}_t$ is known is restrictive but this issue can be ameliorated by estimating it consistently and accommodating its influence on the null limit distribution of the QLR test statistic.

The inferential procedure developed by Cho et al. (2020c) does not provide a method to estimate the number of thresholds in the event that the null hypothesis of no threshold effect is rejected. This observation contributes to the motivation for Cho et al.'s (2020d) analysis of a more general setting of the TARDL process subject to $S-1$ threshold levels. The authors allow for the case in which k may differ from unity, such that there are $\boldsymbol{\tau}_{1*} < \boldsymbol{\tau}_{2*} < \ldots < \boldsymbol{\tau}_{S-1*}$ threshold levels, which means that the long-run relationship between y_t and $\boldsymbol{x}_t$ is captured by:

$$y_t = \gamma_* + \boldsymbol{\beta}_*^{(1)\prime}\boldsymbol{x}_t^{(1)} + \boldsymbol{\beta}_*^{(2)\prime}\boldsymbol{x}_t^{(2)} + \ldots + \boldsymbol{\beta}_*^{(S)\prime}\boldsymbol{x}_t^{(S)} + u_t,$$

where, for each $j = 1, 2, \ldots, S$, $\boldsymbol{x}_t^{(j)} = \sum_{t=0}^{t} \Delta\boldsymbol{x}_j^{(j)}$ with $\Delta\boldsymbol{x}_t^{(1)} := \Delta\boldsymbol{x}_t \circ \mathbb{1}\{\Delta\boldsymbol{x}_t <_\dagger \boldsymbol{\tau}_{1*}\}$, $\Delta\boldsymbol{x}_t^{(S)} := \Delta\boldsymbol{x}_t \circ \mathbb{1}\{\Delta\boldsymbol{x}_t \geq_\dagger \boldsymbol{\tau}_{S-1*}\}$, and for $s = 2, 3, \ldots, S-2$, $\Delta\boldsymbol{x}_t^{(s)} := \Delta\boldsymbol{x}_t \circ \mathbb{1}\{\boldsymbol{\tau}_{s-1*} \leq_\dagger \Delta\boldsymbol{x}_t <_\dagger \boldsymbol{\tau}_{s*}\}$. Here, $\circ$ denotes the Hadamard product, while the subscript '†' following an inequality sign denotes the element-by-element inequality between two vectors. This represents a straightforward generalisation of the long-run relationship between y_t and $\boldsymbol{x}_t$ introduced in Section 3.1 to the case of multiple thresholds. Cho et al. (2020d) consider the following error-correction form of this process, which generalises the TARDL(p, q) process (6):

$$\Delta y_t = \rho_* y_{t-1} + \sum_{s=1}^{S} \boldsymbol{\theta}_*^{(s)\prime}\boldsymbol{x}_{t-1}^{(s)} + \gamma_* + \sum_{j=1}^{p-1} \varphi_{j*}\Delta y_{t-j} + \sum_{j=0}^{q-1}\sum_{s=1}^{S} \boldsymbol{\pi}_{j*}^{(s)\prime}\Delta\boldsymbol{x}_{t-j}^{(s)} + \varepsilon_t^{(S)}, \tag{10}$$

and refer to this as the TARDL($\boldsymbol{\tau}_*$; p, q) process, where $\boldsymbol{\tau}_* := (\boldsymbol{\tau}'_{1*}, \ldots, \boldsymbol{\tau}'_{S-1*})'$. Note that, if $S = 2$, the TARDL($\boldsymbol{\tau}_*$; p, q) process reduces to the TARDL(p, q) process. Furthermore, if $S = 1$, it reduces to the linear ARDL(p, q) process.

The objective of Cho et al. (2020d) is to estimate the number of regimes, S. Not only does estimating the number of regimes provide a means to test the adequacy of models built around an assumed number of regimes (e.g. a NARDL model with two regimes defined using a single threshold value of zero), but it also provides a means to model phenomena that exhibit complex nonlinearities involving several unknown threshold values. For example, suppose that import prices exhibit both sign and size asymmetry with respect to movements in the domestic exchange rate. To capture the sign asymmetry, a threshold value of zero is required. Two additional threshold values, one positive and one negative, would allow for import prices to exhibit size asymmetry with respect to both appreciations and depreciations of the domestic exchange rate. The addition of further threshold values would allow for more complex nonlinear relationships.

Cho et al. (2020d) examine the properties of several popular information criteria by simulation, including the Akaike information criterion (AIC), the Schwarz information criterion (SIC) and the Hannan-Quinn information criterion (HQIC), as well as modified versions of each criterion defined following Pitarakis (2006), denoted pAIC, pSIC and pHQIC. For each $S = 1, 2, \ldots, \bar{S}$, Cho et al. (2020d) first estimate the error variance of (10) by least squares. Let:

$$\widehat{\sigma}^2_{n(S)} := \inf_{\boldsymbol{\lambda}} \frac{1}{T} \sum_{t=1}^{T} \left[\Delta y_t - \rho y_{t-1} - \sum_{s=1}^{S} \boldsymbol{\theta}^{(s)\prime} \boldsymbol{x}^{(s)}_{t-1}(\boldsymbol{\tau}) - \gamma - \sum_{j=1}^{p-1} \varphi_j \Delta y_{t-j} - \sum_{j=0}^{q-1} \sum_{s=1}^{S} \boldsymbol{\pi}^{(s)\prime}_j \Delta \boldsymbol{x}^{(s)}_{t-j}(\boldsymbol{\tau}) \right]^2,$$

where $\boldsymbol{\lambda} := (\rho, \boldsymbol{\theta}', \gamma, \boldsymbol{\varphi}', \boldsymbol{\pi}', \boldsymbol{\tau}')'$, $\boldsymbol{\theta} := (\boldsymbol{\theta}^{(1),\prime}, \ldots, \boldsymbol{\theta}^{(S)\prime})'$, $\boldsymbol{\varphi} := (\varphi_1, \ldots, \varphi_{p-1})'$, and $\boldsymbol{\pi} := (\boldsymbol{\pi}^{(1)\prime}_0, \ldots, \boldsymbol{\pi}^{(S)\prime}_0, \ldots, \boldsymbol{\pi}^{(1)\prime}_{q-1}, \ldots, \boldsymbol{\pi}^{(S)\prime}_{q-1})'$, which generalises the error variance estimators used for the QLR test statistic in the TARDL process. Next, the information criteria are computed as follows:

$$IC_{TARDL(\boldsymbol{\tau},p,q)} := \log(\widehat{\sigma}^2_{n(S)}) + \frac{c_T}{T} n_{TARDL} \quad \text{and} \quad IC^{\circ}_{TARDL(\boldsymbol{\tau},p,q)} := \log(\widehat{\sigma}^2_{n(S)}) + \frac{c_T}{T} n^{\circ}_{TARDL},$$

where c_T is a deterministic penalty term that satisfies $c_T/T \to 0$, n_{TARDL} is the number of parameters in the TARDL($\boldsymbol{\tau}$, p, q) process and n°_{TARDL} is the number of parameters not including $\boldsymbol{\tau}$. Akaike (1973), Schwarz (1978), and Hannan and Quinn (1979) set c_T equal to 2, $\log(T)$, and $2\log(\log(T))$, respectively, while Pitarakis (2006) prefers the use of $IC^{\circ}_{TARDL(\boldsymbol{\tau},p,q)}$ to $IC_{TARDL(\boldsymbol{\tau},p,q)}$.

