Four Australian Banks and the Multivariate Time-Varying Smooth Transition Correlation GARCH model

Anthony D. Hall, Annastiina Silvennoinen and Timo Teräsvirta

CREATES Research Paper 2021-13

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 9352 1936

Four Australian Banks and the Multivariate Time-Varying Smooth Transition Correlation GARCH model

Anthony D. Hall, Annastiina Silvennoinen[*] and Timo Teräsvirta[†‡]
[*]NCER, Queensland University of Technology, Brisbane
[†]CREATES, Aarhus University
[‡]C.A.S.E., Humboldt-Universität zu Berlin

September 27, 2021

Abstract

This paper looks at changes in the correlations of daily returns between the four major banks in Australia. Revelations from the analysis are of importance to investors, but also to government involvement, due to the large proportion of the highly concentrated financial sector relying on the stability of the Big Four. For this purpose, a methodology for building Multivariate Time-Varying STCC–GARCH models is developed. The novel contributions in this area are the specification tests related to the correlation component, the extension of the general model to allow for additional correlation regimes, and a detailed exposition of the systematic, improved modelling cycle required for such nonlinear models. There is an R-package that includes the steps in the modelling cycle. Simulations evidence the robustness of the recommended model building approach. The empirical analysis reveals an increase in correlations of the Australia's four largest banks that coincides with the stagnation of the home loan market, technology changes, the mining boom, and Basel II alignment, increasing the exposure of the Australian financial sector to shocks.

Acknowledgements. This research has been supported by Center for Research in Econometric Analysis of Time Series (CREATES). Part of the work for this paper was carried out when the third author was visiting School of Economics and Finance of Queensland University of Technology, Brisbane, whose kind hospitality is gratefully acknowledged. We would also like to thank Katherine Uylangco and Natalie Elms for insightful comments, and Glen Wade for his work with the R-package. Material from this paper has been presented at the 26th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, Tokyo, March 2018, the workshop 'Frontiers in Econometrics', Queensland University of Technology, Brisbane, July 2018, the

Quantitative Methods in Finance 2018 Conference (UTS), Sydney, December 2018, and the 13th International Conference of the ERCIM WG on Computational and Methodological Statistics, London, December 2020. Comments from participants are gratefully acknowledged. Responsibility for any errors and shortcomings in this work remains ours.

JEL Classification Codes: C32; C52; C58

Keywords: Unconditional correlation; modelling volatility; modelling correlations; multivariate autoregressive conditional heteroskedasticity.

1 Introduction

The Australian banking market is an oligopoly dominated by four banks. Listed in the descending order of market share, they are the Commonwealth Bank of Australia (CBA), Westpac Banking Corporation (WBC), National Australia Bank (NAB) and Australia and New Zealand Banking Group (ANZ), commonly called the Big Four. They currently represent 19% of the market value of the ASX200 share index (see Figure 1) and hold about 80% of the home loan market in Australia. Consequently, the banking sector is a major component for many Australian superannuation and other investment funds. This is due not only to its size but also to its relative secure status. The latter originated from the so-called Four Pillars Policy established by the government in 1990 and the extensive government support over the past three decades.

Correlations between the banks are important because they make up a large percentage of the ASX200 market cap. But perhaps even more importantly, the Big Four are systemically important in Australia's financial system. Over the past decade they represent over 60% of the ASX200 Financials Index. They are also highly interconnected which means that financial contagion is likely to occur between the four banks and hence the failure of any one of these banks would have a significant impact on the other three institutions as well as on individuals and company that borrow from (i.e. have loans with), lend to (i.e. have deposits with) these banks, as well as shareholders of the banks. This is essentially the 'too big to fail' argument that was bandied around during the GFC. Hence, having a better understanding the correlation of these banks over time is crucial for Australia's financial system.

From the investors' point of view the Big Four is therefore an important object of study. For instance, many large superannuation funds have the banking sector as a major component in their portfolios. Because of this and the role of the Big Four in the Australian economy in general, analysing their stock return volatility and correlations between returns is of considerable interest. Our interest lies in investigating long run movements in both of them in the Big Four framework.

As to the Big Four daily returns, their volatilities cannot automatically be assumed stationary. Furthermore the correlations, even when time-varying, cannot *a priori* be assumed to fluctuate around a constant level, which is one of the assumptions in many popular multivariate GARCH models. For these reasons, a model

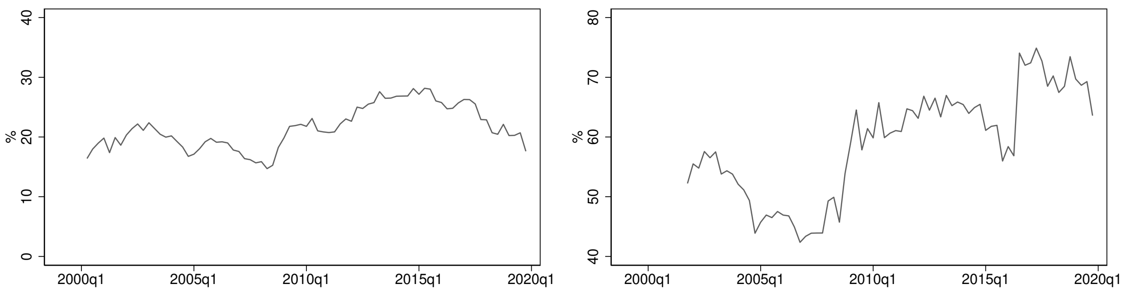

Figure 1: The market capitalisation of the Big Four as percentage of ASX200 (left) and of ASX200 Financials Index (right).

with deterministically time-varying component that describes the smoothly changing dispersion in univariate GARCH models is required. In addition, this feature has to be combined with a correlation matrix in which the time-varying correlations are not restricted to fluctuating around a constant level. The modelling framework we adopt here is based on earlier work by Silvennoinen and Teräsvirta (2021). Their Multiplicative Time-Varying Smooth Transition Correlation GARCH model, MTV for short, places the emphasis on smooth long-run movements in correlations which are the main object of study in this application. Statistical inference within the MTV model framework is based on the asymptotic results in Silvennoinen and Teräsvirta (2021). The focus in that paper is on maximum likelihood estimation of the parameters of the MTV model. Here we add to that work by concentrating on specification and evaluation of the model. Before actually estimating an MTV model, the model builder has to make a number of data-driven decisions while specifying the structure of the model. After estimation, the structure has to be tested to reveal its potential weaknesses.

To accompany the aforementioned modelling cycle, we have built an R-package that can assist in applying the MTV model in practice. The package includes the estimation routines, as well as the specification and evaluation stage tests. It should be noted that the modelling process is data driven. It requires user input and consists of several steps. An automated general approach is not feasible for the highly nonlinear MTV family of models if reliable estimates are desired.

The plan of the paper is as follows. The broad strokes of the Australian landscape for superannuation, housing markets, and the Big Four are, up to their relevance to the topic of the paper, provided in Section 2. The MTV model is introduced in Section 3, followed by details of the stages and procedures related to the model building in Section 4. Model specification is considered in Section 5, estimation in Section 6 and evaluation in Section 7. Section 8 is devoted to the estimation results in relation to the correlations between the Big Four and the implications of the findings on the investor risk profile. Conclusions can be found in Section 9. There are also appendices containing material such as details of the test statistics, simulation results, details of the estimation algorithm and estimated equations.

2 Background information and data

In 1990 the Australian government adopted an intervention policy called 'six pillars'. It covered the four biggest Australian banks and two insurers (AMP and National Mutual) and stated that further mergers of these institutions would not be accepted. The basic idea was to ensure a competitive banking market. In 1997 the policy became 'four pillars' as the insurers were left outside the arrangement. The policy has mostly enjoyed the bipartisan support since its establishment, and the proponents of the policy have argued it has contributed to the stability and strength of the Australian financial sector. The government also had sympathetic policy settings, which allowed the banks to recapitalize in the 1990s and 2000s. During this period there were financial losses at WBC ($1.6bn loss in 1992 and close to insolvency), ANZ (poorly executed international expansions) and subsequently with NAB (purchase of the US mortgage originator and servicer Homeside led to a $2.2bn in losses 2002). Larger financial concentration due to mergers with other financial institutions is seen as acceptable: in 2008 WestPac and CBA acquired St. George and BankWest, then the fifth and sixth largest banks, respectively. The impact can be seen in Figure 1 as the Big Four's share of the ASX200 Financials Index increased drastically.

The recent history of the pillars policy coincides with a few major incidents and changes. These include the dot-com boom in the late 1990s and early 2000s, and the global financial crisis nearly ten years later, but also events that have had more localised impacts, such as a number of regulatory changes (Basel guidelines), the most recent mining boom that started around 2005 and was interrupted by the GFC, and technology-driven market disruptions (non bank lenders and payment providers). Since the global financial crisis, the banks have enjoyed substantial government support including a deposit guarantee and, as already noticed, have come to dominate the home mortgage market with an 80% market share.

Growth of the housing credit market ceased around 2003 and the market remained stagnant until the early 2010s. The volume of investment loans in particular stagnated, while owner occupier loans saw cyclical patterns, albeit around a very slow positive trend. At the same time the Australian house prices were increasing. Households owning their accommodation were pushed into extremely high level of debts or to finding alternative housing options because inflation did not match the rates. The drying up of the housing credit markets then had an impact on the four banks.

One important aspect that has changed the banking sector has to do with the technological change. By 2004, the 'old-school' bank manager lending was replaced by lending managers and online forms. The concept of customer loyalty to brand was replaced by instant access to lending providers' products, and the long term relations with the loan managers were steadily replaced by online interface. The access to transparency and comparability at an increasing rate pushed the banks to a very different competitive environment. The ease of comparison as well as removal of barriers to changing mortgage providers increased the efficiency in determining going rates and conditions.[1] The competition in the homeloan markets, of which the Big Four have an 80% share, has made investors regard these banks as a rather homogeneous group with increasingly identical in their products, rather than individual entities with sharply different valuations.

In 2003, it was announced that Basel II was to be implemented in Australia. The updated accord was aiming to level inequalities amongst the internationally active banks and setting expectations regarding capital adequacy requirements. The Australian Prudential Regulation Authority (APRA) in charge of overseeing the uptake of the accord worked extensively with numerous ADIs, industry and other relevant bodies during 2005-2007, aiming to ensure adoption of Basel II included all relevant aspects of the implementation process, goals, and impacts. Fear for being subjected to a competitive disadvantage relative to their international counterparts, both within international and domestic operations, coupled with an opportunity for a reduced regulatory capital incentivised the banks to signal early their preference to conform with the accord.[2] As a result, Australia was amongst the first nations to have fully implemented the framework, on 1 January 2008.

The daily return series for the Big Four used in this paper extend from 2 January 1992 to 31 January 2020. The series are plotted in Figure 2. From the plots it is seen that there is a rather tranquil period between 2003 and 2008, except perhaps for WBC, that overlaps with the stagnant housing credit market, and also includes the most recent mining boom. Volatility then substantially increases during the global financial crisis beginning 2008. There is another increase in volatility after 2016. Consequently, the amplitude of volatility clusters varies over time for all four banks. This shows in the autocorrelation functions of squared returns in Figure 3. In all four cases, autocorrelations decay very slowly as a function of the lag length.

These amplitude changes suggest nonstationarity. In its presence, the standard weakly stationary GARCH or, in our case GJR-GARCH, model does not provide an adequate description of the data. Our alternative is the MTV model to which we now turn.

3 The MTV model

The MTV model used in this paper belongs to the family of multivariate GARCH models introduced by Bollerslev (1990). In the original model the conditional correlations were constant, hence the name Constant Conditional Correlation (CCC-GARCH) model. This assumption that made the resulting model rather parsimonious was later found too restrictive in applications, and time-varying correlations were simultaneously introduced by Engle (2002) (dynamic conditional correlations, DCC) and Tse and Tsui (2002) (varying correlations, VC). In these models, conditional variance components are typically assumed stationary, and correlations are

[1] See Submission to the Inquiry into Competition in the Banking and Non-Banking Sectors, Reserve Bank of Australia, 10 July 2008. https://www.rba.gov.au/publications/submissions/financial-sector/inquiry-report-2009-05/

[2] See Explanatory Statement, Banking (prudential standard) determination Nos.5,12,15 of 2007. https://www.legislation.gov.au/Details/F2007L04593/

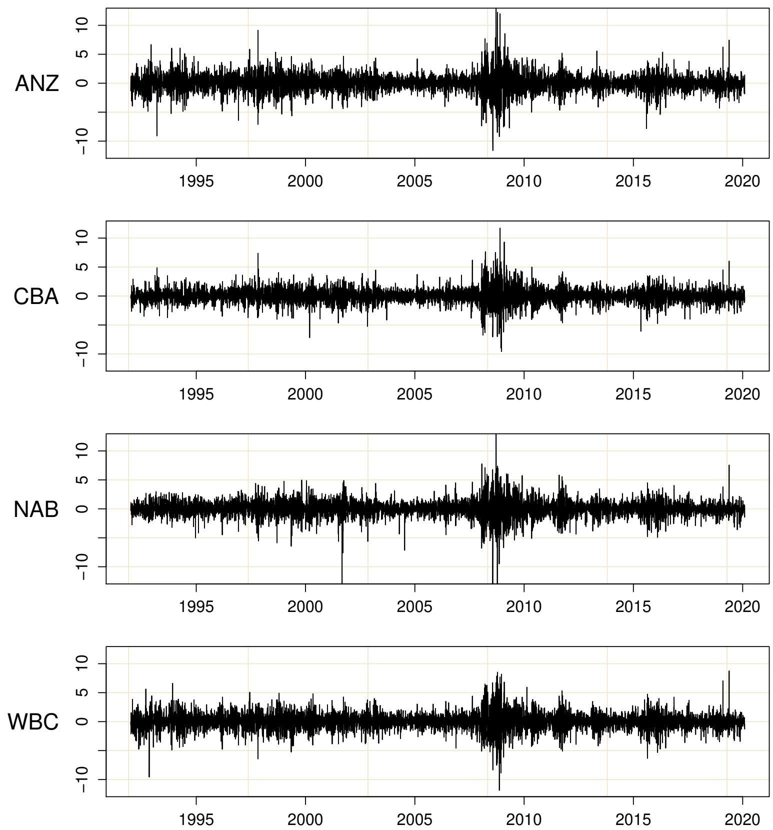

Figure 2: Daily returns of the Big Four, from 2 January 1992 to 31 January 2020. From top to bottom: ANZ, CBA, NAB and WBC

assumed, at least implicitly, to fluctuate around a constant level. In order to consider the MTV model as defined in Silvennoinen and Teräsvirta (2021) we introduce some notation. The observable stochastic $N \times 1$ vector $\boldsymbol{\varepsilon}_t$ is decomposed in a customary fashion as

$$\boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2} \boldsymbol{z}_t = \boldsymbol{S}_t \boldsymbol{D}_t \boldsymbol{P}_t^{1/2} \boldsymbol{\zeta}_t \tag{1}$$

where $\boldsymbol{H}_t = \boldsymbol{S}_t \boldsymbol{D}_t \boldsymbol{P}_t \boldsymbol{D}_t \boldsymbol{S}_t$ is an $N \times N$ covariance matrix, and $\boldsymbol{\zeta}_t \sim \text{iid}(\boldsymbol{0}, \boldsymbol{I}_N)$. We also define $\boldsymbol{z}_t = \boldsymbol{P}_t^{1/2} \boldsymbol{\zeta}_t$, a vector of independent random variables with $\mathsf{E}\boldsymbol{z}_t = \boldsymbol{0}$ and a positive definite deterministically varying covariance matrix $\text{cov}(\boldsymbol{z}_t) = \boldsymbol{P}_t$. The structure of $\boldsymbol{P}_t$ will be defined later. The deterministic matrix $\boldsymbol{S}_t = \text{diag}(g_{1t}^{1/2}, \ldots, g_{Nt}^{1/2})$ has positive diagonal elements for all t, and $\boldsymbol{D}_t = \text{diag}(h_{1t}^{1/2}, \ldots, h_{Nt}^{1/2})$ contains the conditional standard deviations of the elements of $\boldsymbol{S}_t^{-1} \boldsymbol{\varepsilon}_t = (\varepsilon_{1t}/g_{1t}^{1/2}, \ldots, \varepsilon_{Nt}/g_{Nt}^{1/2})'$. As in Silvennoinen and Teräsvirta (2021) and earlier univariate papers, beginning with Amado and Teräsvirta (2008), and in the multivariate time-varying GARCH article by Amado and Teräsvirta (2014), the diagonal elements of $\boldsymbol{S}_t^2$ are defined as follows:

$$g_{it} = g_i(t/T) = \delta_{i0} + \sum_{j=1}^{r_i} \delta_{ij} G_{ij}(t/T, \gamma_{ij}, \boldsymbol{c}_{ij}) \tag{2}$$