Cho et al. (2020d) conduct extensive simulation studies to determine which of the six information criteria performs best. Setting $k = 1$, they separately examine the cases in which x_t has no time drift and where x_t is driven by a time drift. Note that if Δx_t has a non-zero population mean, then x_t is driven by a time drift. The simulations reveal that the relative performance of the information criteria is influenced by a combination of factors, including the sample size, the nature of the data generating process and whether or not x_t is driven by a time drift. In general, the standard SIC outperforms the other information criteria if Δx_t has zero population mean. However, if x_t is driven by a time drift, the picture is more complex. In this case, the modified SIC advocated by Pitarakis (2006) outperforms the other information criteria when T is small. However, as the sample size increases, the standard SIC and Pitarakis's SIC converge in terms of performance, with both dominating the other four information criteria. In addition, when the regime-specific parameters of the TARDL process are similar, such that the TARDL process is locally close to the simpler ARDL process, then Pitarakis's SIC has a tendency to outperform relative to all competitors, including the standard SIC.

4 ARDL Estimation by Quantile Regression

Following Xiao's (2009) seminal analysis of quantile cointegration, a growing theoretical and empirical literature has considered applications of quantile regression to integrated time series. Cho et al. (2015) were the first to bring quantile regression to the ARDL literature, yielding the quantile autoregressive distributed lag (QARDL) process. The QARDL process captures both the long-run and short-run relationships between y_t and $\boldsymbol{x}_t$ at any desired location in the conditional distribution. Not only does this allow researchers to focus on specific quantiles of interest but, by providing a framework to test the equality of the QARDL coefficients across quantiles, it provides a means to study locational asymmetry, which arises if the QARDL parameters vary across quantiles. Cho et al. (2015) illustrate the nature of locational asymmetry with an application to dividend policy in the US, showing that dividend smoothing is stronger and that the long-run target payout ratio is higher at the upper quantiles of the conditional distribution of dividends.

Starting with the ARDL model (1), for each $\xi \in (0,1)$, we may write:

$$y_t = \gamma_*(\xi) + \sum_{j=1}^{p} \phi_{j*}(\xi) y_{t-j} + \sum_{j=0}^{q} \boldsymbol{\theta}_{j*}(\xi)' \boldsymbol{x}_{t-j} + \varepsilon_t(\xi), \tag{11}$$

where $(\gamma_*(\xi), \phi_{1*}(\xi), \ldots, \phi_{p*}(\xi), \boldsymbol{\theta}_{0*}(\xi)', \ldots, \boldsymbol{\theta}_{q*}(\xi)')'$ is the quantile coefficient of the QARDL process, and $\varepsilon_t(\xi)$ is the quantile error, from which the linear quantile long-run relationship can be obtained as $y_t = \mu_*(\xi) + \boldsymbol{\beta}_*(\xi)' \boldsymbol{x}_t + u_t(\xi)$, where $\mu_*(\xi)$, $\boldsymbol{\beta}_*$, and $u_t(\xi)$ will be defined in Section 4.1, below. Estimation of the quantile coefficients enables the researcher to conduct inference on y_t and $\boldsymbol{x}_t$ at any desired conditional quantile. Cho et al. (2015) provide relevant theory on the estimation and inference on the parameters in (11) along with an empirical application to post-war dividend smoothing in the US.

Cho et al. (2020a) propose a quantile version of the NARDL process in (6). That is, for each $\xi \in (0,1)$, the quantile nonlinear autoregressive distributed lag (QNARDL) process is defined as:

$$y_t = \gamma_*(\xi) + \sum_{j=1}^{p} \phi_{j*}(\xi) y_{t-j} + \sum_{j=0}^{q} (\boldsymbol{\theta}_{j*}^{+}(\xi)' \boldsymbol{x}_{t-j}^{+} + \boldsymbol{\theta}_{j*}^{-}(\xi)' \boldsymbol{x}_{t-j}^{-}) + \varepsilon_t(\xi), \tag{12}$$

where $\boldsymbol{x}_t^+$ and $\boldsymbol{x}_t^-$ are the same as in (5) and the quantile coefficients $(\gamma_*(\xi), \phi_{1*}(\xi), \ldots, \phi_{p*}(\xi), \boldsymbol{\theta}_{0*}^{+}(\xi)',$ $\ldots, \boldsymbol{\theta}_{q*}^{+}(\xi)', \boldsymbol{\theta}_{0*}^{-}(\xi)', \ldots, \boldsymbol{\theta}_{q*}^{-}(\xi)')'$ can be used to explore quantile variation in the asymmetric relationship embodied by the NARDL process. The QNARDL process represents a synthesis of the QARDL and NARDL processes. For some coefficients $\boldsymbol{\beta}_*^+(\xi)$ and $\boldsymbol{\beta}_*^-(\xi)$ and a stationary variable $u_t(\xi)$ as defined in Section 4.2, the long-run relationship is given by $y_t = \boldsymbol{\beta}_*^+(\xi)' \boldsymbol{x}_t^+ + \boldsymbol{\beta}_*^-(\xi)' \boldsymbol{x}_t^- + u_t(\xi)$. As with the NARDL model, the use of partial sum processes introduces sign asymmetry in both the long-run and the short-run. Meanwhile, as with the QARDL model, the use of quantile regression allows for locational asymmetry. Continuing with the example of dividend policy from Cho et al. (2015), the switch from a QARDL specification to a QNARDL specification would allow for sign asymmetry with respect to earnings news coupled with locational asymmetry across the conditional distribution of dividends.

4.1 Estimation and Inference for the QARDL Model

Cho et al. (2015) note that the QARDL process (10) can be equivalently converted into two different forms. First, following Pesaran and Shin (1998), the QARDL process can be expressed as follows:

$$y_t = \gamma_*(\xi) + \boldsymbol{x}_t' \boldsymbol{\gamma}_*(\xi) + \sum_{j=1}^{p} \phi_{j*}(\xi) y_{t-j} + \sum_{j=0}^{q-1} \Delta \boldsymbol{x}_{t-j}' \boldsymbol{\delta}_{j*}(\xi) + \varepsilon_t(\xi), \tag{13}$$

where $\boldsymbol{\gamma}_*(\xi) := \sum_{j=0}^{q} \boldsymbol{\theta}_{j*}(\xi)$ and $\boldsymbol{\delta}_{j*}(\xi) := -\sum_{i=j+1}^{q} \boldsymbol{\theta}_{i*}(\xi)$. The long-run quantile relationship is $y_t = \mu_*(\xi) + \boldsymbol{x}_t' \boldsymbol{\beta}_*(\xi) + u_t(\xi)$, where $\mu_*(\xi) := \gamma_*(\xi) \left(1 - \sum_{i=1}^{p} \phi_{i*}(\xi)\right)^{-1}$, $\boldsymbol{\beta}_*(\xi) := \boldsymbol{\gamma}_*(\xi) \left(1 - \sum_{i=1}^{p} \phi_{i*}(\xi)\right)^{-1}$ and $u_t(\xi)$ is a stationary process defined by $\{\Delta \boldsymbol{x}_t, \varepsilon_t(\xi), \Delta \boldsymbol{x}_{t-1}, \varepsilon_{t-1}(\xi), \ldots\}$. Further, the following error-correction form can be derived from (13):

$$\Delta y_t = \gamma_*(\xi) + \rho_*(\xi)(y_{t-1} - \boldsymbol{\beta}_*(\xi)' \boldsymbol{x}_{t-1}) + \sum_{j=1}^{p-1} \varphi_{j*}(\xi) \Delta y_{t-j} + \sum_{j=0}^{q-1} \boldsymbol{\pi}_{j*}(\xi)' \Delta \boldsymbol{x}_{t-j} + \varepsilon_t(\xi), \tag{14}$$

where, for $\ell = 1, 2, \ldots, p-1$ and $j = 1, 2, \ldots, q-1$:

$$\rho_*(\xi) := \sum_{j=1}^{p} \phi_{j*}(\xi) - 1, \quad \boldsymbol{\pi}_{0*}(\xi) := \boldsymbol{\theta}_{0*}(\xi), \quad \varphi_{\ell*}(\xi) := -\sum_{i=\ell+1}^{p} \phi_{i*}(\xi), \quad \text{and} \quad \boldsymbol{\pi}_{j*} := -\sum_{i=j+1}^{p} \boldsymbol{\theta}_{*i}(\xi).$$

Cho et al. (2015) exploit these two different representations to estimate the long-run parameter, $\boldsymbol{\beta}_*(\xi)$, and the other short-run parameters in (14). Specifically, they first estimate the parameters in (13) by quantile regression following Koenker and Bassett (1978). That is, the quantile-specific parameters can be consistently estimated by $(\widetilde{\gamma}_T(\xi), \widetilde{\phi}_{T,1}(\xi), \ldots, \widetilde{\phi}_{T,p}(\xi), \widetilde{\boldsymbol{\theta}}_{T,1}(\xi)', \ldots, \widetilde{\boldsymbol{\theta}}_{T,q}(\xi)')'$, which minimises:

$$\sum_t \zeta_\xi \left(y_t - \gamma(\xi) - \sum_{j=1}^{p} \phi_j(\xi) y_{t-j} - \sum_{j=0}^{q} \boldsymbol{\theta}_j(\xi)' \boldsymbol{x}_{t-j} \right),$$

where $\zeta_\xi(\cdot)$ is the check function. Next, the quantile-specific long-run parameters are estimated by $\widehat{\boldsymbol{\beta}}_T(\xi) := \left(\sum_{j=0}^{q} \widetilde{\boldsymbol{\theta}}_{T,j}(\xi)\right) \left(1 - \sum_{i=1}^{p} \widetilde{\phi}_{T,i}(\xi)\right)^{-1}$. The authors show that $\widehat{\boldsymbol{\beta}}_T(\xi)$ is T-consistent and asymptotically follows a mixed normal distribution. Estimating the short-run parameters in (14) is then straightforward. After replacing $\boldsymbol{\beta}_*(\xi)$ with $\widehat{\boldsymbol{\beta}}_T(\xi)$, the following is optimised with respect to the short-run parameters:

$$\sum_t \zeta_\xi \left(\Delta y_t - \gamma(\xi) + \rho(\xi)(y_{t-1} - \widehat{\boldsymbol{\beta}}_T(\xi)' \boldsymbol{x}_{t-1}) + \sum_{j=1}^{p-1} \varphi_j(\xi) \Delta y_{t-j} + \sum_{j=0}^{q-1} \boldsymbol{\pi}_j(\xi)' \Delta \boldsymbol{x}_{t-j} \right).$$

Hence, the short-run parameters in (14) can be consistently estimated at rate $\sqrt{T}$ and are asymptotically normal.

In addition to the conditions given by Pesaran and Shin (1998), Cho et al. (2015) assume that $\mathbb{E}[\Delta \boldsymbol{x}_t] = \boldsymbol{0}$ and that ε_t has a continuous and finite marginal probability density function so that quantile regression can be applied. Furthermore, they assume that $\boldsymbol{z}_t := [\Delta \boldsymbol{x}_t', \varepsilon_t(\xi), \zeta_\xi(\varepsilon_t(\xi))]'$ is sufficiently regular to apply the functional central limit theorem to the series that is linearly transformed from $\{\ldots, \boldsymbol{z}_{t-1}', \boldsymbol{z}_t', \boldsymbol{z}_{t+1}', \ldots\}$ with linear coefficients forming an absolutely converging series. Under these assumptions, the authors show that the limit distributions of the short-run and long-run parameter estimator are asymptotically normal and a functional of Brownian motions, respectively.

Cho et al. (2015) extend the scope of the QARDL model by letting $\boldsymbol{\xi}$ be a set of multiple quantile levels and examine the large sample properties of the long-run parameter estimators for $\boldsymbol{\xi}$. By extending the univariate quantile structure to the multivariate level, Cho et al. (2015) show that the long-run parameter estimator converges to the unknown long-run parameter at rate T, even when the long-run parameters are estimated for multiple quantile levels. Furthermore, the authors demonstrate that the joint limit distribution of the long-run parameter estimator is mixed-normal, such that a test statistic constructed using the Wald (1943) testing principle asymptotically follows a χ^2 distribution under the null.

4.2 Estimation and Inference for the QNARDL Model

The QNARDL process can be written in error correction form as:

$$\begin{aligned}\Delta y_t = {} & \rho_*(\xi) y_{t-1} + \boldsymbol{\theta}_*^+(\xi)' \boldsymbol{x}_{t-1}^+ + \boldsymbol{\theta}_*^-(\xi)' \boldsymbol{x}_{t-1}^- + \gamma_*(\xi) \\ & + \sum_{j=1}^{p-1} \varphi_{j*}(\xi) \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^+(\xi)' \Delta \boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_{j*}^-(\xi)' \Delta \boldsymbol{x}_{t-j}^- \right) + \varepsilon_t(\xi), \end{aligned} \tag{15}$$

for each $\xi \in (0,1)$, where the definition of the QNARDL quantile coefficients is similar to the NARDL case and so, for brevity, we do not repeat it here. Letting $u_{t-1}(\xi) := y_{t-1} - \boldsymbol{\beta}_*^+(\xi)' \boldsymbol{x}_{t-1}^+ - \boldsymbol{\beta}_*^-(\xi)' \boldsymbol{x}_{t-1}^-$ be the quantile cointegration error, we can re-write (15) as follows:

$$\Delta y_t = \rho_*(\xi) u_{t-1}(\xi) + \gamma_*(\xi) + \sum_{j=1}^{p-1} \varphi_{j*}(\xi) \Delta y_{t-j} + \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_{j*}^+(\xi)' \Delta \boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_{j*}^-(\xi)' \Delta \boldsymbol{x}_{t-j}^- \right) + \varepsilon_t(\xi), \tag{16}$$

where $\boldsymbol{\beta}_*^+(\xi) := -\boldsymbol{\theta}_*^+(\xi)/\rho_*(\xi)$ and $\boldsymbol{\beta}_*^-(\xi) := -\boldsymbol{\theta}_*^-(\xi)/\rho_*(\xi)$. This error correction form of the QNARDL process enables the researcher to estimate the unknown coefficients in a manner parallel to the NARDL case outlined above. However, if the unknown long-run parameters are estimated by quantile regression directly, then the singularity problem arises. Thus, Cho et al. (2020a) recommend that the QNARDL process should be re-parameterised prior to estimation. Specifically, for each $\xi \in (0,1)$, let the long-run relationship be represented as follows:

$$y_t = \varsigma_*(\xi) + \boldsymbol{\lambda}_*(\xi)' \boldsymbol{x}_t^+ + \boldsymbol{\eta}_*(\xi)' \boldsymbol{x}_t + u_t(\xi),$$

where $\boldsymbol{x}_t \equiv \boldsymbol{x}_t^+ + \boldsymbol{x}_t^-$. The long-run parameters $(\boldsymbol{\lambda}_*(\xi)', \boldsymbol{\eta}_*(\xi)')'$ can be estimated by quantile regression for $k=1$. With this estimator denoted by $(\widehat{\boldsymbol{\lambda}}_T(\xi)', \widehat{\boldsymbol{\eta}}_T(\xi)')'$, the long-run parameters are obtained by $\widehat{\boldsymbol{\beta}}_T^+(\xi) := \widehat{\boldsymbol{\lambda}}_T(\xi) + \widehat{\boldsymbol{\eta}}_T(\xi)$ and $\widehat{\boldsymbol{\beta}}_T^-(\xi) := \widehat{\boldsymbol{\eta}}_T(\xi)$. Under this re-parameterisation, the singularity problem does not arise and the long-run parameter estimator converges to the true parameter value at rate T. Consequently, the short-run parameters in (16) can be estimated by replacing u_{t-1} with $\widehat{u}_{t-1}(\xi) := y_{t-1} - \widehat{\boldsymbol{\beta}}_T^+(\xi)' \boldsymbol{x}_{t-1}^+ -$