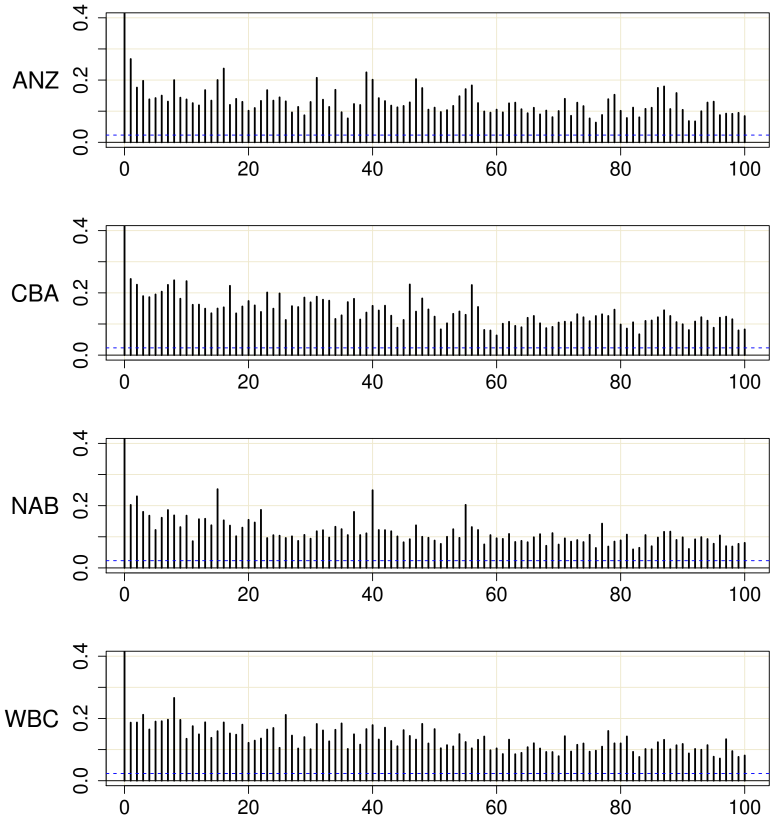

Figure 3: The first 100 autocorrelations of squared returns. From the top: ANZ, CBA, NAB and WBC

$i = 1, \ldots, N$, where $\delta_{i0} > 0$ is a known constant, $\delta_{ij} \neq 0$, $j = 1, \ldots, r_i$, and the (generalised) logistic function

$$G_{ij}(t/T, \gamma_{ij}, \boldsymbol{c}_{ij}) = (1 + \exp\{-\gamma_{ij} \prod_{k=1}^{K_{ij}} (t/T - c_{ijk})\})^{-1} \tag{3}$$

where $\gamma_{ij} > 0$ and $\boldsymbol{c}_{ij} = (c_{ij1}, \ldots, c_{ijK_{ij}})'$ such that $c_{ij1} \leq \ldots \leq c_{ijK_{ij}}$. Both $\gamma_{ij} > 0$, $c_{ij1} \leq \ldots \leq c_{ijK_{ij}}$, and $\delta_{ij} \neq 0$, $j = 1, \ldots, r_i$ are identification restrictions. Assuming δ_{i0} in (2) known is another one. Furthermore, to prevent exchangeability of components in (2), restrictions are needed on $\boldsymbol{c}_{ij}$. As an example, if $K_{ij} = 1$ for $j = 1, \ldots, r_i$, one can assume (for instance) that $c_{ij1} < \ldots < c_{ir_i1}$.

The conditional variances have a GARCH or GJR-GARCH(1,1) structure, see Glosten, Jagannathan and Runkle (1993) for the latter:

$$h_{it} = \alpha_{i0} + \alpha_{i1}\varepsilon_{i,t-1}^2 + \kappa_{i1} I(\varepsilon_{t-1} < 0)\varepsilon_{i,t-1}^2 + \beta_{i1} h_{i,t-1} \tag{4}$$

where $I(A)$ is an indicator function: $I(A) = 1$ when A occurs, zero otherwise. A higher-order structure is possible, although there do not seem to exist applications of the GJR-GARCH model of order greater than one.

As discussed in earlier papers, the idea of g_{it} is to normalise or rescale the observations. Left-multiplying (1) by $\boldsymbol{S}_t^{-1}$ yields

$$\boldsymbol{\phi}_t = \boldsymbol{S}_t^{-1}\boldsymbol{\varepsilon}_t = \boldsymbol{D}_t\boldsymbol{z}_t$$

where each element of $\boldsymbol{\phi}_t$ is assumed to have a standard weakly stationary GARCH representation while the conditional covariance matrix $\mathsf{E}\{\boldsymbol{\phi}_t\boldsymbol{\phi}_t'|\mathcal{F}_{t-1}\} = \boldsymbol{D}_t\boldsymbol{P}_t\boldsymbol{D}_t$. In order to describe the correlation structure, we employ the Double Smooth Transition Conditional Correlation (DSTCC) model by Silvennoinen and Teräsvirta (2009). In that model, assuming that the transition variable is t/T throughout, the time-varying correlation matrix $\boldsymbol{P}_t$ is defined as

$$\begin{aligned}\boldsymbol{P}_t =&(1 - G_2(t/T,\gamma_2,\boldsymbol{c}_2))\{(1 - G_1(t/T,\gamma_1,\boldsymbol{c}_1))\boldsymbol{P}_{(11)} + G_1(t/T,\gamma_1,\boldsymbol{c}_1)\boldsymbol{P}_{(21)}\}\\ &+ G_2(t/T,\gamma_2,\boldsymbol{c}_2)\{(1 - G_1(t/T,\gamma_1,\boldsymbol{c}_1))\boldsymbol{P}_{(12)} + G_1(t/T,\gamma_1,\boldsymbol{c}_1)\boldsymbol{P}_{(22)}\} \qquad (5)\end{aligned}$$

where $\boldsymbol{P}_{(ij)}$, $i,j = 1,2$, are four positive definite correlation matrices not equal to each other, and

$$G_i(t/T,\gamma_i,\boldsymbol{c}_i) = (1 + \exp\{-\gamma_i \prod_{k=1}^{K_i}(t/T - c_{ik})\})^{-1},\ \gamma_i > 0 \qquad (6)$$

where $\boldsymbol{c}_i = (c_{i1},...,c_{iK_i})$, $c_{i1} < ... < c_{iK_i}$, $i = 1,2$. For the Big Four application we simplify the definition (5) slightly by assuming $\boldsymbol{P}_{(12)} = \boldsymbol{P}_{(22)}$, so (5) becomes

$$\begin{aligned}\boldsymbol{P}_t =&(1 - G_2(t/T,\gamma_2,\boldsymbol{c}_2))\{(1 - G_1(t/T,\gamma_1,\boldsymbol{c}_1))\boldsymbol{P}_{(1)} + G_1(t/T,\gamma_1,\boldsymbol{c}_1)\boldsymbol{P}_{(2)}\}\\ &+ G_2(t/T,\gamma_2,\boldsymbol{c}_2)\boldsymbol{P}_{(3)} \qquad (7)\end{aligned}$$

where re-indexing the matrices highlights the interpretation that there are two transitions over time. One is from $\boldsymbol{P}_{(1)}$ to $\boldsymbol{P}_{(2)}$ and the other one from a convex combination of these two to $\boldsymbol{P}_{(3)}$. Since $\boldsymbol{P}_{(1)}$, $\boldsymbol{P}_{(2)}$, and $\boldsymbol{P}_{(3)}$ are positive definite, $\boldsymbol{P}_t$ is positive definite as a convex combination of the three matrices. This simplified version of the DSTCC model is especially useful when modelling correlations that shift from one state to the next as a function of time. To that end, the obvious extension to n such transitions is best expressed as a recursion

$$\begin{aligned}&\boldsymbol{P}_t^{(0)} = \boldsymbol{P}_{(1)}\\ &\boldsymbol{P}_t^{(n)} = (1 - G_n(t/T,\gamma_n,\boldsymbol{c}_n))\boldsymbol{P}_t^{(n-1)} + G_n(t/T,\gamma_n,\boldsymbol{c}_n)\boldsymbol{P}_{(n+1)} \qquad (8)\end{aligned}$$

When $G_2(t/T,\gamma_2,\boldsymbol{c}_2) \equiv 1$ and $N = 2$, (5) and (7) collapse into the smooth transition correlation GARCH model by Berben and Jansen (2005) or, if the transition variable in G_1 is stochastic and $N \geq 2$, into the smooth transition conditional correlation GARCH model of Silvennoinen and Teräsvirta (2005, 2015). An MTV-Conditional Correlation GARCH model with GARCH equations similar to the ones here but differently defined stochastic $\boldsymbol{P}_t$ was discussed in Amado and Teräsvirta (2014). It may be noted that Feng (2006) introduced another multivariate Conditional Correlation type GARCH model with deterministically varying correlations. In this model the variation is described nonparametrically, and the model can be viewed as a generalisation of the univariate model in Feng (2004).

4 The three stages of model building

The MTV model is rather general and nests many models. To take just one example, fitting an MTV model when a nested CCC-GARCH model actually generates the data leads to inconsistent parameter estimates. For this reason, building adequate MTV models requires care, and a systematic approach is necessary. Selecting a candidate from this family of models is a data-driven process, and statistical inference has to be used to obtain an acceptable model such that it passes the available misspecification tests.

In this work we follow the classical approach to model building advocated by Box and Jenkins (1970) and later applied to nonlinear models of the conditional mean, see for example Teräsvirta, Tjøstheim and Granger (2010, Ch. 16). It has also been applied to building single-equation MTV-GARCH models; see Amado and Teräsvirta (2017) and Amado, Silvennoinen and Teräsvirta (2017). The idea is to first specify the model (select a member from the family of MTV models), and once this has been done estimate its parameters. At the evaluation stage the estimated model is subjected to a battery of misspecification tests. These three stages, specification, estimation and evaluation, will be considered in the next sections. The emphasis will be on specification and evaluation as maximum likelihood estimation of the parameters of the MTV model has already been considered in Silvennoinen and Teräsvirta (2021).

5 Specification of the MTV model

5.1 Specification of the univariate variance equations

Specification of the MTV model is begun by specifying the univariate volatility equations. This was first discussed in Amado and Teräsvirta (2017). The idea is to begin with a GARCH(1,1) model by Bollerslev (1986) or the GJR-GARCH model by Glosten et al. (1993) and test the hypothesis that the multiplicative deterministic component is constant. The single-equation MTV-GARCH model has the following form:

$$\varepsilon_{it} = z_{it} h_{it}^{1/2} g_{it}^{1/2} \tag{9}$$

where $z_{it} \sim \mathrm{iid}(0,1)$. The conditional variance equals

$$h_{it} = \alpha_0 + \alpha_1 \phi_{i,t-1}^2 + \kappa_{i1} I(\phi_{i,t-1} < 0)\phi_{i,t-1}^2 + \beta_{i1} h_{i,t-1} \tag{10}$$

where $\phi_{it} = \varepsilon_{it}/g_{it}^{1/2}$. The deterministic positive-valued function $g_{it} = g_i(t/T) = g_i(t/T; \boldsymbol{\theta}_1)$ is defined as in (2) and (3).

Positivity of (2) imposes the following restrictions on δ_{ij}, $j = 1, \ldots, r_i$:

$$\delta_{i0} + \sum_{j=1}^{r_i} \delta_{ij} G_{ij}(r, \gamma_{ij}, \boldsymbol{c}_{ij}) > 0$$

for all $r \in$ [0,1].

Typically in applications, $K_{ij} = 1, 2$. There are two specification issues, determining r_i and choosing K_{ij}, $j = 1, \ldots, r_i$. It is possible that $g(t/T; \gamma, \boldsymbol{c}) = \delta_0 > 0$, that is, $g(t/T; \gamma, \boldsymbol{c})$ is a positive constant. In this case the MTV-GARCH model collapses into a standard GARCH or GJR-GARCH equation.

Amado and Teräsvirta (2017) solved the problem of choosing r_i by first estimating the GARCH model and testing the hypothesis of a constant $g_i(t/T)$ against the alternative $r_i = 1$ in (2) thereafter using a Lagrange multiplier type test. The test can be viewed as a misspecification test of the estimated GARCH model. If the null hypothesis is rejected, an MTV-GARCH model with a single transition is estimated, and the hypothesis $r = 1$ is tested against $r_i = 2$. Sequential testing continues until the first non-rejection of the null hypothesis. The number of transitions is determined in this order because of an identification problem: the model with $r_i + 1$ transitions is not identified if the true number of transitions is r_i. The shape of the logistic function, controlled by the parameter K_{ij}, can be determined using the sequence of tests familiar from the specification of smooth transition autoregressive (STAR) models, see Teräsvirta (1994) or Teräsvirta et al. (2010, Chapter 16). Details can be found in Amado and Teräsvirta (2017).

More recently, Silvennoinen and Teräsvirta (2016) considered testing the constancy of $g_i(t/T)$ before estimating the GARCH model, that is, assuming $h_{it} = 1$ in (9). This implies that the size of the test is distorted because conditional heteroskedasticity is ignored, so the size of the test has to be adjusted by simulation. It turned out that power of the size-adjusted test improved considerably compared to the case where the test is a misspecification test. Reasons for this are discussed in Silvennoinen and Teräsvirta (2016).

A major difficulty with this approach is that while in simulations the parameters of the conditional variance component h_{it} under the null hypothesis are known, in practice this is not the case. The underlying 'null' GARCH process has to be generated artificially. In so doing, special attention is to be placed on the persistence of the (GJR-)GARCH process, measured by $\alpha_{i1} + \kappa_{i1}/2 + \beta_{i1}$ in (10) when $g_{it} \equiv 1$. In fact, the asymmetry parameter has no practical importance for the purpose of calibrating the test statistic distribution, and it is therefore sufficient to restrict attention to the standard GARCH process. Other features, such as implied kurtosis or relative sizes of α and β corresponding to a particular level of persistence only have a negligible effect on the performance of the test.

A practical problem is that it is not possible to estimate this measure of persistence when the null hypothesis does not hold, that is, when g_{it} is not constant over time. How this difficulty is handled has an effect on the power of the test. We study two approaches that are discussed more in detail in Appendix B.1. The first one consists of visually identifying a period of time where there appears to be no change in the overall level of baseline volatility. A standard GARCH(1,1) is estimated over this subperiod. The second approach is to use rolling window variance targeting. This means that the intercept in the GARCH equation is time-varying, and its value at each point in time is calculated such that it matches the unconditional variance obtained from a window around that point in time. Simulations discussed in Appendix B.1 experiment with the choice of window size. Both of these methods provide GARCH parameter and persistence estimates that are used for calibrating the null distribution of the test statistic and calculating p-values.

5.2 Specification of time-varying correlations

After the MTV-GARCH equations have been specified and estimated assuming the errors are uncorrelated, the next step is to specify the time-varying correlations. This is done by sequential testing, First, constancy of correlation tested against the model with a single transition, i.e., $G_2(t/T, \gamma_2, c_2) \equiv 1$ in (5). The null hypothesis is that the model is a MTV-Constant Correlation GARCH model as in Bollerslev (1990), except that the GARCH equations are MTV-GARCH equations. If it is rejected, the one-transition model estimated and tested against (5) or (7). If, again, rejected, the alternative with two transitions is estimated. This is repeated until no further evidence for time-variation in the correlations is detected.

As discussed in Silvennoinen and Teräsvirta (2005, 2015), the MTV model with one transition is only identified under the alternative, which invalidates the standard asymptotic inference. The identification problem can be circumvented by approximating the transition function (6) by its Taylor expansion around the null hypothesis, H_0: $\gamma_1 = 0$. The form of the expansion depends on the order of the exponent in (6).