$\widehat{\boldsymbol{\beta}}_T^-(\xi)\boldsymbol{x}_{t-1}^-$ and then applying quantile regression. That is, one can consistently estimate the short-run parameters by optimising:

$$\sum_t \zeta_\xi \left(\Delta y_t - \rho(\xi)\widehat{u}_{t-1}(\xi) - \gamma(\xi) - \sum_{j=1}^{p-1} \varphi_j(\xi)\Delta y_{t-j} - \sum_{j=0}^{q-1} \left(\boldsymbol{\pi}_j^+(\xi)'\Delta \boldsymbol{x}_{t-j}^+ + \boldsymbol{\pi}_j^-(\xi)'\Delta \boldsymbol{x}_{t-j}^- \right) \right),$$

with respect to the short-run parameters.

Cho et al. (2020a) note that it is not straightforward to derive the limiting distribution of the long-run parameter estimator, which hampers hypothesis testing. Consequently, in keeping with the estimation procedure advanced by Cho et al. (2019), Cho et al. (2020a) first estimate $(\boldsymbol{\lambda}_*(\xi)', \boldsymbol{\eta}_*(\xi)')'$ by applying the FM-OLS estimator of Phillips and Hansen (1990) to the quantile regression, and next estimate the long-run parameters as described above. The resulting long-run parameter estimator is T-consistent, and asymptotically follows a mixed normal distribution. Therefore, test statistics constructed using the Wald (1943) principle converge to the standard χ^2 distribution. Consequently, inference on the QNARDL model proceeds in a similar manner to the NARDL model.

The regularity conditions assumed by Cho et al. (2020a) are almost identical to those in Cho et al. (2019). The main difference stems from the fact that parameter estimation is conducted by minimising the check function. Instead of imposing regularity conditions formed by regression errors, the authors assume the regularity conditions formed by the check function applied to the regression error and impose the regularity conditions in parallel to Cho et al. (2019). Lastly, if $k > 1$, one can estimate the trend models first and next estimate the long and short-run equations sequentially, by analogy to Cho et al. (2020b). Due to the parallel structure of these procedures, we do not repeat the discussion here in the interest of brevity.

5 Panel Data Extensions of the ARDL Model

In addition to the asymmetric and nonlinear ARDL variants discussed above, ARDL specifications and their derivatives have also been applied in the context of dynamic panel data analysis. Two simple estimators for panels with many time series observations for each of N groups are the traditional pooled estimators (including the fixed and random effects estimators) and the Mean Group (MG) estimator of Pesaran and Smith (1995). The first approach involves pooling the data together and estimating a single model under the assumption that all of the model parameters aside from the intercepts are homogeneous across groups. At the other extreme, the MG estimator of Pesaran and Smith (1995) involves estimating N separate models, where all of the parameters are allowed to vary over groups. To summarise the distribution of these group-specific parameters, one may simply take their mean value across groups. MG estimation of ARDL and NARDL models has been applied in the analysis of exchange rate pass-through into import prices by Brun-Aguerre et al. (2017).

Both the pooled and MG frameworks have limitations; the assumption of parameter homogeneity that is central to the pooled estimators is restrictive, while the MG estimator does not make use of any homogeneity restrictions, even if they are valid. This leads Pesaran et al. (1999) to pursue an intermediate approach that they refer to as the Pooled Mean Group (PMG) estimator, which mimics the structure of the ARDL model in a panel setting under the assumption of long-run homogeneity, while allowing the short-run parameters and error variances to differ across groups. The authors note that there are often good reasons to believe that equilibrium relations should be common across groups but that the same is not typically true of short-run dynamic parameters. This reasoning justifies their pursuit of a hybrid approach to estimation that makes use of pooling for the estimation of long-run parameters and averaging for the estimation of short-run parameters. Consequently, the interpretation of the PMG model follows easily from that of the linear ARDL model of Pesaran and Shin (1998), simply transferred to a panel data setting under long-run homogeneity. Due to its intuitive structure and relative ease of implementation, the PMG estimator has been highly influential, with approximately 4,000 citations to date according to Google Scholar.

5.1 Estimation and Inference for the PMG Model

Suppose that one wishes to estimate an $ARDL\,(p, q, q, \ldots, q)$ model of the following form using a panel dataset with groups indexed by $i = 1, 2, \ldots, N$ and time periods indexed by $t = 1, \ldots, T$, where T is sufficiently large to allow the model to be estimated consistently for each group:[2]

$$y_{it} = \sum_{j=1}^{p} \lambda_{ij*} y_{i,t-j} + \sum_{j=0}^{q} \boldsymbol{\delta}'_{ij*} \boldsymbol{x}_{i,t-j} + \boldsymbol{\gamma}'_{i*} \boldsymbol{d}_t + \varepsilon_{it}, \tag{17}$$

where $\boldsymbol{x}_{it}$ and $\boldsymbol{d}_t$ are $k \times 1$ and $s \times 1$ vectors of regressors, respectively, while the λ_{ij*}s are unknown scalars and the $\boldsymbol{\delta}_{ij*}$s and $\boldsymbol{\gamma}_{i*}$s are $k \times 1$ and $s \times 1$ vectors of unknown parameters to be estimated. Note that (17)

[2] It is straightforward to allow for different lag orders associated with each of the variables in $\boldsymbol{x}_{it}$.

can be re-parameterised as follows:

$$\Delta y_{it} = \phi_{i*} y_{i,t-1} + \boldsymbol{\beta}'_{i*} \boldsymbol{x}_{it} + \sum_{j=1}^{p-1} \lambda^*_{ij} \Delta y_{i,t-j} + \sum_{j=0}^{q-1} \boldsymbol{\delta}^{*\prime}_{ij} \Delta \boldsymbol{x}_{i,t-j} + \boldsymbol{\gamma}'_{i*} \boldsymbol{d}_t + \varepsilon_{it}, \tag{18}$$

where $\phi_{i*} := -(1 - \sum_{j=1}^{p} \lambda_{ij*})$, $\boldsymbol{\beta}_{i*} := \sum_{j=0}^{q} \boldsymbol{\delta}_{ij*}$, $\lambda^*_{ij} := -\sum_{m=j+1}^{p} \lambda_{im*}$, $j = 1, \dots, p-1$ and $\boldsymbol{\delta}^*_{ij} := -\sum_{m=j+1}^{q} \boldsymbol{\delta}_{im*}$, $j = 1, \dots, q-1,\ i = 1, \dots, N$. By stacking the time series observations for each group to obtain $\boldsymbol{y}_i := (y_{i1}, \dots, y_{iT})'$ and $\mathbf{X}_i := (\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT})'$, (18) can be re-written as follows:

$$\Delta \boldsymbol{y}_i = \phi_{i*} \boldsymbol{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta}_{i*} + \sum_{j=1}^{p-1} \lambda^*_{ij} \Delta \boldsymbol{y}_{i,-j} + \sum_{j=0}^{q-1} \Delta \mathbf{X}_{i,-j} \boldsymbol{\delta}^*_{ij} + \mathbf{D} \gamma_{i*} + \boldsymbol{\varepsilon}_i, \tag{19}$$

where $\mathbf{D} := (\boldsymbol{d}_1, \dots, \boldsymbol{d}_T)'$ is a $T \times s$ matrix of observations on the deterministic regressors, such as intercepts and time trends, and $\boldsymbol{y}_{i,-1}$ and $\Delta \mathbf{X}_{i,-j}$ are $T \times 1$ and $T \times k$ matrices obtained by stacking $y_{i,t-1}$ and $\boldsymbol{x}_{i,t-j}$, respectively.

The long-run coefficients on $\mathbf{X}_i$ can be obtained as $\boldsymbol{\theta}_{i*} = -\boldsymbol{\beta}_{i*}/\phi_{i*}$. Pesaran et al. (1999) assume long-run homogeneity, such that $\boldsymbol{\theta}_{i*} = \boldsymbol{\theta}_*$ for every $i = 1, \dots, N$. Consequently, the long-run relationship for every cross-section unit has the same structure as the long-run relationship embedded in the ARDL model of Pesaran and Shin (1998). This allows (19) to be re-written compactly as:

$$\Delta \boldsymbol{y}_i = \phi_{i*} \boldsymbol{\xi}_i(\boldsymbol{\theta}_*) + \mathbf{W}_i \boldsymbol{\kappa}_{i*} + \boldsymbol{\varepsilon}_i, \qquad i = 1, \dots, N, \tag{20}$$

where:

$$\boldsymbol{\xi}_i(\boldsymbol{\theta}_*) := \boldsymbol{y}_{i,-1} - \mathbf{X}_i \boldsymbol{\theta}_* \quad i = 1, \dots, N,$$

is the error correction component, $\mathbf{W}_i = (\Delta \boldsymbol{y}_{i,-1}, \dots, \Delta \boldsymbol{y}_{i,-p+1}, \Delta \mathbf{X}_i, \Delta \mathbf{X}_{i,-1}, \dots, \Delta \mathbf{X}_{i,-q+1}, \mathbf{D})$, and $\boldsymbol{\kappa}_{i*} := (\lambda^*_{i1}, \dots, \lambda^*_{i,p-1}, \boldsymbol{\delta}^{*\prime}_{i0}, \boldsymbol{\delta}^{*\prime}_{i1}, \dots, \boldsymbol{\delta}^{*\prime}_{i,q-1}, \boldsymbol{\gamma}'_{i*})'$.

Estimation of (20) is complicated by three factors: (i) the equation for each group is nonlinear in ϕ_{i*} and $\boldsymbol{\theta}_*$; (ii) the long-run homogeneity assumption introduces cross-equation parameter restrictions; and (iii) the error variances exhibit heterogeneity across groups. Pesaran et al. (1999) propose a maximum likelihood estimation framework in which the homogeneous long-run parameters are estimated by pooling, while group-wide mean estimates of the heterogeneous short-run parameters and error-correction coefficients are obtained by taking averages across groups. Hence the nomenclature "pooled mean group" estimation.

Pesaran et al. (1999) develop separate asymptotic theories for the PMG estimators in the case that the regressors, $\boldsymbol{x}_{it}$, are stationary and non-stationary. Under stationarity, the authors demonstrate the consistency and asymptotic normality of the PMG estimators and show that the long- and short-run parameter estimators share a common convergence rate of $\sqrt{T}$. By contrast, under the assumption that the regressors are first-order integrated processes, the asymptotic analysis is complicated by the fact that the ML estimators of the long-run and the short-run parameters converge to their true values at different rates (T and $\sqrt{T}$, respectively). In this case, for a fixed N and as $T \to \infty$, the PMG estimator asymptotically follows a mixed-normal distribution.

These results are obtained under standard conditions on the behaviour of the disturbance terms, as well conditions that ensure the dynamic stability of the model and allow for identification. By analogy to the ARDL model of Pesaran and Shin (1998), in the case of non-stationarity, Pesaran et al. (1999) further assume that the variables in $\boldsymbol{x}_{it}$ are not cointegrated among themselves. The strongest assumption that is invoked is the long-run homogeneity condition from which the PMG estimator draws its name. Not only is this assumption testable using Hausman (1978)-type tests but it is defensible in many practical applications, as heterogeneity across groups may often be confined to the dynamic parameters and the error variances. The authors further note that, in practice, the most challenging aspect of working with the PMG estimator is often the interpretation of the heterogeneity, as it can be difficult to exclude the possibility of group-specific omitted variables and/or measurement errors correlated with the regressors.

6 The Spatio-Temporal ARDL Model

The final ARDL variant that we consider is another panel data implementation of ARDL, this time motivated by the increasing availability of large spatial time series datasets. The spatio-temporal autoregressive distributed lag (STARDL) model proposed by Shin and Thornton (2019) extends the popular spatial dynamic panel data model by allowing both spatial and temporal coefficients to differ across spatial units. The STARDL(p, q) model is specified as follows:

$$y_{it} = \sum_{\ell=1}^{p} \phi_{i\ell *} y_{i,t-\ell} + \sum_{\ell=0}^{p} \phi_{i\ell}^{*} y_{i,t-\ell}^{*} + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell *}' \boldsymbol{x}_{i,t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\pi}_{i\ell}^{*\prime} \boldsymbol{x}_{i,t-\ell}^{*} + \alpha_{i*} + u_{it}, \tag{21}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where y_{it} is a scalar dependent variable for the ith spatial unit at time t, $\boldsymbol{x}_{it} = (x_{it}^{1}, \ldots, x_{it}^{K})'$ is a $K \times 1$ vector of exogenous regressors, $\boldsymbol{\pi}_{i\ell *} = (\pi_{i\ell *}^{1}, \ldots, \pi_{i\ell *}^{K})'$ is a $K \times 1$ vector of parameters and likewise for $y_{i,t-\ell}$ and $\boldsymbol{x}_{i,t-\ell}$ and their associated parameters. Spatial interactions between units, both contemporaneously and with lags, are captured via the spatial variables, y_{it}^* and $\boldsymbol{x}_{it}^*$, defined as:

$$y_{it}^* \equiv \sum_{j=1}^{N} w_{ij} y_{jt} = \boldsymbol{w}_i' \boldsymbol{y}_t \text{ with } \underset{N\times 1}{\boldsymbol{y}_t} = (y_{1t}, \ldots, y_{Nt})', \tag{22}$$

$$\underset{K\times 1}{\boldsymbol{x}_{it}^*} = \left(x_{it}^{1*}, \ldots, x_{it}^{K*}\right)' \equiv \left(\sum_{j=1}^{N} w_{ij} x_{jt}^1, \ldots, \sum_{j=1}^{N} w_{ij} x_{jt}^K\right)' = \left(\boldsymbol{w}_i' \otimes \boldsymbol{I}_K\right) \boldsymbol{x}_t; \underset{NK\times 1}{\boldsymbol{x}_t} = \begin{pmatrix} \boldsymbol{x}_{1t} \\ \vdots \\ \boldsymbol{x}_{Nt} \end{pmatrix}, \tag{23}$$

where $\boldsymbol{w}_i' = (w_{i1}, \ldots, w_{iN})$ denotes a $1 \times N$ vector of non-stochastic spatial weights, which are assumed to be known a priori, with $w_{ii} = 0$.