The test can be constructed along the lines presented in the appendix of Silvennoinen and Teräsvirta (2005).[3] See also Silvennoinen and Teräsvirta (2021). To derive the test statistic, consider the first-order Taylor expansion of (6) around $\gamma_1 = 0$ assuming $K_i = 2$. It has the following form:

$$\begin{aligned} G_i(t/T, \gamma_i, c_i) &= (1 + \exp\{-\gamma_i \prod_{k=1}^{K_i} (t/T - c_{ik})\})^{-1} \\ &= \frac{1}{2} + \frac{1}{4}(t/T - c_{i1})(t/T - c_{i2})\gamma_i + R_2(t/T; \gamma_i) \end{aligned} \tag{11}$$

where $R_2(t/T; \gamma_i)$ is the remainder. Using (11), (5) becomes

$$\begin{aligned} \boldsymbol{P}_t =& (\boldsymbol{P}_{(1)} - \boldsymbol{P}_{(2)})(\frac{1}{2} + \frac{\gamma_i c_{i1} c_{i2}}{4}) + \boldsymbol{P}_{(2)} - (t/T)(\boldsymbol{P}_{(1)} - \boldsymbol{P}_{(2)})\frac{\gamma_i(c_{i1} + c_{i2})}{4} \\ &+ (t/T)^2(\boldsymbol{P}_{(1)} - \boldsymbol{P}_{(2)})\frac{\gamma_i}{4} + (\boldsymbol{P}_{(1)} - \boldsymbol{P}_{(2)})R_2(t/T; \gamma_i) \\ =& \boldsymbol{P}_{(A0)} + (t/T)\boldsymbol{P}_{(A1)} + (t/T)^2\boldsymbol{P}_{(A2)} + (\boldsymbol{P}_{(1)} - \boldsymbol{P}_{(2)})R_2(t/T; \gamma_i) \end{aligned}$$

where $\boldsymbol{P}_{(1)} \neq \boldsymbol{P}_{(2)}$. The main diagonals of $\boldsymbol{P}_{(A1)}$ and $\boldsymbol{P}_{(A2)}$ consist of zeroes. Setting $\boldsymbol{\rho}_A = (\boldsymbol{\rho}'_{A0}, \boldsymbol{\rho}'_{A1}, \boldsymbol{\rho}'_{A2})'$, where $\boldsymbol{\rho}_{Ai} = \mathrm{vecl}(\boldsymbol{P}_{(Ai)})$, $i = 0, 1, 2$, the new null hypothesis is H_0: $\boldsymbol{\rho}_{A1} = \boldsymbol{\rho}_{A2} = \mathbf{0}_{N(N-1)/2}$. Note that a simpler version of the test assumes $K_i = 1$ and yields a similar approximation although without the term $(t/T)^2\boldsymbol{P}_{(A2)}$. The new null in this case is H_0: $\boldsymbol{\rho}_{A1} = \mathbf{0}_{N(N-1)/2}$. This version of the test is more powerful than the former in case time-variation in the correlations is monotonic.

[3] Available also in http://econ.au.dk/research/research-centres/creates/research/research-papers/ supplementary-downloads.

However, and especially with longer time horizons, this may not always be the case, and the square term of the expansion is able to capture at least some nonmonotonic changes.

The details of the ensuing LM-type test statistics are presented in the Appendix A.3 and A.4.

6 Estimation of the MTV model

After specifying the deterministic components of the model, both in GARCH equations and correlations, one can estimate the complete model with conditional heteroskedasticity included. The log-likelihood of the MTV-STCC-GARCH model has the form

$$\begin{aligned}\ln f(\boldsymbol{\zeta}_t|\boldsymbol{\theta}) \propto &-(1/2)\sum_{i=1}^{N}\ln g_i(t/T)-(1/2)\sum_{i=1}^{N}\ln h_{it}-(1/2)\ln|\boldsymbol{P}_t| \\ &-(1/2)\boldsymbol{\varepsilon}_t'\{\boldsymbol{S}_t\boldsymbol{D}_t\boldsymbol{P}_t\boldsymbol{D}_t\boldsymbol{S}_t\}^{-1}\boldsymbol{\varepsilon}_t. \qquad (12)\end{aligned}$$

Since $\boldsymbol{D}_t\boldsymbol{z}_t = \boldsymbol{S}_t^{-1}\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}/g_{1t}^{1/2},\ldots,\varepsilon_{Nt}/g_{Nt}^{1/2})'$, it is seen from (4) that the conditional variance components in (12) are

$$h_{it} = \alpha_{i0} + \alpha_{i1}\phi_{i,t-1}^2 + \kappa_{i1}I(\phi_{t-1}<0)\phi_{i,t-1}^2 + \beta_{i1}h_{i,t-1}$$

$i = 1,\ldots,N$. We make the following assumptions, see Silvennoinen and Teräsvirta (2021):

AN1. In (4), $\alpha_{i0} > 0$, either $\alpha_{i1} > 0$ and $\alpha_{i1}+\kappa_{i1} \geq 0$ or $\alpha_{i1} \geq 0$ and $\alpha_{i1}+\kappa_{i1} > 0$, $\beta_{i1} \geq 0$, and $\alpha_{i1} + \kappa_{i1}/2 + \beta_{i1} < 1$ for $i = 1,\ldots,N$.

AN2. The parameter subspaces $\{\alpha_{i0} \times \kappa_i \times \alpha_i \times \beta_i\}$, $i = 1,\ldots,N$, are compact, the whole space Θ_h is compact, and the true parameter value $\boldsymbol{\theta}_h^0$ is an interior point of Θ_h.

AN3. $\boldsymbol{\zeta}_t \sim \mathrm{iid}N(\mathbf{0}, \boldsymbol{I}_N)$.

AN1 is the necessary and sufficient weak stationarity condition for the ith first-order GJR-GARCH equation. Assumption AN2 is a standard regularity condition required for proving asymptotic normality of maximum likelihood estimators of $\boldsymbol{\theta}_{hi}$, $i = 1,\ldots,N$. Assuming normality (A3) is a strong condition, but it is needed here; see Silvennoinen and Teräsvirta (2021).

These assumptions are sufficient for the maximum likelihood estimators of the GARCH parameters in single-equation GARCH models to be consistent and asymptotically normal. Rewrite (12) indicating the parameters in the relevant functions:

$$\begin{aligned}\ln f(\boldsymbol{\zeta}_t|\boldsymbol{\theta}) \propto &-(1/2)\sum_{i=1}^{N}\ln g_{it}(\boldsymbol{\theta}_{gi})-(1/2)\sum_{i=1}^{N}\ln h_{it}(\boldsymbol{\theta}_{hi})-(1/2)\ln|\boldsymbol{P}_t(\boldsymbol{\theta}_P)| \\ &-(1/2)\boldsymbol{\varepsilon}_t'\{\boldsymbol{S}_t(\boldsymbol{\theta}_g)\boldsymbol{D}_t(\boldsymbol{\theta}_g,\boldsymbol{\theta}_h)\boldsymbol{P}_t(\boldsymbol{\theta}_P)\boldsymbol{D}_t(\boldsymbol{\theta}_g,\boldsymbol{\theta}_h)\boldsymbol{S}_t(\boldsymbol{\theta}_g)\}^{-1}\boldsymbol{\varepsilon}_t.\end{aligned}$$

The parameters are estimated in turn: first estimate $\boldsymbol{\theta}_{gi}$ to obtain starting-values to joint estimation of $\boldsymbol{\theta}_g$ and $\boldsymbol{\theta}_P$. This is done assuming $h_{it}(\boldsymbol{\theta}_{hi}) \equiv 1$, $i = 1, \ldots, N$. Amado and Teräsvirta (2013) showed in the single-equation GJR-GARCH case that under regularity conditions the maximum likelihood estimator of $\boldsymbol{\theta}_{gi}$ is consistent and asymptotically normal. Silvennoinen and Teräsvirta (2021) generalised this result to MTV models. That means that joint estimation of $\boldsymbol{\theta}_g$ and $\boldsymbol{\theta}_P$ by maximum likelihood produces consistent estimates of these parameter vectors. If $\widehat{\boldsymbol{\theta}}_g$ and $\widehat{\boldsymbol{\theta}}_P$ are consistent and Assumptions AN1, AN2 and AN3 hold, then by Theorem 3.3 of Song, Fan and Kalbfleisch (2005), the maximum likelihood estimator of $\boldsymbol{\theta}_h$ is consistent and asymptotically normal. After estimating $\boldsymbol{\theta}_h$, the parameter vectors $\boldsymbol{\theta}_g$ and $\boldsymbol{\theta}_P$ are re-estimated. Iteration continues until convergence. Song et al. (2005) showed that the final maximum likelihood estimator of $\boldsymbol{\theta}$ is consistent and asymptotically normal. A more detailed description of the maximisation by parts applied to the present situation can be found in Appendix D, see also Silvennoinen and Teräsvirta (2021).

7 Evaluation of the MTV model

Once the model has been specified and estimated, it is worth having a final glance over to identify areas of misspecification. The tests in Section 5.1 were used to guide the choice of the functional form of the deterministic component, and a rejection of the null was seen as evidence of the current model still lacking in its specification. In that sense, the tests in Section 5.1 are seen as both specification and evaluation tests. It is worth reiterating that these specification tests were constructed at the stage when the GARCH part was not yet specified, that is, $h_t = 1$ in (10). However, when the deterministic part passes these tests, and an MTV-GARCH equation is subsequently estimated, there is room for additional checks in terms of model misspecification, beyond the presently final model specification. The tests in Amado and Teräsvirta (2017) are available for this purpose. They fall into three categories.

In the first one, the deterministic component is additively misspecified. In the context of the current MTV-GARCH model, the relevant case is a test for yet another transition in (2). The second test assesses the GARCH equation for additive misspecifiation. The concern here is validity of the maximum lags p or q. The final test is the 'test of no remaining ARCH', which is based on the idea of a sufficiently well-specified model managing to clear any autocorrelation from the squared standardised residuals. The test that suits each of these situations (or its robustified version to avoid the assumption of normality) is conveniently carried out following a set of steps outlined in Appendix.

It is worth pointing out that the tests here are applied to $\hat{\boldsymbol{P}}_t^{-1/2}\boldsymbol{\varepsilon}_t$, one series at a time. Efforts towards completing the tests in the complete N-variate system simultaneously would open up a vast number of permutations of various misspecification options. To manage the task, the recommendation is to focus on the univariate specifications one at a time, even with the acknowledgment of some potential for deviating from the asymptotically exact results. Simulations in Section B.2 indicate that applying the tests on the pre-filtered data has very little impact on the distributions of the test statistics. While the standard form of the misspecification tests suffers from minor oversizing, this is mostly corrected when using the robust version of the test.

The test for an additional correlation in Section 5.2 may also be used as an evaluation test. It is based on the completely specified univariate and correlation components, and therefore its role as a misspecifiation test of a complete model is just. The number of degrees of freedom in this test quickly becomes large with increasing N. One way of restricting this growth would be to assume that under the alternative only the eigenvalues of the correlation matrix are changing over time. The alternative would be a correlation matrix only if all correlations were identical, but an LM test can nevertheless be built on this assumption. Write the correlation matrix as $\boldsymbol{P}_t = \boldsymbol{Q}_t \boldsymbol{\Lambda}_t \boldsymbol{Q}_t'$, where $\boldsymbol{P}_t$ is defined as in (5), $\boldsymbol{\Lambda}_t$ is the matrix of eigenvalues and $\boldsymbol{Q}_t$ contains the corresponding eigenvectors. Simplify this by assuming $\boldsymbol{Q}_t = \boldsymbol{Q}$ and approximate $\boldsymbol{\Lambda}_t$, the eigenvalue matrix of (5), by $\boldsymbol{\Psi}_t = \sum_{k=0}^{K} \boldsymbol{\Psi}_k (t/T)^k$. Under the null hypothesis, $K = 0$. The resulting test statistic is derived and its small-sample properties studied in Silvennoinen and Teräsvirta (2017).

8 Big Four results

8.1 Modelling the error variances

As suggested in the Introduction, the return series may not be adequately described by a weakly stationary GARCH or GJR-GARCH model. From the plots in Figure 2 it is seen that the amplitude of clusters varies for all four banks, in particular during and after the financial crisis beginning 2008. There also seems to be a rather tranquil period between 2004 and 2008. This variation also shows in the autocorrelation functions of squared returns in Figure 3. In all four cases, autocorrelations decay very slowly as a function of the lag length. For this reason, modelling the returns has to be initiated by testing the stationarity hypothesis. As discussed in Section 5.1, the slow moving volatility specification is done first, followed by the TV-GJR-GARCH estimation. The test statistic from Section A.1 is calibrated using methods from Appendix B.1. In the first one, the period of calm spans spans November 2003 – October 2007. In the other, the window size is 400, chosen based on the simulations, see discussion in Appendix B.1. To increase the performance of the test, the entire sample of over 7000 observations is broken into subsections. Once one transition is found and estimated, the test is applied to both before and after this transition to see if there is another transition on either side. The process is continued until the null of no transition is not rejected at the 5% level. After this specification stage, the GJR-GARCH equations are estimated together with the time-varying component as a complete TV-GJR-GARCH model. The estimated equations are then checked for signs of misspecification using the tests from Amado and Teräsvirta (2017).

Estimated TV-GJR-GARCH equations results appear in Table 1 and the deterministic components in Appendix E. For interest, a GJR-GARCH equations without the deterministic component are also estimated, and the results appear in

		α_{i0}	α_{i1}	κ_{i1}	β_{i1}	pers	kurt
ANZ	GJR	0.020	0.039	0.044	0.929	0.991	3.76
		(0.005)	(0.007)	(0.008)	(0.008)		
	TV-GJR	0.111	0.015	0.046	0.792	0.831	3.02
		(0.016)	(0.005)	(0.007)	(0.027)		
CBA	GJR	0.035	0.060	0.063	0.886	0.977	3.66
		(0.005)	(0.008)	(0.011)	(0.010)		
	TV-GJR	0.107	0.021	0.065	0.813	0.867	3.06
		(0.014)	(0.006)	(0.010)	(0.020)		
NAB	GJR	0.065	0.077	0.075	0.850	0.964	3.68
		(0.009)	(0.010)	(0.014)	(0.014)		
	TV-GJR	0.152	0.021	0.058	0.731	0.780	3.03
		(0.019)	(0.006)	(0.009)	(0.030)		
WBC	GJR	0.031	0.045	0.058	0.910	0.985	3.70
		(0.006)	(0.007)	(0.010)	(0.009)		
	TV-GJR	0.079	0.015	0.041	0.829	0.864	3.02
		(0.011)	(0.004)	(0.006)	(0.020)		

Table 1: Univariate estimation results for the four banks. GJR is the GJR-GARCH(1,1) equation, TV-GJR is the TV-GJR-GARCH equation, standard deviations in parentheses

the same Table. It is seen that in all four cases the persistence strongly decreases after rescaling with the TV component. This is also indirectly obvious from the autocorrelations in Figure 4 that are considerably smaller than the ones in Figure 3. The main cause for this decrease lies in the coefficient of the lagged conditional variance whose estimate becomes smaller in the process. Estimates of the asymmetry parameter κ_{i1} slightly increase, so asymmetry becomes more pronounced when nonstationarity is properly taken care of. The table also contains the kurtosis estimates for the two GJR-GARCH processes, obtained using definitions in He and Teräsvirta (1999).

Figure 5 contains the estimated transitions. There are two conspicuous features in the figure. One is the downward shift around 2004, which for WBC is a long and rather smooth decline. This calm period coincides with the most recent mining boom, which was interrupted by the GFC. For all four banks the deterministic component remains higher after 2010 than it was before 2008. For WBC, the deterministic component slowly but steadily declines after the crisis. For the three others it remains constant. This can be seen from the estimated equations behind the figures in Appendix E.

Effects of the deterministic component g_{it} on the GARCH equations also become obvious by comparing the conditional variances from the GJR-GARCH equations (Figure 6) and from the TV-GJR-GARCH ones (Figure 7). It is seen that the nonstationarity around 2008–2010 in Figure 6 is no longer visible in Figure 7. This is true for all four banks.

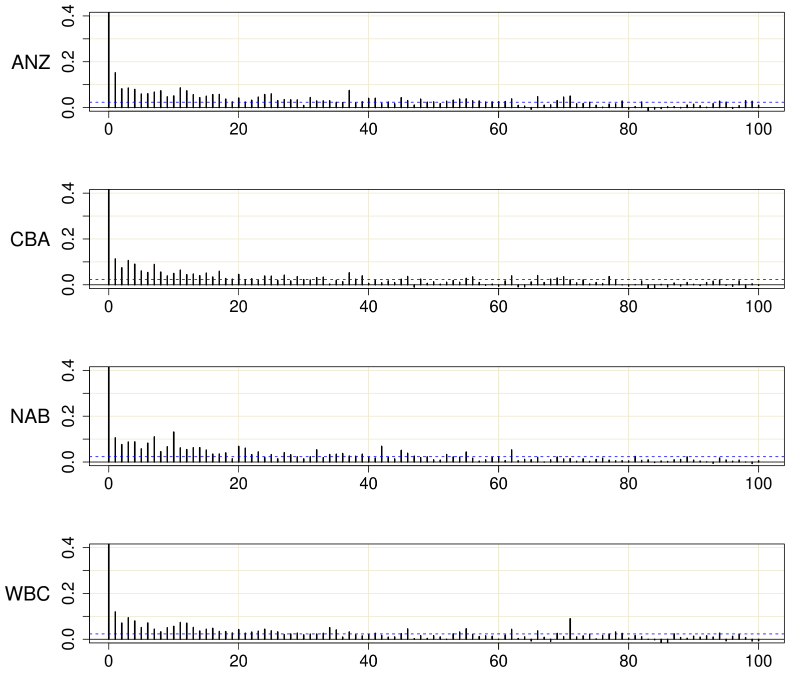

Figure 4: The first 100 autocorrelations of squared standardised returns $\varepsilon_{it}^2/g_{it}$. From the top: ANZ, CBA, NAB and WBC.