Stacking the N individual STARDL (p, q) equations in (21), Shin and Thornton (2019) obtain the following spatial system:

$$\boldsymbol{y}_t = \sum_{\ell=1}^{p} \boldsymbol{\Phi}_{\ell *} \boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{p} \boldsymbol{\Phi}_{\ell}^* \mathbf{W} \boldsymbol{y}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_{\ell *} \boldsymbol{x}_{t-\ell} + \sum_{\ell=0}^{q} \boldsymbol{\Pi}_{\ell}^* \left(\mathbf{W} \otimes \mathbf{I}_K\right) \boldsymbol{x}_{t-\ell} + \boldsymbol{\alpha}_* + \boldsymbol{u}_t, \tag{24}$$

where $\mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_N)'$ denotes the $N \times N$ spatial weights matrix, $\boldsymbol{\alpha}_* = (\alpha_{1*}, \ldots, \alpha_{N*})'$, $\boldsymbol{\Phi}_{\ell *} = diag(\phi_{1\ell *}, \cdots, \phi_{N\ell *})$ for $\ell = 1, \ldots, p$ and $\boldsymbol{\Phi}_{\ell}^* = diag(\phi_{1\ell}^*, \cdots, \phi_{N\ell}^*)$ for $\ell = 1, \ldots, p$, while $\boldsymbol{\Pi}_{\ell *} = diag(\boldsymbol{\pi}_{1\ell *}', \cdots, \boldsymbol{\pi}_{N\ell *}')$ and $\boldsymbol{\Pi}_{\ell}^* = diag(\boldsymbol{\pi}_{1\ell *}', \cdots, \boldsymbol{\pi}_{N\ell *}')$ for $\ell = 0, 1, \ldots, q$. The STARDL model is both relatively simple to estimate and also highly adaptable, as it nests several popular spatial dynamic panel data models, including the dynamic spatial Durbin model analysed by Lee and Yu (2010) and Elhorst (2014) and the heterogeneous spatial autoregressive panel data model of Aquaro et al. (2021).

6.1 Estimation and Inference for the STARDL Model

Shin and Thornton (2019) propose both quasi-maximum likelihood (QML) and control function (CF) based instrumental variables estimators. Both estimators are $\sqrt{T}$-consistent at the individual level and both asymptotically follow normal distributions. The QML estimation algorithm is standard and similar to that recently proposed by Aquaro et al. (2021). For the CF based estimator, Shin and Thornton (2019) begin by re-writing (21) as follows:

$$y_{it} = \phi_{i0}^* y_{it}^* + \boldsymbol{\theta}_i' \boldsymbol{\chi}_{it} + u_{it},$$

where $\boldsymbol{\chi}_{it} = (y_{i,t-1}, \ldots, y_{i,t-p}, y_{i,t-1}, \ldots, y^*_{i,t-p}, \boldsymbol{x}'_{it}, \ldots, \boldsymbol{x}'_{i,t-q}, \boldsymbol{x}^{*\prime}_{it}, \ldots, \boldsymbol{x}^{*\prime}_{i,t-q}, 1)'$ and $\boldsymbol{\theta}_i$ denotes the vector of associated parameters. Estimation proceeds in two steps. In the first step, one obtains the reduced form residuals:

$$\hat{v}_{it} = y^*_{it} - \boldsymbol{\psi}'_i \boldsymbol{z}_{it},$$

where $\boldsymbol{z}_{it}$ are instruments obtained internally from the model.[3] In the second step, one estimates the following by OLS:

$$y_{it} = \phi^*_{i0} y^*_{it} + \boldsymbol{\theta}'_i \boldsymbol{\chi}_{it} + \rho_i \hat{v}_{it} + u_{it}.$$

The authors show that this procedure yields consistent and asymptotically normal estimates of the STARDL parameters.

The properties of both the QML and CF based estimators are obtained under standard conditions on the behaviour of the disturbance terms, as well as identification conditions and conditions related to dynamic stability of the model. The variables entering the model are assumed to be stationary for reasons of expediency; the extension to the case of non-stationarity in the context of spatial data would bring with it several challenges and would represent a substantial contribution to the literature. The disturbance terms are assumed to be serially uncorrelated but need not be i.i.d.. In particular, they are allowed to exhibit heteroskedasticity. The serial correlation assumption is not onerous, as the inclusion of sufficient lags in the STARDL model should typically be sufficient to capture the autocorrelation structure in the data. The spatial weights are assumed to be non-stochastic and are subject to common normalisation conventions, while an addition condition limiting the degree of contemporaneous spatial feedback is required to ensure that the variance of the disturbance terms is finite. These conditions are in widespread use in the theoretical literature on spatial panel data models and are not unduly restrictive.

Spatial models, including the STARDL model, have an implicit network structure and so many of the popular tools of network analysis, including centrality statistics and clustering algorithms, can be used to facilitate their interpretation. Their use is particularly advantageous in the context of the STARDL model, where spatial dynamic interactions are governed by an array of parameters, unlike simple homogeneous parameter models in which interest primarily centres on a single spatial parameter. Shin and Thornton (2019) make use of two quantities for this purpose: (i) individual spatio-temporal dynamic multipliers based

[3] For example, one may obtain the instrumental variables internally from the spatially lagged exogenous and predetermined variables, e.g. $(\mathbf{W}^2\boldsymbol{y}_{t-1}, \mathbf{W}^2\boldsymbol{x}_t, \cdots)$.

on work by Shin et al. (2014); and (ii) system diffusion multipliers. The diffusion multipliers represent a novel technique to measure the joint impacts of $\boldsymbol{x}_t$ on $\boldsymbol{y}_{t+h}$ in space and time for $h = 0, 1, 2, \ldots$. To obtain the diffusion multipliers, it is first necessary to re-write (24) as follows:

$$\tilde{\boldsymbol{\Phi}}(L)\,\boldsymbol{y}_t = \tilde{\boldsymbol{\Pi}}(L)\,\boldsymbol{x}_t + \tilde{\boldsymbol{u}}_t, \tag{25}$$

where $\widetilde{\boldsymbol{\Phi}}(L) = \mathbf{I}_N - \sum_{\ell=1}^{p} \widetilde{\boldsymbol{\Phi}}_{\ell*} L^{\ell}$, $\widetilde{\boldsymbol{\Pi}}(L) = \sum_{\ell=0}^{q} \widetilde{\boldsymbol{\Pi}}_{\ell*} L^{\ell}$, $\tilde{\boldsymbol{u}}_t = (\boldsymbol{I}_N - \boldsymbol{\Phi}_0^{*}\boldsymbol{W}(L))^{-1}\boldsymbol{u}_t$ and $\widetilde{\boldsymbol{\Phi}}_{\ell*}$ and $\widetilde{\boldsymbol{\Pi}}_{\ell*}$ are the coefficients of the endogenous and exogenous variables obtained from (24), respectively. Premultiplying (25) by $\left[\widetilde{\boldsymbol{\Phi}}(L)\right]^{-1}$, one obtains:

$$\boldsymbol{y}_t = \mathbf{B}(L)\,\boldsymbol{x}_t + \left[\widetilde{\boldsymbol{\Phi}}(L)\right]^{-1}\widetilde{\boldsymbol{u}}_t,$$

where $\mathbf{B}(L) := \left[\widetilde{\boldsymbol{\Phi}}(L)\right]^{-1}\widetilde{\boldsymbol{\Pi}}(L) = \sum_{j=0}^{\infty} \mathbf{B}_{j*} L^{j}$ and where the diffusion multipliers, $\mathbf{B}_{j*}$ for $j = 0, 1, \ldots$, can be evaluated as follows:

$$\mathbf{B}_{j*} = \widetilde{\boldsymbol{\Phi}}_{1*}\mathbf{B}_{j-1*} + \widetilde{\boldsymbol{\Phi}}_{2*}\mathbf{B}_{j-2*} + \cdots + \widetilde{\boldsymbol{\Phi}}_{j-1*}\mathbf{B}_{1*} + \widetilde{\boldsymbol{\Phi}}_{j*}\mathbf{B}_{0*} + \widetilde{\boldsymbol{\Pi}}_{j*},\ j = 1, 2, \ldots,$$

where $\mathbf{B}_{0*} := \widetilde{\boldsymbol{\Pi}}_{0*}$ and $\mathbf{B}_{j*} = \mathbf{0}$ for $j < 0$ by construction. The $N \times NK$ matrix of cumulative diffusion multiplier effects is given by:

$$\boldsymbol{d}_{x*}^{H} = \sum_{h=0}^{H} \frac{\partial \boldsymbol{y}_{t+h}}{\partial \boldsymbol{x}_t'} = \sum_{h=0}^{H} \mathbf{B}_{h*},\ H = 0, 1, 2, \ldots$$

As with the use of cumulative dynamic multipliers in the NARDL literature, the spatio-temporal dynamic multipliers and diffusion multipliers proposed by Shin and Thornton (2019) can be used to illuminate the traverse to equilibrium following a perturbation to the system, thereby providing an intuitive insight into both the dynamic and long-run properties of the STARDL model. Shin and Thornton (2019) leverage recent developments in the econometric analysis of networks and connectedness due to Diebold and Yilmaz (2014) and Greenwood-Nimmo, Nguyen, and Shin (in press) to provide a simple and intuitive summary of the impacts of $\{x_{jt}\}_{j=1}^{N}$ on $\{y_{it}\}_{i=1}^{N}$. In this way, Shin and Thornton show that the diffusion multipliers can be used to obtain two simple and easily interpreted measures of the role of each node within the network: (i) its *external motivation*, which captures the extent and direction to which each node is steered by the network;

and (ii) its *systemic influence*, which captures the relative importance of each node within the network. The authors apply their framework to analyse the effect of enemy casualties on civilian deaths across the 18 Governorates of Iraq in the aftermath of the 2003 invasion. Their results indicate that violence is used both as a means to maintain influence in emergent political institutions and also in reprisal for past acts.

7 Concluding Remarks

In this paper, we survey the literature on the ARDL model. Given the existence of several excellent surveys focusing on the case of stationary distributed lags (e.g. Griliches, 1967; Nerlove, 1972; Hendry et al., 1984; Wickens and Breusch, 1988), our starting point is the more challenging setting in which the ARDL specification is applied to cointegrated non-stationary time series (Pesaran and Shin, 1998) or time series with mixed orders of integration (Pesaran et al., 2001). We present several recent extensions of this model, including the NARDL and TARDL models associated with Shin et al. (2014) and Cho et al. (2019, 2020b,c,d), the QARDL and QNARDL models developed by Cho et al. (2015) and Cho et al. (2020a), the pooled mean group panel data estimator of Pesaran et al. (1999) and the spatio-temporal ARDL model proposed by Shin and Thornton (2019).

The large body of research that we survey above highlights the adaptability of the ARDL specification but it should not be taken as evidence that all of the worthwhile avenues for development have already been explored. There are three areas in particular that have yet to be developed and that hold considerable promise. The first is to develop theoretical and applied methods that combine the ARDL and NARDL specifications with other popular regime-switching mechanisms, such as Markov-switching and smooth transition models. The second objective is the development of a system extension of the NARDL/TARDL model, which would provide a valuable framework for the analysis of asymmetric and nonlinear phenomena in multivariate systems. The last is the development of new panel ARDL models that can jointly accommodate both spatial and factor dependence. Developments in this area promise substantial contributions to the fast-growing literature on the unified modelling of cross-section dependence in panels (e.g. Bai and Li, 2013; Shi and Lee, 2017; Kuersteiner and Prucha, 2020; Chen, Shin, and Zheng, 2020).

References

AKAIKE, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle," in *Proceedings of the Second International Symposium on Information Theory*, ed. by B. N. Petrov and F. Csáki, Budapest: Akademiai Kiado, 267–281.

ALMON, S. (1965): "The Distributed Lag Between Capital Appropriations and Expenditures," *Econometrica*, 33, 178–196.

AQUARO, M., N. BAILEY, AND M. H. PESARAN (2021): "Estimation and Inference for Spatial Models with Heterogeneous Coefficients: An Application to US House Prices," *Journal of Applied Econometrics*, 36, 18–44.

BAEK, Y. I., J. S. CHO, AND P. C. PHILLIPS (2015): "Testing Linearity Using Power Transforms of Regressors," *Journal of Econometrics*, 187, 376–384.

BAI, J. AND K. LI (2013): "Spatial Panel Data Models with Common Shocks," MPRA Paper 52786, University Library of Munich, Germany.

BOX, M., K. GRATZER, AND X. LIN (2018): "The Asymmetric Effect of Bankruptcy Fraud in Sweden: A Long-Term Perspective," *Journal of Quantitative Criminology*, in press.

BRUN-AGUERRE, R. X., A.-M. FUERTES, AND M. J. GREENWOOD-NIMMO (2017): "Heads I Win; Tails You Lose: Asymmetry In Exchange Rate Pass-through into Import Prices," *Journal of the Royal Statistical Society Series A*, 180, 587–612.

CHEN, J., Y. SHIN, AND C. ZHENG (2020): "Estimation and Inference in Heterogeneous Spatial Panel Data Models with a Multifactor Error Structure," University of York Discussion Papers in Economics 20/03, York.

CHEONG, T., J. S. CHO, AND H. WHITE (2011): "Experience with the Weighted Bootstrap in Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Economic Theory and Econometrics*, 22, 248–260.

CHO, J. S., M. J. GREENWOOD-NIMMO, T.-H. KIM, AND Y. SHIN (2020a): "Hawks, Doves and

Asymmetry in US Monetary Policy: Evidence from a Dynamic Quantile Regression Model," Mimeo: University of York.

CHO, J. S., M. J. GREENWOOD-NIMMO, AND Y. SHIN (2019): "Two-Step Estimation of the Nonlinear Autoregressive Distributed Lag Model," Working Paper 2019rwp-154, Yonsei University, Seoul.

——— (2020b): "Estimating the Nonlinear Autoregressive Distributed Lag Model for Time Series Data with Drifts," Mimeo: University of York.

——— (2020c): "Testing for the Threshold Autoregressive Distributed Lag Model," Mimeo: University of York.

——— (2020d): "The Threshold Autoregressive Distributed Lag Model," Mimeo: University of York.

CHO, J. S. AND I. ISHIDA (2012): "Testing for the Effects of Omitted Power Transformations," *Economics Letters*, 117, 287–290.

CHO, J. S., I. ISHIDA, AND H. WHITE (2011): "Revisiting Tests for Neglected Nonlinearity Using Artificial Neural Networks," *Neural Computation*, 23, 1133–1186.