8.2 Modelling the error correlations

Stability of correlations over time is tested using the test statistic (A.5) in Section 5.2. The p-value of the test equals 0.000, so the null hypothesis is rejected. An STCC model is then estimated with a single monotonic time transition. Whether this specification is sufficient is tested with the test for an additional transition. The resulting p-value is 0.467, and the single time-transition is deemed sufficient. Estimation results of the MTV model in Table 2, see also Figure 8, indicate that the correlations between the standardised residuals have been stable, around 0.49–0.60, from 1992 until the mid to late 2007. At that point, the correlations begin their steady increase to the range of 0.78–0.83, which they reach by early 2008. The final correlations are not only quite large but also remarkably similar.

The time-varying correlation structure perhaps reflects the four pillars policy, the financial concentration which was further emphasised by the acquisition of the next largest banks by CBA and WBC in 2008. The competition between the banks was increased by the implementation of Basel II, the technological developments and the easing of restrictions that directly impacted the home mortgage market which the Big Four now dominated, and the stagnation of the housing credit market. As a further effect of the government intervention policy, the government had sympathetic policy settings, which allowed the banks to recapitalize in the 1990s and 2000s when three of the banks suffered financial losses and the fourth one (CBA) was privatised. Since the global financial crisis (GFC), the banks have enjoyed substantial government support (deposit guarantee) – as a result they have, in a sense, become

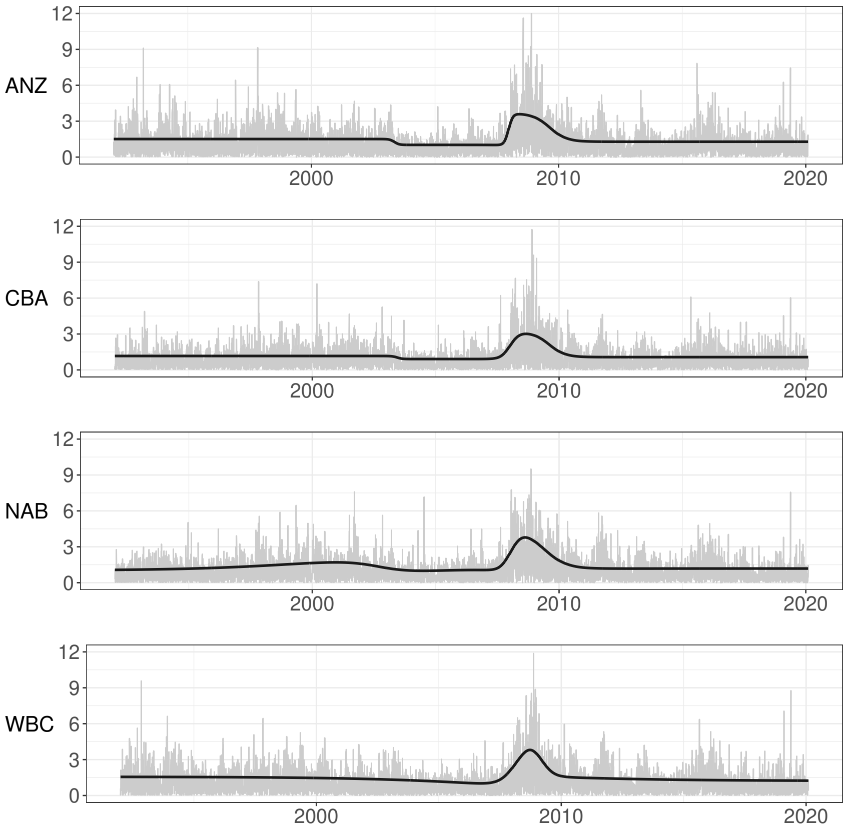

Figure 5: Estimated multiplicative component $\widehat{g}_{it}^{1/2}$ (solid curve) and the absolute returns $|\varepsilon_{it}|$ (grey area). From top to bottom: ANZ, CBA, NAB and WBC.

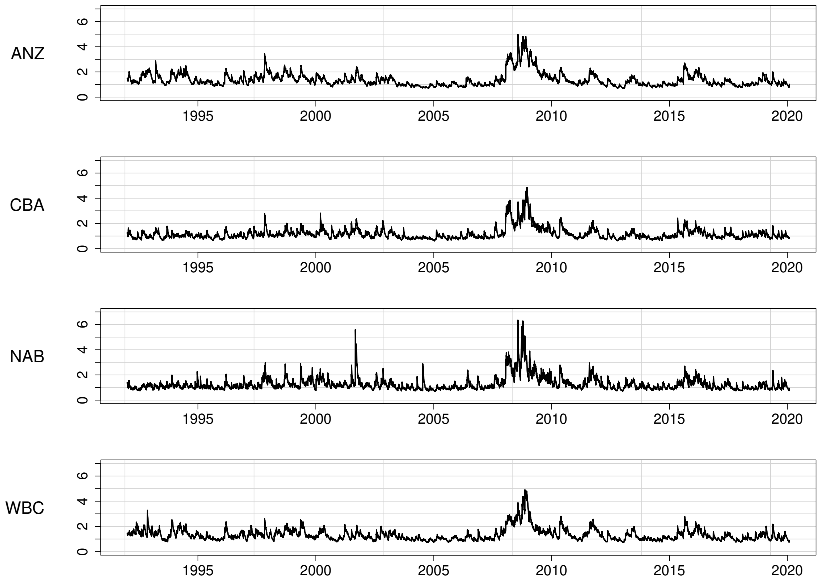

Figure 6: Estimated conditional variance $\widehat{h}_{it}$ from the GJR-GARCH model. From top to bottom: ANZ, CBA, NAB and WBC.

and remained similar.

The fact that serious effort has been made to model the volatilities and correlations separately allows for observations on the timing and magnitude of those features without the cross-contamination occurring if covariances were examined instead. It is often noted that correlations increase during turbulent times. In the case of the Big Four, the situation is the opposite. The calm period 2003-2007 is around the middle of the mining boom. It is only towards the end of this period when the correlations have smoothly increased over a span of 16 months. Furthermore, it is notable that the events around the GFC have tremendous impact on the volatilities, whereas the correlations have by then settled to their high levels, and exhibit no further increase. This does, however, highlight the potentially fragile state of the Australia's financial sector, which may have become exposed to shocks that are likely to have contagious impacts on the financial markets, and Australia's economy as a whole.

9 Conclusions

In this paper we describe the steps forming the modelling cycle for MTV models. An R-package has been built to assist in this process. Our analysis of the Big Four reveals a systematic increase in similarity of the four banks as assets, reducing their contribution towards diversification of risk in an investment portfolio. This shift in similarity occurs before the GFC amid the prevalent mining boom and tranquil markets. The regulatory changes and technological development have paved the way

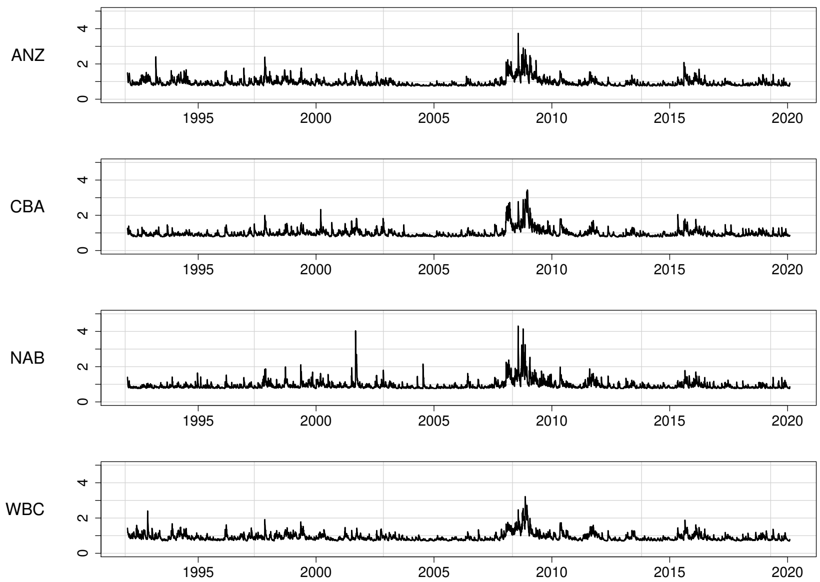

Figure 7: Estimated conditional variance $\widehat{h}_{it}$ from the TV-GJR-GARCH model. From top to bottom: ANZ, CBA, NAB and WBC.

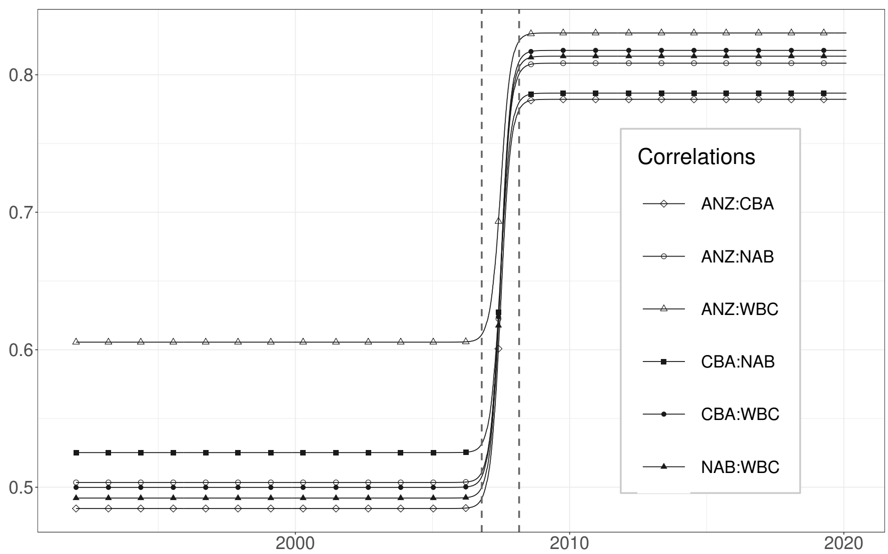

Figure 8: Estimated correlations. Vertical lines correspond to Oct-2007 and Feb-2008.

	$\boldsymbol{P}_1$				$\boldsymbol{P}_2$		
	ANZ	CBA	NAB		ANZ	CBA	NAB
CBA	0.485 (0.011)	1.000		CBA	0.782 (0.006)	1.000	
NAB	0.503 (0.010)	0.525 (0.010)	1.000	NAB	0.808 (0.005)	0.787 (0.005)	1.000
WBC	0.606 (0.009)	0.500 (0.011)	0.492 (0.011)	WBC	0.830 (0.004)	0.818 (0.005)	0.814 (0.005)

transition parameters:	c	η
	0.552 (0.002)	5.020 (0.162)

Table 2: Estimation results for the four banks' time-varying correlations. 90% of the estimated transition is between the dates 2006/Oct/18 and 2008/Feb/28. The centerpoint of the location corresponds to 2007/Jun/28, with ± two standard deviations range of 2007/May/11 – 2007/Aug/13.

for increased competition prior to the shift in the correlation. In short, the highly concentrated financial sector is also highly correlated. The government intervention in terms of the tightening of the lending rules has opened the home loan market to more competition, which has reduced the share of the Big Four within that market and forced differentiation between the banks. The recent Royal Commission (2017-2019) has been speculated to have further induced the Big Four's willingness to be different. However, our estimation results suggest that the attempts by the government to diversify the Big Four as assets in an investor portfolio have not yet carried fruit.

References

Amado, C., Silvennoinen, A. and Teräsvirta, T.: 2017, Modelling and forecasting WIG20 daily returns, *Central European Journal of Economic Modelling and Econometrics* **9**, 173–200.

Amado, C. and Teräsvirta, T.: 2008, Modelling conditional and unconditional heteroskedasticity with smoothly time-varying structure, *SSE/EFI Working Paper Series in Economics and Finance 691*, Stockholm School of Economics.

Amado, C. and Teräsvirta, T.: 2013, Modelling volatility by variance decomposition, *Journal of Econometrics* **175**, 153–165.

Amado, C. and Teräsvirta, T.: 2014, Conditional correlation models of autoregressive conditional heteroscedasticity with nonstationary GARCH equations, *Journal of Business and Economic Statistics* **32**, 69–87.

Amado, C. and Teräsvirta, T.: 2017, Specification and testing of multiplicative time-varying GARCH models with applications, *Econometric Reviews* **36**, 421–446.

Berben, R.-P. and Jansen, W. J.: 2005, Comovement in international equity markets: A sectoral view, *Journal of International Money and Finance* **24**, 832–857.

Bollerslev, T.: 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307–327.

Bollerslev, T.: 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *Review of Economics and Statistics* **72**, 498–505.

Box, G. E. P. and Jenkins, G. M.: 1970, *Time Series Analysis, Forecasting And Control*, Holden-Day, San Francisco.

Chan, F. and Theoharakis, B.: 2011, Estimating m-regimes STAR–GARCH model using QMLE with parameter transformation, *Mathematics and Computers in Simulation* **81**, 1385–1396.

Ekner, L. E. and Nejstgaard, E.: 2013, Parameter identification in the logistic STAR model, *Discussion Paper 13-07*, Department of Economics, University of Copenhagen.

Engle, R. F.: 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* **20**, 339–350.

Feng, Y.: 2004, Simultaneously modeling conditional heteroskedasticity and scale change, *Econometric Theory* **20**, 563–596.

Feng, Y.: 2006, A local dynamic conditional correlation model, *MPRA Paper 1592*, Maxwell Institute for Mathematical Sciences, Heriot-Watt University.

Glosten, L. W., Jagannathan, R. and Runkle, D. E.: 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**, 1779–1801.

Godfrey, L. G.: 1988, *Misspecification Tests in Econometrics*, Cambridge University Press, Cambridge.

Goodwin, B. K., Holt, M. T. and Prestemon, J. P.: 2011, North American oriented strand board markets, arbitrage activity, and market price dynamics: A smooth transition approach, *American Journal of Agricultural Economics* **93**, 993–1014.

He, C. and Teräsvirta, T.: 1999, Properties of moments of a family of GARCH processes, *Journal of Econometrics* **92**, 173–192.

Lütkepohl, H.: 1996, *Handbook of Matrices*, John Wiley & Sons, Chichester.

Luukkonen, R., Saikkonen, P. and Teräsvirta, T.: 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Silvennoinen, A. and Teräsvirta, T.: 2005, Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations, SSE/EFI Working Paper Series in Economics and Finance No. 577.

Silvennoinen, A. and Teräsvirta, T.: 2009, Modelling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model, *Journal of Financial Econometrics* **7**, 373–411.

Silvennoinen, A. and Teräsvirta, T.: 2015, Modeling conditional correlations of asset returns: A smooth transition approach, *Econometric Reviews* **34**, 174–197.

Silvennoinen, A. and Teräsvirta, T.: 2016, Testing constancy of unconditional variance in volatility models by misspecification and specification tests, *Studies in Nonlinear Dynamics and Econometrics* **20**, 347–364.

Silvennoinen, A. and Teräsvirta, T.: 2017, A parsimonious test of constancy of a positive definite correlation matrix in a multivariate time-varying GARCH model, *work in progress*, Queensland University of Technology, Brisbane.

Silvennoinen, A. and Teräsvirta, T.: 2021, Consistency and asymptotic normality of maximum likelihood estimators of the multiplicative time-varying smooth transition correlation GARCH model, *Econometrics and Statistics* (in press) .

Song, P. X., Fan, Y. and Kalbfleisch, J. D.: 2005, Maximization by parts in likelihood inference, *Journal of the American Statistical Association* **100**, 1145–1158.

Teräsvirta, T.: 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J.: 2010, *Modelling Nonlinear Economic Time Series*, Oxford University Press, Oxford.

Tse, Y. K. and Tsui, K. C.: 2002, A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations, *Journal of Business and Economic Statistics* **20**, 351–362.

Online Appendix

This Appedix contains additional material to the paper. Section A provides details of the TVV-model specification, the MTV-GARCH model evaluation, the test of constant correlations, and finally the test for an additional transition in the correlations. The simulations studies in Section B explore aspects of the specification and evaluation of the GARCH equations, and the size and sensitivity of the test of constant correlations. Proof of Lemma 1 is presented in Section C. Section D presents the details of maximisation by parts. Estimated deterministic components of the Four Banks' transition equations are presented in Section E. Finally, Sections F and G provide tabulated results and figures related to the simulation studies.