——— (2014): "Testing for Neglected Nonlinearity Using Twofold Unidentified Models under the Null and Hexic Expansions," in *Essays on Nonlinear Time Series Econometrics: A Festschrift in Honor of Timo Teräsvirta*, ed. by N. Haldrup, M. Meitz, and P. Saikkonen, Oxford: Oxford University Press, chap. 1, 3–27.

CHO, J. S., T.-H. KIM, AND Y. SHIN (2015): "Quantile Cointegration in the Autoregressive Distributed Lag Modeling Framework," *Journal of Econometrics*, 188, 281–300.

CHO, J. S. AND P. C. PHILLIPS (2018): "Sequentially Testing Polynomial Model Hypothesis Using the Power Transform of Regressors," *Journal of Applied Econometrics*, 33, 141–159.

CHO, J. S. AND H. WHITE (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

——— (2010): "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157, 458–480.

CLAUS, E. AND V. H. NGUYEN (2019): "Monetary Policy Shocks from the Consumer Perspective," *Journal of Monetary Economics*, in press.

DAVIES, R. B. (1977): "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 64, 247–254.

——— (1987): "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 74, 33–43.

DHRYMES, P. J. (1971): *Distributed Lags: Problems of Estimation and Formulation*, Edinburgh: Oliver and Boyd.

DIEBOLD, F. X. AND K. YILMAZ (2014): "On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms," *Journal of Econometrics*, 182, 119–134.

ELHORST, J. P. (2014): *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Berlin: Springer.

ENGLE, R. F. AND C. W. GRANGER (1987): "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

FEDOSEEVA, S. (2013): "(A)symmetry, (Non)linearity and Hysteresis of Pricing-To-Market: Evidence from German Sugar Confectionery Exports," *Journal of Agricultural and Food Industrial Organization*, 11, 69–85.

GREENWOOD-NIMMO, M. J., V. H. NGUYEN, AND Y. SHIN (in press): "Measuring the Connectedness of the Global Economy," *International Journal of Forecasting*.

GRILICHES, Z. (1967): "Distributed Lags: A Survey," *Econometrica*, 35, 16–49.

HAMMOUDEH, S., A. LAHIANI, D. K. NGUYEN, AND R. M. SOUSA (2015): "An Empirical Analysis of Energy Cost Pass-Through to CO2 Emission Prices," *Energy Economics*, 49, 149–156.

HANNAN, E. J. AND B. G. QUINN (1979): "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 41, 190–195.

HANSEN, B. E. (1996): "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

HAUSMAN, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

HE, Z. AND F. ZHOU (2018): "Time-Varying and Asymmetric Eeffects of the Oil-Specific Demand Shock on Investor Sentiment," *PLOS ONE*, 31.

HENDRY, D. F. AND G. E. MIZON (1978): "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England," *Economic Journal*, 88, 549–563.

HENDRY, D. F., A. R. PAGAN, AND J. D. SARGAN (1984): "Dynamic Specifications," in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, Amsterdam: North Holland, vol. 2, chap. 18, 1023–1100.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–255.

KOENKER, R. AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

KOYCK, L. M. (1954): *Distributed Lags and Investment Analysis*, Amsterdam: North-Holland.

KUERSTEINER, G. M. AND I. R. PRUCHA (2020): "Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity," *Econometrica*, 88, 2109–2146.

LEE, L. AND J. YU (2010): "Some Recent Developments in Spatial Panel Data Models," *Regional Science and Urban Economics*, 40.

MADDALA, G. S. (1977): *Econometrics*, New York: McGraw-Hill.

NERLOVE, M. L. (1972): "Lags in Economic Behavior," *Econometrica*, 40, 221–251.

PAL, D. AND S. MITRA (2015): "Asymmetric Impact of Crude Price on Oil Product Pricing in the United States: An Application of Multiple Threshold Nonlinear Autoregressive Distributed Lag Model," *Economics Modelling*, 51, 436–443.

PESARAN, M. H. AND Y. SHIN (1998): "An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis," in *Econometrics and Economic Theory: The Ragnar Frisch Centennial Symposium*, ed. by S. Strom, Cambridge: Cambridge University Press, Econometric Society Monographs, 371–413.

PESARAN, M. H., Y. SHIN, AND R. J. SMITH (2001): "Bounds Testing Approaches to the Analysis of Level Relationships," *Journal of Applied Econometrics*, 16, 289–326.

PESARAN, M. H., Y. SHIN, AND R. P. SMITH (1999): "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels," *Journal of the American Statistical Association*, 94, 621–634.

PESARAN, M. H. AND R. P. SMITH (1995): "Estimating Long-run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics*, 68, 79–113.

PHILLIPS, P. C. AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variable Regression with I(1) Processes," *Review of Economic Studies*, 57, 99–125.

PHILLIPS, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

PITARAKIS, J.-Y. (2006): "Model Selection Uncertainty and Detection of Threshold Effects," *Studies in Nonlinear Dynamics and Econometrics*, 10, 1–30.

SARGAN, J. D. (1964): "Wages and Prices in the United Kingdom: A Study in Econometric Methodology (with Discussion)," in *Econometric Analysis for National Economic Planning*, ed. by P. Hart, G. Mills, and J. Whitaker, London: Butterworth and Co., 25–63.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

SEONG, D., J. S. CHO, AND T. TERÄSVIRTA (2019): "Comprehensive Testing of Linearity against the Smooth Transition Autoregressive Model," CREATES Research Papers 2019-17, Aarhus University, Aarhus.

SHI, W. AND L. LEE (2017): "Spatial Dynamic Panel Data Models with Interactive Fixed Effects," *Journal of Econometrics*, 197, 323–347.

SHIN, Y. AND M. THORNTON (2019): "The Spatio-Temporal Autoregressive Distributed Lag Modelling Approach to an Analysis of the Spatial Heterogeneity and Diffusion Dependence," Mimeo: University of York.

SHIN, Y., B. YU, AND M. J. GREENWOOD-NIMMO (2014): "Modelling Asymmetric Cointegration and Dynamic Multipliers in a Nonlinear ARDL Framework," in *Festschrift in Honor of Peter Schmidt:*

Econometric Methods and Applications, ed. by W. Horrace and R. Sickles, New York (NY): Springer Science & Business Media, 281–314.

SIMS, C. A. (1974): “Distributed Lags,” in *Frontiers of Quantitative Economics*, ed. by D. A. Kendrick and M. D. Intriligator, Amsterdam: North Holland, vol. 2.

STOCK, J. H. AND M. W. WATSON (1993): “A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems,” *Econometrica*, 61, 783–820.

SÜSSMUTH, B. AND U. WOITEK (2013): “Estimating Dynamic Asymmetries in Demand at the Munich Oktoberfest,” *Tourism Economics*, 19, 653–674.

THOMAS, J. (1977): “Some Problems in the use of Almon’s Technique in the Estimation of Distributed Lags,” *Empirical Economics*, 2, 175–193.

WALD, A. (1943): “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large,” *Transactions of the American Mathematical Society*, 54, 426–482.

WALLIS, K. F. (1969): “Some Recent Developments in Applied Econometrics: Dynamic Models and Simultaneous Equation Systems,” *Journal of Economic Literature*, 7, 771–796.

WICKENS, M. R. AND T. S. BREUSCH (1988): “Dynamic Specification, the Long Run Estimation of the Transformed Regression Models,” *Economic Journal*, 98, 189–205.

XIAO, Z. (2009): “Quantile Cointegrating Regression,” *Journal of Econometrics*, 150, 248–260.

ZELLNER, A. (1979): “Statistical Analysis of Econometric Models,” *Journal of the American Statistical Association*, 74, 628–643.