A Test statistics

A.1 Test statistic for TVV-model specification

In order to specify g_t we not only test constancy but even specify the number of transitions before estimating the GARCH component of the model. Amado and Teräsvirta (2013) showed that maximum likelihood estimators of the corresponding time-varying variance (TVV) model, assuming that there is no conditional heteroskedasticity, are consistent and asymptotically normal. This forms the base for constructing Lagrange multiplier type tests for testing r against $r+1$ transitions. For notational simplicity consider testing one transition against two. Omitting the subscript i for simplicity, the TVV model is (9) with $h_t = 1$, and

$$g_t = \delta_0 + \delta_1 G_1(t/T, \gamma_1, \boldsymbol{c}_1) + \delta_2 G_2(t/T, \gamma_2, \boldsymbol{c}_2),\ \gamma_i > 0,\ i = 1, 2.$$

The null hypothesis is $\gamma_2 = 0$, in which case $G_2(t/T, \gamma_2, \boldsymbol{c}_2) \equiv 1/2$. To circumvent the identification problem (the model with one transition is only identified when the alternative $\gamma_2 > 0$ is true) we follow Luukkonen, Saikkonen and Teräsvirta (1988) and approximate the second transition by a third-order Taylor expansion around the null hypothesis. After reparameterisation this yields

$$g_t = \delta_0^{*0} + \delta_1 G_1(t/T, \gamma_1, \boldsymbol{c}_1) + \psi_1 t/T + \psi_2 (t/T)^2 + \psi_3 (t/T)^3,\ \gamma_1 > 0. \tag{A.1}$$

We may call (9) with (A.1) the auxiliary TVV model. The parameters $\psi_i = \gamma_2 \widetilde{\psi}_i$, where $\widetilde{\psi}_i \neq 0$, $i = 1, 2, 3$. The new null hypothesis in (A.1) equals H_0': $\psi_1 = \psi_2 = \psi_3 = 0$. The remainder term of the expansion can be ignored because when we construct a Lagrange multiplier test, the model is only estimated under H_0 (or H_0') and under this hypothesis the order of the Taylor expansion equals zero. The remainder is present only under the alternative, and so ignoring it when H_0 is valid does not affect the asymptotic size of the test. It does make a positive contribution to the power of the test when H_0 does not hold.

Assume (again for notational simplicity) that $K_1 = 1$ in (A.1), so $\boldsymbol{c}_1 = c_1$ (a scalar). The log-likelihood for observation t of the auxiliary TVV model equals

$$\ell_t = k - (1/2)\ln g_t - (1/2)\frac{\varepsilon_t^2}{g_t}$$

and the corresponding element of the score is

$$\frac{\partial \ell_t}{\partial \boldsymbol{\theta}_1} = \frac{1}{2}(\frac{\varepsilon_t^2}{g_t} - 1)\frac{1}{g_t}\frac{\partial g_t}{\partial \boldsymbol{\theta}_1} \tag{A.2}$$

where $\boldsymbol{\theta}_1 = (\delta_0^{*0}, \delta_1, \gamma_1, c_1, \psi_1, \psi_2, \psi_3)'$. Denoting $G_1(t/T) = G_1(t/T, \gamma_1, \boldsymbol{c}_1)$, the partial derivative in (A.2) is $\partial g_t / \partial \boldsymbol{\theta}_1 = (\boldsymbol{g}_1'(t/T), \boldsymbol{\tau}_t')'$where

$$\boldsymbol{g}_1(t/T) = (1, G_1(t/T), G_{1\gamma}(t/T), G_{1c}(t/T)G_{1\gamma}(t/T))'$$

with $G_{1\gamma}(t/T) = G_1(t/T)(1 - G_1(t/T))(t/T - c_1)$, $G_{1c}(t/T) = -\gamma_1 G_1(t/T)(1 - G_1(t/T))$, and $\boldsymbol{\tau}_t = (t/T, (t/T)^2, (t/T)^3)$. Define the true parameter vector under H_0 as $\boldsymbol{\theta}_1^0 = (\delta_0^{*0}, \delta_1^0, \gamma_1^0, c_1^0, 0, 0, 0)'$. If $\boldsymbol{z}_t$ is normally distributed, the corresponding element of the information matrix under H_0 has the form

$$\boldsymbol{B}_t = \frac{1}{4}\mathsf{E}(\frac{\varepsilon_t^2}{g_t} - 1)^2 \begin{bmatrix} \boldsymbol{B}_{11t} & \boldsymbol{B}_{12t} \\ \boldsymbol{B}_{21t} & \boldsymbol{B}_{22t} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \boldsymbol{B}_{11t} & \boldsymbol{B}_{12t} \\ \boldsymbol{B}_{21t} & \boldsymbol{B}_{22t} \end{bmatrix}.$$

where, letting $g_t^0 = g^0(t/T) = \delta_0^{*0} + \delta_1^0 G_1(t/T, \gamma_1^0, \boldsymbol{c}_1^0)$ and denoting $G_1^0(t/T) = G_1(t/T, \gamma_1^0, \boldsymbol{c}_1^0)$,

$$\boldsymbol{B}_{11t} = \frac{1}{2(g^0(t/T))^2}\boldsymbol{g}_1^0(t/T)\boldsymbol{g}_1^0(t/T)',\ \boldsymbol{B}_{12t} = \frac{1}{2(g^0(t/T))^2}\boldsymbol{g}_1^0(t/T)\boldsymbol{\tau}_t'$$

and

$$\boldsymbol{B}_{22t} = \frac{1}{2(g^0(t/T))^2}\boldsymbol{\tau}_t\boldsymbol{\tau}_t'.$$

Let

$$\boldsymbol{g}_1^0(r) = (1, G_1^0(r), G_{1\gamma}^0(r), G_{1c}^0(r))'$$

and $\boldsymbol{r} = (r, r^2, r^3)'$. We state the following lemma:

Lemma 1 *Under the null hypothesis and assuming $z_t \sim iid\mathcal{N}(0,1)$, the information matrix*

$$\boldsymbol{B} = \lim_{T\to\infty}\frac{1}{2T}\sum_{t=1}^{T}\begin{bmatrix} \boldsymbol{B}_{11t} & \boldsymbol{B}_{12t} \\ \boldsymbol{B}_{21t} & \boldsymbol{B}_{22t} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix}$$

where

$$\boldsymbol{B}_{11} = \frac{1}{2}\int_0^1 (g^0(r))^{-2}\boldsymbol{g}_1^0(r)\boldsymbol{g}_1^0(r)'dr,\ \boldsymbol{B}_{12} = \frac{1}{2}\int_0^1 (g^0(r))^{-2}\boldsymbol{g}_1^0(r)\boldsymbol{r}'dr$$

and

$$\boldsymbol{B}_{22} = \frac{1}{2}\int_0^1 (g^0(r))^{-2}\boldsymbol{r}\boldsymbol{r}'dr.$$

Proof. See Appendix C.

Since the maximum likelihood estimators of the parameters of the auxiliary TVV model under H_0 are consistent, we may construct the LM test for the hypothesis H_0': $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3)' = \mathbf{0}$. Denoting the relevant block of the score by

$$\boldsymbol{s}_2(\widehat{\boldsymbol{\theta}}_1) = \frac{1}{2T}\sum_{t=1}^{T}(\frac{\varepsilon_t^2}{\widehat{g}_t} - 1)\frac{1}{\widehat{g}_t}\frac{\partial g_t}{\partial \boldsymbol{\psi}}$$

where $\partial g_t/\partial \boldsymbol{\psi} = \boldsymbol{\tau}_t$ and, assuming $z_t = \varepsilon_t/g_t^{1/2}$ is standard normal under H_0,

$$\widehat{g}_t = \widehat{\delta}_0 + \widehat{\delta}_1(1 + \exp\{-\widehat{\gamma}_1(t/T - \widehat{c}_1)\})^{-1}$$

the test statistic has the following form:

$$LM_T^1 = (T/2)\boldsymbol{s}_2'(\widehat{\boldsymbol{\theta}}_1)(\boldsymbol{B}_{22} - \boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12})^{-1}\boldsymbol{s}_2(\widehat{\boldsymbol{\theta}}_1) \quad \text{(A.3)}$$

where $\widehat{\boldsymbol{\theta}}_1 = (\widehat{\delta}_0^*, \widehat{\delta}_1, \widehat{\gamma}_1, \widehat{c}_1, 0, 0, 0)'$, see for example Godfrey (1988, p. 14). In order to make (A.3) operational, the blocks of $\boldsymbol{B}$ are replaced by their consistent counterparts.

When constancy of the error variance is tested against a single transition, $g_t \equiv \delta_0$, $\boldsymbol{g}_1(t/T) = 1$ (scalar), and $\partial g_t/\partial \boldsymbol{\psi} = \boldsymbol{\tau}_t$ as before. Then $\boldsymbol{B}_{11} = (2(\delta_0^0)^2)^{-1}$,

$$\boldsymbol{B}_{12} = \frac{1}{2(\delta_0^0)^2}\int_0^1 \boldsymbol{r}'\mathrm{d}r \text{ and } \boldsymbol{B}_{22} = \frac{1}{2(\delta_0^0)^2}\int_0^1 \boldsymbol{r}\boldsymbol{r}'\mathrm{d}r.$$

The test statistic (A.3) becomes

$$LM_T^0 = \frac{T(\delta_0^0)^2}{2\widehat{\delta}_0^2}\boldsymbol{s}_2'(\widehat{\boldsymbol{\theta}}_1)(\boldsymbol{B}_{22} - \boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12})^{-1}\boldsymbol{s}_2(\widehat{\boldsymbol{\theta}}_1) \quad \text{(A.4)}$$

where

$$\boldsymbol{s}_2(\widehat{\boldsymbol{\theta}}_1) = \frac{1}{2T}\sum_{t=1}^{T}(\frac{\varepsilon_t^2}{\widehat{\delta}_0} - 1)\boldsymbol{\tau}_t.$$

When the elements of the covariance matrix are replaced by their consistent estimators in (A.4), the ratio $(\delta_0^0)^2/\widehat{\delta}_0^2$ equals unity.

As already mentioned, conditional heteroskedasticity is ignored in setting up the test. For this reason, the test statistic (A.3) is likely to be size distorted when applied to financial time series of sufficiently high frequency, that is, GARCH-type volatility clustering is present. In applications its size thus has to be adjusted by calibrating its distribution to reflect the persistence of the GARCH effect present in the data. This is the topic of discussion in Section B.1.

A.2 Test statistic for MTV-GARCH model evaluation

In this Section the evaluation tests of the univariate MTV-GARCH equations are presented in an easy to implement fashion. The full details can be found in Amado and Teräsvirta (2017).

The test statistic is computed based on the following components: $\hat{\zeta}_t$, $\boldsymbol{r}_{1t}$, and $\boldsymbol{r}_{2t}$. $\hat{\zeta}_t = \varepsilon_t/\sqrt{\hat{h}_t\hat{g}_t}$ are the residuals, $\boldsymbol{r}_{1t}$ contains the derivatives of the h_t and g_t functions with respect to the parameters that govern the MTV-GARCH model under the null, $\boldsymbol{\theta}_g$ and $\boldsymbol{\theta}_h$:

$$\boldsymbol{r}_{1t} = (\hat{g}_t^{-1}\frac{\partial g_t}{\partial \hat{\boldsymbol{\theta}}_g} + \hat{h}_t^{-1}\frac{\partial h_t}{\partial \hat{\boldsymbol{\theta}}_g}, \hat{h}_t^{-1}\frac{\partial h_t}{\partial \hat{\boldsymbol{\theta}}_h}),$$

evaluated at the estimated parameters $\hat{\boldsymbol{\theta}}_g$ and $\hat{\boldsymbol{\theta}}_h$. These are recursively calculated, and depend on the prevailing MTV-GARCH model under the null. For example, $g_t = \delta_0 + \delta_1 G_1(t/T, \gamma_1, c_1) + \delta_2 G_2(t/T, \gamma_2, (c_{21}, c_{22})')$ and $h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2/g_{t-1} + \kappa_1 I(\varepsilon_{t-1} < 0)\varepsilon_{t-1}^2/g_{t-1} + \beta_1 h_{t-1}$. Here, $\boldsymbol{\theta}_g = (\delta_1, \gamma_1, c_1, \delta_2, \gamma_2, c_{21}, c_{22})'$ and $\boldsymbol{\theta}_h = (\alpha_0, \alpha_1, \kappa_1, \beta_1)'$.

Then
$$\frac{\partial g_t}{\partial \boldsymbol{\theta}_g} = (G_1, \delta_1 \frac{\partial G_1}{\partial \gamma_1}, \delta_1 \frac{\partial G_1}{\partial c_1}, G_2, \delta_2 \frac{\partial G_2}{\partial \gamma_2}, \delta_2 \frac{\partial G_2}{\partial c_{21}}, \delta_2 \frac{\partial G_2}{\partial c_{22}})'$$
where
$$\begin{aligned}
\frac{\partial G_1}{\partial \gamma_1} &= G_1(1-G_1)(t/T - c_1) \\
\frac{\partial G_1}{\partial c_1} &= -G_1(1-G_1)\gamma_1 \\
\frac{\partial G_2}{\partial \gamma_2} &= G_2(1-G_2)(t/T - c_{21})(t/T - c_{22}) \\
\frac{\partial G_2}{\partial c_{21}} &= -G_2(1-G_2)\gamma_2(t/T - c_{22}) \\
\frac{\partial G_2}{\partial c_{21}} &= -G_2(1-G_2)\gamma_2(t/T - c_{21})
\end{aligned}$$
The GARCH equation derivatives are formed recursively as
$$\frac{\partial h_t}{\partial \boldsymbol{\theta}_g} = -g_t^{-1}(\alpha_1 \varepsilon_{t-1}^2/g_{t-1} + \kappa_1 I(\varepsilon_{t-1} < 0)\varepsilon_{t-1}^2/g_{t-1})\frac{\partial g_{t-1}}{\partial \boldsymbol{\theta}_g} + \beta_1 \frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_g}$$
and
$$\frac{\partial h_t}{\partial \boldsymbol{\theta}_h} = (\varepsilon_{t-1}^2/g_{t-1}, I(\varepsilon_{t-1} < 0)\varepsilon_{t-1}^2/g_{t-1}, h_{t-1})' + \beta_1 \frac{\partial h_{t-1}}{\partial \boldsymbol{\theta}_h}$$
From this example, it should be easy to extend the null model to include more additive deterministic terms and / or have a higher order GARCH equation with or without asymmetric terms.

One extension regarding the deterministic part should be mentioned though. It is often convenient to replace the slope parameter γ with e^{η}. In this case, $\boldsymbol{\theta}_g = (\delta_1, \eta_1, c_1, \delta_2, \eta_2, c_{21}, c_{22})'$, and
$$\begin{aligned}
\frac{\partial G_1}{\partial \eta_1} &= G_1(1-G_1)e^{\eta_1}(t/T - c_1) \\
\frac{\partial G_1}{\partial c_1} &= -G_1(1-G_1)e^{\eta_1} \\
\frac{\partial G_2}{\partial \eta_2} &= G_2(1-G_2)e^{\eta_2}(t/T - c_{21})(t/T - c_{22}) \\
\frac{\partial G_2}{\partial c_{21}} &= -G_2(1-G_2)e^{\eta_2}(t/T - c_{22}) \\
\frac{\partial G_2}{\partial c_{21}} &= -G_2(1-G_2)e^{\eta_2}(t/T - c_{21})
\end{aligned}$$

$\boldsymbol{r}_{2t}$ contains the derivatives of the misspecified part. Details of it in the most commonly encountered situations are given shortly. The number of variables (columns) in $\boldsymbol{r}_{2t}$ defines the degrees of freedom in the χ^2 distribution for the test statistic under the null.

Given the three components, the LM-test is carried out as follows:

1. Compute the $SSR_0 = \sum_{t=1}^{T}(\hat{\zeta}_t^2 - 1)^2$.

2. Regress $\hat{\zeta}_t^2 - 1$ on $(\boldsymbol{r}_{1t}, \boldsymbol{r}_{2t})$, and form the sum of squared residuals SSR_1.

3. Compute the test statistic $LM = T\frac{SSR_0 - SSR_1}{SSR_0}$.

The robust version that does not rely on the normality of the error term is formed as follows:

1. Compute the $SSR_0 = \sum_{t=1}^{T}(\hat{\zeta}_t^2 - 1)^2$.
2. Regress ζ_t^2 on $\boldsymbol{r}_{1t}$, and obtain residuals w_t. When $\boldsymbol{r}_{2t}$ has more than one variable, run the regression for each of them separately, and thereby obtain a set of residuals $\boldsymbol{w}_t$.
3. Regress $\mathbf{1}$ on $(\hat{z}_t^2 - 1)\boldsymbol{w}_t$, and form the sum of squared residuals SSR_R.
4. Compute the test statistic $LM_R = T - SSR_R$

The first case seeks to find evidence of misspecification of the determininstic part of the MTV-GARCH model. That is, the conditional variance is of the form

$$\sigma_t^2 = h_t(g_t + f_t)$$

where the additive term f_t is zero under the null of the model being correctly specified. The case we consider here is the one of testing r against $r+1$ transitions in the deterministic part. The additive term is linearised and reparameterised, after which it becomes

$$f_t = \delta_0^* + \delta_1^* t/T + \delta_2^*(t/T)^2 + \delta_3^*(t/T)^3$$

The derivative component for the alternative is then

$$\boldsymbol{r}_{2t} = \hat{g}_t^{-1}(1, t/T, (t/T)^2), (t/T)^3))$$

The second case addresses misspecification in the GARCH part:

$$\sigma_t^2 = (h_t + f_t)g_t$$

where the additive term f_t is again zero under the null. A common scenario is when f_t may increase either the ARCH or the GARCH order (but not both). An example of the former is GARCH(1,1) vs GARCH(2,1), in which case $f_t = \alpha_2\varepsilon_{t-2}^2/g_{t-2}$, and therefore

$$\boldsymbol{r}_{2t} = \hat{h}_t^{-1}\varepsilon_{t-2}^2/\hat{g}_{t-2}$$

If the model is a GJR one, and the potential increase in the order of the ARCH term extends to the asymmetric terms as well, then $f_t = \alpha_2\varepsilon_{t-2}^2/g_{t-2} + \kappa_2 I(\varepsilon_{t-2} < 0)\varepsilon_{t-2}^2/g_{t-2}$, and

$$\boldsymbol{r}_{2t} = \hat{h}_t^{-1}(\varepsilon_{t-2}^2/\hat{g}_{t-2}, I(\varepsilon_{t-2} < 0)\varepsilon_{t-2}^2/\hat{g}_{t-2})$$

An example of the latter is GARCH(1,1) vs GARCH(2,1), which leads to $f_t = \beta_2 h_{t-2}$, and so

$$\boldsymbol{r}_{2t} = \hat{h}_t^{-1}\hat{h}_{t-2}$$

The third case is the test of no remaining ARCH. This is a test against multiplicative misspecification,

$$\sigma_t^2 = h_t g_t f_t$$

where $f_t = 1$ under the null. If the alternative is that there is ARCH of order m, then

$$\boldsymbol{r}_{2t} = (\hat{z}^2_{t-1}, \ldots, \hat{z}^2_{t-m})$$

A.3 Test of constant correlations

The log-likelihood of the auxiliary MTV model for observation t assuming $K = 2$ equals

$$\begin{aligned}\ln f_A(\boldsymbol{\zeta}_t|\boldsymbol{\theta}) = &- (1/2)\sum_{i=1}^{N} \ln g_{it} - (1/2)\sum_{i=1}^{N} \ln h_{it} - (1/2)\ln|\boldsymbol{P}_{At}| \\ &- (1/2)\boldsymbol{\varepsilon}'_t\{\boldsymbol{S}_t\boldsymbol{D}_t\boldsymbol{P}_{At}\boldsymbol{D}_t\boldsymbol{S}_t\}^{-1}\boldsymbol{\varepsilon}_t\end{aligned}$$

where

$$\boldsymbol{P}_{At} = \boldsymbol{P}_{(A0)} + (t/T)\boldsymbol{P}_{(A1)} + (t/T)^2\boldsymbol{P}_{(A2)}$$

and $g_{it} = \delta_{i0} + \delta_{i1}G_{i1}(t/T, \gamma_{i1}, c_{i1})$; only one transition for notational simplicity, and h_{it} is as in (10). The first sub-block of the score corresponding to the deterministic variance component under H_0 becomes

$$\boldsymbol{s}_t(\boldsymbol{\theta}_{gi}) = -\frac{1}{2}(g_{it}^{-1}\frac{\partial g_{it}}{\partial \boldsymbol{\theta}_{gi}} + h_{it}^{-1}\frac{\partial h_{it}}{\partial \boldsymbol{\theta}_{gi}})(1 - \boldsymbol{e}'_i\boldsymbol{P}^{-1}_{(A0)}\boldsymbol{z}_t\boldsymbol{z}'_t\boldsymbol{e}_i)$$

where $\boldsymbol{e}_i = (\boldsymbol{0}'_{i-1}, 1, \boldsymbol{0}'_{N-i})'$, $i = 1, \ldots, N$, and $\boldsymbol{0}_0$ is an empty set. The sub-block corresponding to the GARCH parameters under H_0 is

$$\boldsymbol{s}_t(\boldsymbol{\theta}_{hi}) = -\frac{1}{2}(h_{it}^{-1}\frac{\partial h_{it}}{\partial \boldsymbol{\theta}_{hi}})(1 - \boldsymbol{e}'_i\boldsymbol{P}^{-1}_{(A0)}\boldsymbol{z}_t\boldsymbol{z}'_t\boldsymbol{e}_i)$$

$i = 1, \ldots, N$. The remaining sub-blocks under H_0 equal

$$\begin{aligned}\boldsymbol{s}_t(\boldsymbol{\rho}_{Aj}) &= -\frac{1}{2}\frac{\partial \mathrm{vec}(\boldsymbol{P}_{At})'}{\partial \boldsymbol{\rho}_{Aj}}\{\mathrm{vec}(\boldsymbol{P}^{-1}_{(A0)}) - (\boldsymbol{P}^{-1}_{(A0)} \otimes \boldsymbol{P}^{-1}_{(A0)})\mathrm{vec}(\boldsymbol{z}_t\boldsymbol{z}'_t)\} \\ &= -\frac{1}{2}(t/T)^j\boldsymbol{U}'\{\mathrm{vec}(\boldsymbol{P}^{-1}_{(A0)}) - (\boldsymbol{P}^{-1}_{(A0)} \otimes \boldsymbol{P}^{-1}_{(A0)})\mathrm{vec}(\boldsymbol{z}_t\boldsymbol{z}'_t)\}\end{aligned}$$

$j = 0, 1, 2$, where $\boldsymbol{U} = \partial\mathrm{vec}(\boldsymbol{P}_{(Aj)})/\partial\boldsymbol{\rho}'_{Aj}$ consists of zeroes and ones and is identical for all j. The $N^2 \times N(N-1)/2$ matrix $\boldsymbol{U}$ is a columnwise collection of vectorised indicator matrices that identify the locations of the particular correlation parameter within the matrix $\boldsymbol{P}_{At}$. For example, the first correlation parameter in $\boldsymbol{\rho}_{Aj}$ is located in positions (2,1) and (1,2) in $\boldsymbol{P}_{At}$. An indicator matrix corresponding to this parameter has ones in those positions, and zeros elsewhere. This vectorised indicator matrix is then the first column of matrix $\boldsymbol{U}$, and so on. Consequently, the $3N(N-1)/2 \times N^2$ matrix $\partial\mathrm{vec}(\boldsymbol{P}_{At})'/\partial\boldsymbol{\rho}_A$ equals

$$\frac{\partial \mathrm{vec}(\boldsymbol{P}_{At})'}{\partial \boldsymbol{\rho}_A} = \begin{bmatrix} 1 \\ (t/T) \\ (t/T)^2 \end{bmatrix} \otimes \boldsymbol{U}'.$$

The information matrix for observation t under H_0 is quite similar to but simpler than the corresponding one in Silvennoinen and Teräsvirta (2021). In order to give the matrix a proper expression, we need the commutation matrix $\boldsymbol{K}$, an $N^2 \times N^2$ matrix whose (i, j) block equals $\boldsymbol{e}_j\boldsymbol{e}'_i$, that is, $[\boldsymbol{K}]_{ij} = \boldsymbol{e}_j\boldsymbol{e}'_i$, see for example Lütkepohl (1996, pp. 115–118).

Let the superscript 0 indicate that the corresponding entity is evaluated under H_0 (for example, g_{it}^0 equals $g_{it}|_{\mathrm{H}_0}$, and $\partial g_{it}^0/\partial\boldsymbol{\theta}_{gi}$ equals $\partial g_{it}/\partial\boldsymbol{\theta}_{gi}|_{\mathrm{H}_0}$). The matrix is defined in the following lemma.

Lemma 2 *The expectations of the nine blocks of the information matrix at (rescaled) time t/T under H_0: $\boldsymbol{\rho}_{A1} = \boldsymbol{\rho}_{A2} = \mathbf{0}_{N(N-1)/2}$ are*

$$\boldsymbol{B}_t^0 = \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}^0)\boldsymbol{s}_t'(\boldsymbol{\theta}^0) = \mathsf{E}\begin{bmatrix} \boldsymbol{s}_t(\boldsymbol{\theta}_g^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_g^0) & \boldsymbol{s}_t(\boldsymbol{\theta}_g^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_h^0) & \boldsymbol{s}_t(\boldsymbol{\theta}_g^0)\boldsymbol{s}_t'(\boldsymbol{\rho}_A) \\ \boldsymbol{s}_t(\boldsymbol{\theta}_h^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_g^0) & \boldsymbol{s}_t(\boldsymbol{\theta}_h^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_h^0) & \boldsymbol{s}_t(\boldsymbol{\theta}_h^0)\boldsymbol{s}_t'(\boldsymbol{\rho}_A) \\ \boldsymbol{s}_t(\boldsymbol{\rho}_A)\boldsymbol{s}_t'(\boldsymbol{\theta}_g^0) & \boldsymbol{s}_t(\boldsymbol{\rho}_A)\boldsymbol{s}_t'(\boldsymbol{\theta}_h^0) & \boldsymbol{s}_t(\boldsymbol{\rho}_A)\boldsymbol{s}_t'(\boldsymbol{\rho}_A) \end{bmatrix}$$

The (i,j) sub-block of $\boldsymbol{B}_{11t} = \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_g^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_g^0)$, $i \neq j$, equals

$$\begin{aligned}[\boldsymbol{B}_{11t}]_{ij} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_{gi}^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_{gj}^0) \\ &= \frac{1}{4}(\frac{1}{g_{it}^0}\frac{\partial g_{it}^0}{\partial\boldsymbol{\theta}_{gi}} + \frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{gi}})(\frac{1}{g_{jt}^0}\frac{\partial g_{jt}^0}{\partial\boldsymbol{\theta}_{gj}'} + \frac{1}{h_{jt}^0}\frac{\partial h_{jt}^0}{\partial\boldsymbol{\theta}_{gj}'})\boldsymbol{e}_i'\boldsymbol{P}_{(A0)}^{-1}\boldsymbol{e}_j\boldsymbol{e}_i'\boldsymbol{P}_{(A0)}\boldsymbol{e}_j.\end{aligned}$$

When $i = j$,

$$\begin{aligned}[\boldsymbol{B}_{11t}]_{ii} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_{gi}^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_{gi}^0) \\ &= \frac{1}{4}(\frac{1}{g_{it}^0}\frac{\partial g_{it}^0}{\partial\boldsymbol{\theta}_{gi}} + \frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{gi}})(\frac{1}{g_{it}^0}\frac{\partial g_{it}^0}{\partial\boldsymbol{\theta}_{gi}'} + \frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{gi}'})(1 + \boldsymbol{e}_i'\boldsymbol{P}_{(A0)}^{-1}\boldsymbol{e}_i).\end{aligned}$$

The (i,j) sub-block of $\boldsymbol{B}_{22t} = \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_h^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_h^0)$, $i \neq j$, equals

$$\begin{aligned}[\boldsymbol{B}_{22t}]_{ij} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_{hi}^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_{hj}^0) \\ &= \frac{1}{4}(\frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{hi}})(\frac{1}{h_{jt}^0}\frac{\partial h_{jt}^0}{\partial\boldsymbol{\theta}_{hj}'})\boldsymbol{e}_i'\boldsymbol{P}_{(A0)}^{-1}\boldsymbol{e}_j\boldsymbol{e}_i'\boldsymbol{P}_{(A0)}\boldsymbol{e}_j.\end{aligned}$$

When $i = j$,

$$\begin{aligned}[\boldsymbol{B}_{22t}]_{ii} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_{hi}^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_{hi}^0) \\ &= \frac{1}{4}(\frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{hi}})(\frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{hi}'})(1 + \boldsymbol{e}_i'\boldsymbol{P}_{(A0)}^{-1}\boldsymbol{e}_i).\end{aligned}$$

The (i,j) sub-block of $\boldsymbol{B}_{12t} = \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_g^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_h^0)$, $i \neq j$, equals

$$\begin{aligned}[\boldsymbol{B}_{12t}]_{ij} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_{gi}^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_{hj}^0) \\ &= \frac{1}{4}(\frac{1}{g_{it}^0}\frac{\partial g_{it}^0}{\partial\boldsymbol{\theta}_{gi}} + \frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{gi}})(\frac{1}{h_{jt}^0}\frac{\partial h_{jt}^0}{\partial\boldsymbol{\theta}_{hj}'})\boldsymbol{e}_i'\boldsymbol{P}_{(A0)}^{-1}\boldsymbol{e}_j\boldsymbol{e}_i'\boldsymbol{P}_{(A0)}\boldsymbol{e}_j.\end{aligned}$$

When $i = j$,

$$\begin{aligned}[\boldsymbol{B}_{12t}]_{ii} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_{gi}^0)\boldsymbol{s}_t'(\boldsymbol{\theta}_{hi}^0) \\ &= \frac{1}{4}(\frac{1}{g_{it}^0}\frac{\partial g_{it}^0}{\partial\boldsymbol{\theta}_{gi}} + \frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{gi}})(\frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial\boldsymbol{\theta}_{hi}'})(1 + \boldsymbol{e}_i'\boldsymbol{P}_{(A0)}^{-1}\boldsymbol{e}_i).\end{aligned}$$

Furthermore, the (i,j) *sub-block of* $\mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_g^0)s_t'(\boldsymbol{\rho}_A)$ *equals*

$$\begin{aligned}[\boldsymbol{B}_{13t}]_{ij} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_{gi}^0)\boldsymbol{s}_t'(\boldsymbol{\rho}_{Aj}) \\ &= \frac{1}{4}(t/T)^j(\frac{1}{g_{it}^0}\frac{\partial g_{it}^0}{\partial \boldsymbol{\theta}_{gi}} + \frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial \boldsymbol{\theta}_{gi}})\{(\boldsymbol{e}_i \otimes \boldsymbol{e}_i)'(\boldsymbol{P}_{(A0)}^{-1} \otimes \boldsymbol{I}_N) + (\boldsymbol{e}_i \otimes \boldsymbol{e}_i)'(\boldsymbol{I}_N \otimes \boldsymbol{P}_{(A0)}^{-1})\}\boldsymbol{U}\end{aligned}$$

$i = 1, \ldots, N$; $j = 0, 1, 2$. *The* (i,j) *sub-block of* $\mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_h^0)s_t'(\boldsymbol{\rho}_A)$ *equals*

$$\begin{aligned}[\boldsymbol{B}_{23t}]_{ij} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\theta}_{hi}^0)\boldsymbol{s}_t'(\boldsymbol{\rho}_{Aj}) \\ &= \frac{1}{4}(t/T)^j(\frac{1}{h_{it}^0}\frac{\partial h_{it}^0}{\partial \boldsymbol{\theta}_{hi}})\{(\boldsymbol{e}_i \otimes \boldsymbol{e}_i)'(\boldsymbol{P}_{(A0)}^{-1} \otimes \boldsymbol{I}_N) + (\boldsymbol{e}_i \otimes \boldsymbol{e}_i)'(\boldsymbol{I}_N \otimes \boldsymbol{P}_{(A0)}^{-1})\}\boldsymbol{U}\end{aligned}$$

$i = 1, \ldots, N$; $j = 0, 1, 2$. *Finally, the* (i,j) *sub-block of the last block is equal to*

$$\begin{aligned}[\boldsymbol{B}_{33t}]_{ij} &= \mathsf{E}\boldsymbol{s}_t(\boldsymbol{\rho}_{Ai})\boldsymbol{s}_t'(\boldsymbol{\rho}_{Aj}) \\ &= \frac{1}{4}(t/T)^{i+j}\boldsymbol{U}'\boldsymbol{M}_A\boldsymbol{U}\end{aligned}$$

$i, j = 0, 1, 2$, *where*

$$\boldsymbol{M}_A = \boldsymbol{P}_{(A0)}^{-1} \otimes \boldsymbol{P}_{(A0)}^{-1} + (\boldsymbol{P}_{(A0)}^{-1} \otimes \boldsymbol{I}_N)\boldsymbol{K}(\boldsymbol{P}_{(A0)}^{-1} \otimes \boldsymbol{I}_N).$$

Proof. See the appendix of Silvennoinen and Teräsvirta (2005), or Silvennoinen and Teräsvirta (2021).

In order to define the test statistic, let $\boldsymbol{B}_{13\cdot j}$ be the (i,j) blocks of $\boldsymbol{B}_{13}$ where $i \in \{1, \ldots, N\}$, that is,

$$\boldsymbol{B}_{13\cdot j} = \left[[\boldsymbol{B}_{13}']_{j1}, \ldots, [\boldsymbol{B}_{13}']_{jN}\right], \quad j = 0, 1, 2$$

and define $\boldsymbol{B}_{23\cdot j}$ similarly. Partition the matrix $\boldsymbol{B}$ as follows:

$$\tilde{\boldsymbol{B}}_{11} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \boldsymbol{B}_{13\cdot 0} \\ \boldsymbol{B}_{12}' & \boldsymbol{B}_{22} & \boldsymbol{B}_{23\cdot 0} \\ \boldsymbol{B}_{13\cdot 0}' & \boldsymbol{B}_{23\cdot 0}' & [\boldsymbol{B}_{33}]_{00} \end{bmatrix}$$

$$\tilde{\boldsymbol{B}}_{12} = \begin{bmatrix} \boldsymbol{B}_{13\cdot 1} & \boldsymbol{B}_{13\cdot 2} \\ [\boldsymbol{B}_{33}]_{01} & [\boldsymbol{B}_{33}]_{02} \end{bmatrix}$$

and

$$\tilde{\boldsymbol{B}}_{33} = \begin{bmatrix} [\boldsymbol{B}_{33}]_{11} & [\boldsymbol{B}_{33}]_{12} \\ [\boldsymbol{B}_{33}']_{12} & [\boldsymbol{B}_{33}]_{22} \end{bmatrix}.$$

Next, define

$$\widehat{\boldsymbol{x}}_{jt} = -\frac{1}{2}(\frac{t}{T})^j\boldsymbol{U}'\{\text{vec}(\boldsymbol{P}_{(A0)}^{-1}) - (\widehat{\boldsymbol{P}}_{(A0)}^{-1} \otimes \widehat{\boldsymbol{P}}_{(A0)}^{-1})\text{vec}(\widehat{\boldsymbol{z}}_t\widehat{\boldsymbol{z}}_t')\}$$

$j = 1, 2$, where $\widehat{\boldsymbol{z}}_t$ and $\widehat{\boldsymbol{P}}_{(A0)}$ equal $\boldsymbol{z}_t$ and $\boldsymbol{P}_{(A0)}$ estimated under H_0, respectively. The test statistic

$$LM_T = T(\frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{x}}'_{1t}, \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{x}}'_{2t})\{\tilde{\boldsymbol{B}}_{22} - \tilde{\boldsymbol{B}}'_{12}(\tilde{\boldsymbol{B}}^0_{11})^{-1}\tilde{\boldsymbol{B}}_{12}\}^{-1}(\frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{x}}'_{1t}, \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{x}}'_{2t})' \quad \text{(A.5)}$$

has an asymptotic χ^2-distribution with $N(N-1)$ degrees of freedom when H_0 holds. To make the test statistic operational, the sub-blocks of the information matrix in (A.5) have to be replaced by consistent plug-in estimators.

A.4 Test for an additional transition in the correlations

The test statistic for an additional transition is constructed in the same way as in Section A.3, and the blocks related to the volatility components are identical. However, all blocks related to the correlation need modifications to include the parameters governing the time-varying correlation that exists under the null. This includes both the parameters in the correlation matrices under the null, and their corresponding transition parameters.

Let us define $\boldsymbol{x}_{hit} = h_{it}^{-1}\frac{\partial h_{it}}{\partial \boldsymbol{\theta}_{hi}}$, $\boldsymbol{x}_{git} = g_{it}^{-1}\frac{\partial g_{it}}{\partial \boldsymbol{\theta}_{gi}} + h_{it}^{-1}\frac{\partial h_{it}}{\partial \boldsymbol{\theta}_{gi}}$. Let us also partition the linearised correlation model as $\boldsymbol{P}_{At} = \boldsymbol{P}_{At0} + t/T\boldsymbol{P}_{(A1)} + (t/T)^2\boldsymbol{P}_{(A2)}$, where $\boldsymbol{P}_{At0}$ contains the time-varying correlation model under the null. When testing L transitions against $L+1$ transitions, $\boldsymbol{P}_{At0}$ contains $L+1$ correlation matrices $\boldsymbol{P}_{(1)}, \ldots, \boldsymbol{P}_{(L+1)}$ and L transition functions $G_l(t/T, \gamma_l, c_l)$ (here we assume $K_l = 1$ for simplicity), $l = 1, \ldots, L$. The information matrix is approximated by its consistent estimator

$$\hat{\boldsymbol{B}} = T^{-1}\sum_{t=1}^{T}\mathsf{E}_{t-1}[\boldsymbol{s}_t(\boldsymbol{\theta}^0)\boldsymbol{s}_t(\boldsymbol{\theta}^0)']$$

where $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_g, \boldsymbol{\theta}_h, \boldsymbol{\theta}_G, \boldsymbol{\theta}_{\boldsymbol{\rho}_{A0}}, \boldsymbol{\theta}_{\boldsymbol{\rho}_{A1}})'$, where $\boldsymbol{\theta}_G$ contains the transition parameters from the L transitions that are present under the null, $\boldsymbol{\theta}_{\boldsymbol{\rho}_{A0}} = (\boldsymbol{\rho}_{(1)}, \ldots, \boldsymbol{\rho}_{(L+1)})'$ and $\boldsymbol{\theta}_{\boldsymbol{\rho}_{A1}} = (\boldsymbol{\rho}_{(A1)}, \boldsymbol{\rho}_{(A2)})'$. From here on, the expressions are evaluated at the true parameter values under the null (we omit the additional superscripts of 0 to keep the notation simple).

With this notation, the (i,j) sub-block of $\hat{\boldsymbol{B}}_{11}$, $i \neq j$, equals

$$[\hat{\boldsymbol{B}}_{11}]_{ij} = \frac{1}{4T}\sum_{t=1}^{T}\boldsymbol{x}_{git}\boldsymbol{x}'_{gjt}\boldsymbol{e}'_i\boldsymbol{P}^{-1}_{(At0)}\boldsymbol{e}_j\boldsymbol{e}'_i\boldsymbol{P}_{(At0)}\boldsymbol{e}_j.$$

When $i = j$,

$$[\hat{\boldsymbol{B}}_{11}]_{ii} = \frac{1}{4T}\sum_{t=1}^{T}\boldsymbol{x}_{git}\boldsymbol{x}'_{git}(1 + \boldsymbol{e}'_i\boldsymbol{P}^{-1}_{(At0)}\boldsymbol{e}_i).$$

Similarly, the (i,j) sub-block of $\hat{\boldsymbol{B}}_{22}$, $i \neq j$, is equal to

$$[\hat{\boldsymbol{B}}_{22}]_{ij} = \frac{1}{4T}\sum_{t=1}^{T}\boldsymbol{x}_{hit}\boldsymbol{x}'_{hjt}\boldsymbol{e}'_i\boldsymbol{P}^{-1}_{(At0)}\boldsymbol{e}_j\boldsymbol{e}'_i\boldsymbol{P}_{(At0)}\boldsymbol{e}_j.$$

When $i = j$,

$$[\hat{\boldsymbol{B}}_{22}]_{ii} = \frac{1}{4T}\sum_{t=1}^{T}\boldsymbol{x}_{hit}\boldsymbol{x}'_{hit}(1 + \boldsymbol{e}'_i\boldsymbol{P}^{-1}_{(At0)}\boldsymbol{e}_i).$$

The (i,j) sub-block of $\hat{\boldsymbol{B}}_{12}$, $i \neq j$, equals

$$[\hat{\boldsymbol{B}}_{12}]_{ij} = \frac{1}{4T} \sum_{t=1}^{T} \boldsymbol{x}_{git} \boldsymbol{x}'_{hjt} \boldsymbol{e}'_i \boldsymbol{P}^{-1}_{(At0)} \boldsymbol{e}_j \boldsymbol{e}'_i \boldsymbol{P}_{(At0)} \boldsymbol{e}_j.$$

When $i = j$,

$$[\hat{\boldsymbol{B}}_{12}]_{ii} = \frac{1}{4T} \sum_{t=1}^{T} \boldsymbol{x}_{git} \boldsymbol{x}'_{hit} (1 + \boldsymbol{e}'_i \boldsymbol{P}^{-1}_{(At0)} \boldsymbol{e}_i).$$

The next blocks deal with the transition parameters. Define $\boldsymbol{x}_{Gt} = \frac{\partial \text{vec} \boldsymbol{P}'_{At}}{\partial \boldsymbol{\theta}_G}$. The lth block of $\frac{\partial \text{vec} \boldsymbol{P}'_{At}}{\partial \boldsymbol{\theta}_G}$ is

$$\left(\prod_{i=l+1}^{L} (1 - G_i) \right) \frac{\partial G_i}{\partial \boldsymbol{\theta}_{G_i}} \text{vec}(\boldsymbol{P}_{(l+1)} - \boldsymbol{P}_t^{(l-1)})'$$

using the recursion in (8). The block $\hat{\boldsymbol{B}}_{33}$ is equal to

$$\hat{\boldsymbol{B}}_{33} = \frac{1}{4T} \sum_{t=1}^{T} \boldsymbol{x}_{Gt} \boldsymbol{M}_A \boldsymbol{x}'_{Gt}.$$

The ith sub-block of $\hat{\boldsymbol{B}}_{13}$ equals

$$[\hat{\boldsymbol{B}}_{13}]_i = \frac{1}{4T} \sum_{t=1}^{T} \boldsymbol{x}_{git} (\boldsymbol{e}'_i \boldsymbol{P}^{-1}_{(At0)} \otimes \boldsymbol{e}'_i + \boldsymbol{e}'_i \otimes \boldsymbol{e}'_i \boldsymbol{P}^{-1}_{(At0)}) x'_{Gt}.$$

and the ith sub-block of $\hat{\boldsymbol{B}}_{23}$ equals

$$[\hat{\boldsymbol{B}}_{23}]_i = \frac{1}{4T} \sum_{t=1}^{T} \boldsymbol{x}_{hit} (\boldsymbol{e}'_i \boldsymbol{P}^{-1}_{(At0)} \otimes \boldsymbol{e}'_i + \boldsymbol{e}'_i \otimes \boldsymbol{e}'_i \boldsymbol{P}^{-1}_{(At0)}) x'_{Gt}.$$

Next, we will consider the blocks related to the correlations. The matrix $\partial \text{vec}(\boldsymbol{P}_{At})' / \partial \boldsymbol{\theta}_{\rho_{A0}}$ equals

$$\frac{\partial \text{vec}(\boldsymbol{P}_{At})'}{\partial \boldsymbol{\theta}_{\rho_{A0}}} = \boldsymbol{v} \otimes \boldsymbol{U}' = \begin{bmatrix} \prod_{l=1}^{L} (1 - G_{lt}) \\ G_{1t} \prod_{l=2}^{L} (1 - G_{lt}) \\ G_{2t} \prod_{l=3}^{L} (1 - G_{lt}) \\ \cdots \\ G_{l-1,t} (1 - G_{Lt}) \\ G_{Lt} \end{bmatrix} \otimes \boldsymbol{U}'.$$

and the matrix $\partial \text{vec}(\boldsymbol{P}_{At})' / \partial \boldsymbol{\theta}_{\rho_{A1}}$ equals

$$\frac{\partial \text{vec}(\boldsymbol{P}_{At})'}{\partial \boldsymbol{\rho}_{A1}} = \begin{bmatrix} t/T \\ (t/T)^2 \end{bmatrix} \otimes \boldsymbol{U}'$$

The block $\hat{\boldsymbol{B}}_{44}$ is equal to

$$\hat{\boldsymbol{B}}_{44} = \frac{1}{4T} \sum_{t=1}^{T} \boldsymbol{v}_t \boldsymbol{v}'_t \otimes \boldsymbol{U}' \boldsymbol{M}_A \boldsymbol{U}.$$

and (i,j) sub-block of $\hat{\boldsymbol{B}}_{55}$ is equal to

$$[\hat{\boldsymbol{B}}_{55}]_{ij} = \frac{1}{4T}\sum_{t=1}^{T}(t/T)^{i+j}\boldsymbol{U}'\boldsymbol{M}_A\boldsymbol{U}.$$

for $i,j=1,2$. The ith sub-block of $\hat{\boldsymbol{B}}_{14}$ is equal to

$$[\hat{\boldsymbol{B}}_{14}]_i = \frac{1}{4T}\sum_{t=1}^{T}\boldsymbol{x}_{git}(\boldsymbol{e}_i'\boldsymbol{P}_{(At0)}^{-1}\otimes\boldsymbol{e}_i' + \boldsymbol{e}_i'\otimes\boldsymbol{e}_i'\boldsymbol{P}_{(At0)}^{-1})(\boldsymbol{v}_t'\otimes\boldsymbol{U})$$

and the (i,j) sub-block of $\hat{\boldsymbol{B}}_{15}$ is equal to

$$[\hat{\boldsymbol{B}}_{15}]_{ij} = \frac{1}{4T}\sum_{t=1}^{T}(t/T)^{j}\boldsymbol{x}_{git}(\boldsymbol{e}_i'\boldsymbol{P}_{(At0)}^{-1}\otimes\boldsymbol{e}_i' + \boldsymbol{e}_i'\otimes\boldsymbol{e}_i'\boldsymbol{P}_{(At0)}^{-1})\boldsymbol{U})$$

$j=1,2$. The corresponding sub-blocks of $\hat{\boldsymbol{B}}_{24}$ and $\hat{\boldsymbol{B}}_{25}$ are

$$[\hat{\boldsymbol{B}}_{24}]_i = \frac{1}{4T}\sum_{t=1}^{T}\boldsymbol{x}_{hit}(\boldsymbol{e}_i'\boldsymbol{P}_{(At0)}^{-1}\otimes\boldsymbol{e}_i' + \boldsymbol{e}_i'\otimes\boldsymbol{e}_i'\boldsymbol{P}_{(At0)}^{-1})(\boldsymbol{v}_t'\otimes\boldsymbol{U})$$

and the (i,j) sub-block of $\hat{\boldsymbol{B}}_{15}$ is equal to

$$[\hat{\boldsymbol{B}}_{25}]_{ij} = \frac{1}{4T}\sum_{t=1}^{T}(t/T)^{j}\boldsymbol{x}_{hit}(\boldsymbol{e}_i'\boldsymbol{P}_{(At0)}^{-1}\otimes\boldsymbol{e}_i' + \boldsymbol{e}_i'\otimes\boldsymbol{e}_i'\boldsymbol{P}_{(At0)}^{-1})\boldsymbol{U})$$

$j=1,2$. The block $\hat{\boldsymbol{B}}_{34}$ equals

$$\hat{\boldsymbol{B}}_{34} = \frac{1}{4T}\sum_{t=1}^{T}\boldsymbol{x}_{Gt}\boldsymbol{M}_A(\boldsymbol{v}'\otimes\boldsymbol{U})$$

The ith sub-block of $\hat{\boldsymbol{B}}_{35}$ equals

$$[\hat{\boldsymbol{B}}_{35}]_i = \frac{1}{4T}\sum_{t=1}^{T}(t/T)^{i}\boldsymbol{x}_{Gt}\boldsymbol{M}_A\boldsymbol{U}$$

$i=1,2$ Finally, the The ith sub-block of $\hat{\boldsymbol{B}}_{45}$ is equal to

$$[\hat{\boldsymbol{B}}_{45}]_i = \frac{1}{4T}\sum_{t=1}^{T}(t/T)^{i}(\boldsymbol{v}\otimes\boldsymbol{U}')\boldsymbol{M}_A\boldsymbol{U}$$

Next, define

$$\widehat{\boldsymbol{x}}_{jt} = -\frac{1}{2}(\frac{t}{T})^{j}\boldsymbol{U}'\{\text{vec}(\boldsymbol{P}_{(A0t)}^{-1}) - (\hat{\boldsymbol{P}}_{(A0t)}^{-1}\otimes\hat{\boldsymbol{P}}_{(A0t)}^{-1})\text{vec}(\hat{\boldsymbol{z}}_t\hat{\boldsymbol{z}}_t')\}$$

$j=1,2$, where $\hat{\boldsymbol{z}}_t$ and $\hat{\boldsymbol{P}}_{(A0t)}$ equal $\boldsymbol{z}_t$ and $\boldsymbol{P}_{(A0t)}$ estimated under H_0, respectively. The test statistic

$$LM_T = T(\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{x}}'_{1t}, \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{x}}'_{2t})[\hat{\boldsymbol{B}}^{-1}]_{SW}(\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{x}}'_{1t}, \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{x}}'_{2t})'$$

where $[\hat{\boldsymbol{B}}^{-1}]_{SW}$ is the $N(N-1) \times N(N-1)$ block in the south-west corner of the inverse of $\hat{\boldsymbol{B}}$. Because the matrix $\hat{\boldsymbol{B}}$ can have a large dimension, its inverse could be obtained by using block inversion methods, perhaps applying them recursively. The test statistic has an asymptotic χ^2-distribution with $N(N-1)$ degrees of freedom when H_0 holds.

B Simulations of test statistics

B.1 Tests of GARCH equations

The test for slow moving baseline volatility has a statistic whose distribution is sensitive to the high frequency, GARCH, volatility. For this reason, one cannot use the asymptotic distribution, rather the distribution must be generated via simulation. Further, Silvennoinen and Teräsvirta (2016) showed that the size of the test is distorted if the GARCH parameterisation deviates from the true one. For this reason a few alternative approaches to estimate the GARCH parameters, and especially the persistence, are investigated. It should be noted that estimating GARCH without taking the nonstationarity into account will yield overestimated persistence, thereby impacting the test statistic distribution and thus rendering the test outcomes unreliable. These estimates are given in Table 3.

The baseline volatility may be very different in different data series. Therefore one should not ignore visual inspection of the returns, nor rely on general rules of thumb. If there are sufficiently long sections of data where the general level of volatility remains constant, it is advisable to estimate the GARCH parameters over such subsample. In the case here, there are a couple of relatively constant volatility sections, for example from November 2003 until October 2007. The parameter estimates for that calm subperiod are in Table 3. Comparison with the estimates from the entire period GARCH model, it is clear that the neglected nonstationarity has biased the estimates resulting in high persistence and kurtosis. Because the data set has a sufficiently long span of GARCH type clustering without (visually) significant movement in the general baseline level, more reliable estimates are obtained by using that subsample only.

Another approach consists of estimating the GARCH equation over a rolling window such that the intercept is time-varying, targeting the unconditional volatility over each window, while the other parameters are assumed constant over the entire sample period, and estimated in the usual way. The choice of the window length should take into account the general recommendations regarding the sample size when attempting GARCH estimation. Too long a window will be impacted by the slowly changing baseline volatility level, whereas too short a window will yield very uncertain GARCH estimates. To investigate the properties of this approach, we ran a simulation experiment with a few different baseline volatilities. The window widths varied from 250 to 1000 observations. Figures 9–11 in Section G depict the distributions of the GARCH estimates and the derived persistence and kurtosis measures, as explained in He and Teräsvirta (1999), for a selection of baseline volatilities and window widths. Based on these experiments, we conclude that a window width of 400 observations yields sufficiently robust results for our application. The resulting GARCH estimates are reported in Table 3, and they are quite similar to the ones obtained for the aforementioned calm period. This can be interpreted as support for the rolling window method, especially in situations where visual inspection of data does not reveal a sufficiently long period of constant unconditional volatility.

Overall, it is clear that using simply the GARCH estimates from the entire sample to calibrate the test statistic distribution for the specification of the deterministic component of the volatility is not recommended. For comparison, Table 3 reports also the GARCH estimates from a TV-GARCH model where the TV specification has been completed. The estimated persistence is higher than the ones obtained from the calm period or rolling window variance targeting method, however, as discussed in Silvennoinen and Teräsvirta (2016), underestimation of persistence has less severe impact on the performance of the

TV specification test than overestimation does.

		$\tilde{\alpha}$	$\tilde{\beta}$	persistence	kurtosis
Rolling window 400	ANZ	0.090	0.836	0.926	3.38
	CBA	0.087	0.850	0.937	3.43
	NAB	0.095	0.817	0.912	3.36
	WBC	0.085	0.858	0.943	3.45
Calm period	ANZ	0.073	0.852	0.925	3.24
	CBA	0.081	0.842	0.923	3.29
	NAB	0.066	0.829	0.896	3.14
	WBC	0.091	0.806	0.897	3.28
Entire period GARCH only	ANZ	0.065	0.927	0.992	6.40
	CBA	0.089	0.890	0.979	4.83
	NAB	0.104	0.867	0.971	4.85
	WBC	0.075	0.911	0.986	5.08
Entire period TV-GARCH	ANZ	0.078	0.880	0.957	3.50
	CBA	0.091	0.860	0.950	3.61
	NAB	0.107	0.825	0.931	3.62
	WBC	0.084	0.878	0.962	3.70

Table 3: Specification stage for the deterministic component in volatilities of each of the four banks. $\tilde{\alpha}$ and $\tilde{\beta}$ are the initial estimates used for calibrating the test statistic distribution. The rolling window method allows the GARCH intercept to adjust to target the unconditional variance in a window of size 400. The "calm period" selects the continuous period from Nov-2003 to Oct-2007 which has very little visible variation in the baseline volatility. For comparison, the GARCH estimates from the entire sample period are reported, along with the final estimates from the TV-GARCH model.

B.2 Evaluation tests of GARCH equations

The fact that the evaluation tests discussed in Section A.2 are applied to the pre-filtered data $\hat{\boldsymbol{P}}_t^{-1/2}\boldsymbol{\varepsilon}_t$ is known to potentially alter the distribution of the test statistic. In this Section we present simulation results that show the size of the tests remains practically unchanged, rendering the tests applicable in the proposed way.

The simulation uses 2000 observations on a bivariate TVGARCH model parameterised as $h_t = 0.10 + 0.05\varepsilon_{t-1}^2/g_{t-1} + 0.85h_{t-1}$, $g_t = 1 + 3(1 + \exp\{-e^3(t/T - 0.5)\})^{-1}$. These are coupled with a CCC model with $\rho = 0.5$, and then with an STCC model parameterised as $\rho_{(1)} = 0.3$, $\rho_{(2)} = 0.7$, $G_t = (1 + \exp\{-e^{2.5}(t/T - 0.5)\})^{-1}$. The noise terms are iid standard normal. Two estimation procedures were used, a two-step and a multi-step one.

1st step The individual TVGARCH models are estimated, assuming the series are uncorrelated.

2nd step Estimate the correlation model conditional on the volatility model estimates from the previous step. Then, estimate the TVGARCH models conditional on the correlation estimates.

The misspecification tests are then calculated using the TVGARCH estimates from the 2nd step, and the data is pre-filtered with the correlation estimates from the 2nd step. The multi-step continues repeating the procedure of the 2nd step, until no further improvements are achieved.

		standard			robust		
		10%	5%	1%	10%	5%	1%
CCC two-step	MS1	0.146	0.085	0.020	0.132	0.074	0.016
	MS2 - a	0.122	0.064	0.012	0.101	0.048	0.013
	MS2 - b	0.143	0.080	0.017	0.108	0.051	0.008
	MS3	0.125	0.061	0.010	0.104	0.054	0.010
STCC two-step	MS1	0.134	0.074	0.023	0.121	0.055	0.015
	MS2 - a	0.123	0.059	0.015	0.101	0.045	0.013
	MS2 - b	0.122	0.062	0.019	0.087	0.044	0.010
	MS3	0.115	0.058	0.015	0.100	0.050	0.011
CCC multi-step	MS1	0.145	0.083	0.022	0.133	0.073	0.014
	MS2 - a	0.116	0.062	0.015	0.097	0.052	0.009
	MS2 - b	0.133	0.069	0.018	0.100	0.046	0.010
	MS3	0.120	0.062	0.016	0.107	0.060	0.014
STCC multi-step	MS1	0.147	0.084	0.023	0.135	0.068	0.012
	MS2 - a	0.130	0.059	0.011	0.103	0.046	0.006
	MS2 - b	0.120	0.067	0.016	0.090	0.039	0.005
	MS3	0.112	0.055	0.012	0.104	0.047	0.009

Table 4: Size simulation for the three types of misspecification tests in Amado and Terasvirta (2017). 2000 replications. $T = 2000$, $N = 2$.
MS1: g_t additively misspecified, alternative linearised with first order term only.
MS2 - a: GARCH(1,1) vs GARCH(1,2)
MS2 - b: GARCH(1,1) vs GARCH(2,1)
MS3: test for remaining ARCH, lag 1.

From Table 4 it is evident that the standard form of the tests is slightly oversized. The robust version of the tests on the other hand seems to behave well, and there is no need for any adjustments of the test statistics or their distributions. Therefore the procedure of removing the correlations between the series prior to applying the evaluation tests can be recommended.

B.3 Tests of correlations

The simulation experiment investigates the size of the test in an environment where the multivariate model is correctly specified. The number of data series considered in the system is $N = 2, 5, 10, 20$. The length varies from $T = 25$ for the bivariate systems, which is relevant for time series systems in macro applications, up to $T = 1000$, which in turn is considered to be a fairly small sample size for high frequency returns data. The length of the time series places a constraint on the dimension of the model, that is, the parametric alternative is only feasible if the number of parameters remains comfortably below the amount of available data points. We simulated the test by both assuming that $\boldsymbol{D}_t \equiv \boldsymbol{I}_N$ and that there is conditional heteroskedasticity in the model: $\boldsymbol{D}_t \neq \boldsymbol{I}_N$.

When $\boldsymbol{D}_t \equiv \boldsymbol{I}_N$, it turns out that the results are fairly independent of the structure of the correlations. We used both equicorrelation and Toeplitz matrices in our simulations, and the results remained the same. Table 5 in Section F contains results of a simulation in which $\boldsymbol{D}_t \equiv \boldsymbol{I}_N$, and the $N \times N$ correlation matrix $\boldsymbol{P} = [\rho_{ij}]$ is an equicorrelated one with weak ($\rho = 1/3$) and moderately strong ($\rho = 2/3$) correlation. The Table also reports the results from using a Toeplitz correlation matrix such that $[\rho_{ij}] = \rho^{|i-j|}$, $i, j = 1, ..., N$, with $\rho = 0.5$ representing moderate to weak correlation, and $\rho = 0.9$ representing strong to moderate correlation. It is seen that the empirical size of the test is rather close to the nominal one already when $N = 2$ and $T = 100$. The size holds up across the various correlation patterns.

We next turn to the case $\boldsymbol{D}_t \neq \boldsymbol{I}_N$. Tables 6 and 7 in Section F contain results of size simulations where the sensitivity of the test is examined against combinations for the GARCH persistence and kurtosis, and a selection of strengths of correlations (the equicorrelated and Toepliz ones described above). The test is generally well-sized. However, an interesting aspect is that there is slight oversizing when kurtosis decreases (which means shifting the relative weight from α to β in the GARCH equation, while keeping persistence constant). A change in persistence does not seem to affect the size of the test. As the dimension of the system increases, the test does not perform equally well. Increasing sample size does not seem to be able to counteract this (the simulations use $T = 500, 1000, 2000$).

As an interesting additional simulation a misspecified case is considered. When $\boldsymbol{D}_t \neq \boldsymbol{I}_N$ but the conditional heteroskedasticity is neglected, the test is heavily oversized, as expected (results not reported here). The obvious conclusion is that constancy of correlations can only be tested after specifying and estimating both $\boldsymbol{D}_t$ and $\boldsymbol{S}_t$.

C Proofs

Proof of Lemma 1. The 'sample' information matrix with T observations equals

$$\frac{1}{4T}\sum_{t=1}^{T}\boldsymbol{B}_t = \frac{1}{4T}\sum_{t=1}^{T}\begin{bmatrix} \boldsymbol{B}_{11t} & \boldsymbol{B}_{12t} \\ \boldsymbol{B}_{21t} & \boldsymbol{B}_{22t} \end{bmatrix}.$$

Consider the $(1,2)$ element of $\boldsymbol{B}_{11(T)} = (1/T)\sum_{t=1}^{T}\boldsymbol{B}_{11t}$:

$$[\boldsymbol{B}_{11(T)}]_{12} = (1/T)\sum_{t=1}^{T}(g^0(t/T))^{-2}G_1^0(t/T)$$

which is an average of T values of the logistic cumulative distribution function. Let $[Tr] = t$ be the integer closest to t. Then

$$\begin{aligned}(1/T)\sum_{t=1}^{T}(g^0(t/T))^{-2}G_1^0(t/T) &= \sum_{t=1}^{T}\int_{t/T}^{(t+1)/T}(g^{*0}([Tr]/T)^{-2}G_1^0([Tr]/T)\mathrm{d}r \\ &= \int_{1/T}^{(T+1)/T}(g^{*0}([Tr]/T))^{-2}G_1^0([Tr]/T)\mathrm{d}r \\ &\to \int_0^1(g^{*0}(r))^{-2}G_1^0(r)\mathrm{d}r\end{aligned}$$

as $T \to \infty$. The other elements of $\boldsymbol{B}_{11} = \lim_{T\to\infty}\boldsymbol{B}_{11(T)}$, are derived in a similar fashion. In matrix form,

$$\boldsymbol{B}_{11} = \frac{1}{2}\int_0^1(g^{*0}(r))^{-2}\boldsymbol{g}_1^0(r)\boldsymbol{g}_1^0(r)'\mathrm{d}r$$

The blocks $\boldsymbol{B}_{12}$ and $\boldsymbol{B}_{22}$ are obtained similarly.

D Details of maximisation by parts

This appendix describing the outlines of the estimation algorithm derives from Silvennoinen and Teräsvirta (2021). Estimation proceeds as follows.

1. Assume $\ln h_{it}(\boldsymbol{\theta}_{hi}, \boldsymbol{\theta}_{gi}) = 0$, $i = 1, \ldots, N$, and estimate parameters $\boldsymbol{\theta}_g = (\boldsymbol{\theta}_{g1}, \ldots, \boldsymbol{\theta}_{gN})'$, $i = 1, \ldots, N$, equation by equation, assuming $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho) = \boldsymbol{I}_N$. Denote the estimate $\boldsymbol{S}_t(\widehat{\boldsymbol{\theta}}_g^{(1,1)})$. This means that the deterministic components $g_i(t/T, \boldsymbol{\theta}_{gi})$ have been estimated once, including the intercept δ_{i0} in (2).

2. Estimate $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho)$ given $\boldsymbol{\theta}_g = \widehat{\boldsymbol{\theta}}_g^{(1,1)}$. This requires a separate iteration because $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho)$ is nonlinear in parameters, see (5) and (6). Denote the estimate $\boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(1,1)})$.

3. Re-estimate $\boldsymbol{S}_t(\boldsymbol{\theta}_g)$ assuming $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho) = \boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(1,1)})$. This yields $\boldsymbol{S}_t(\widehat{\boldsymbol{\theta}}_g^{(1,2)})$. Then re-estimate $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho)$ given $\boldsymbol{\theta}_g = \widehat{\boldsymbol{\theta}}_g^{(1,2)}$. Iterate until convergence. Let the result after R_1 iterations be $\boldsymbol{S}_t(\boldsymbol{\theta}_g) = \boldsymbol{S}_t(\widehat{\boldsymbol{\theta}}_g^{(1,R_1)})$ and $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho) = \boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(1,R_1)})$. The resulting estimates are maximum likelihood ones under the assumption $\boldsymbol{D}_t(\boldsymbol{\theta}_h, \boldsymbol{\theta}_g) = \boldsymbol{I}_N$.

4. Estimate $\boldsymbol{\theta}_h$ from $\boldsymbol{D}_t(\boldsymbol{\theta}_h, \widehat{\boldsymbol{\theta}}_g^{(1,R_1)})$ using $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho) = \boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(1,R_1)})$. This is a standard multivariate conditional correlation GARCH estimation step as in Bollerslev (1990), because $\boldsymbol{S}_t(\widehat{\boldsymbol{\theta}}_g^{(1,R_1)})$ is fixed and does not affect the maximum, and $\boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(1,R_1)})$ is known. In total, steps 1–4 form the first iteration of the maximisation algorithm. Denote the estimate $\widehat{\boldsymbol{\theta}}_h^{(1)}$.

5. Estimate $\boldsymbol{\theta}_g$ from $\boldsymbol{S}_t(\boldsymbol{\theta}_g)$ keeping $\boldsymbol{D}_t(\widehat{\boldsymbol{\theta}}_h^{(1)}, \widehat{\boldsymbol{\theta}}_g^{(1,R_1)})$ and $\boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(1,R_1)})$ fixed. This step is analogous to the first part of Step 3. The difference is that $\boldsymbol{D}_t(\widehat{\boldsymbol{\theta}}_h^{(1)}, \widehat{\boldsymbol{\theta}}_g^{(1,R_1)}) \neq \boldsymbol{I}_N$. Denote the estimator $\boldsymbol{S}_t(\widehat{\boldsymbol{\theta}}_g^{(2,1)})$.

6. Estimate $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho)$ given $\boldsymbol{\theta}_g = \widehat{\boldsymbol{\theta}}_g^{(2,1)}$ and $\boldsymbol{\theta}_h = \widehat{\boldsymbol{\theta}}_h^{(1)}$. Denote the estimator $\boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(2,1)})$. Iterate until convergence, R_2 iterations. The result: $\boldsymbol{S}_t(\boldsymbol{\theta}_g) = \boldsymbol{S}_t(\widehat{\boldsymbol{\theta}}_g^{(2,R_2)})$ and $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho) = \boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(2,R_2)})$.

7. Estimate $\boldsymbol{\theta}_h$ from $\boldsymbol{D}_t(\boldsymbol{\theta}_h, \widehat{\boldsymbol{\theta}}_g^{(2,R_2)})$ using $\boldsymbol{P}_t(\boldsymbol{\theta}_\rho) = \boldsymbol{P}_t(\widehat{\boldsymbol{\theta}}_\rho^{(2,R_2)})$ ($\boldsymbol{S}_t(\widehat{\boldsymbol{\theta}}_g^{(2,R_2)})$ is fixed). The result: $\boldsymbol{\theta}_h = \widehat{\boldsymbol{\theta}}_h^{(2)}$. This completes the second full iteration.

8. Repeat steps 5–7 and iterate until convergence.

For identification reasons, δ_{0i}, $i = 1, \ldots, N$, is frozen to $\delta_{i0} = \widehat{\delta}_{i0}^{(1,R_1)}$. This frees the intercepts in $\boldsymbol{\theta}_{hi}$. Any positive constant would do for δ_{i0}, but for numerical reasons the intercepts are fixed to the values they obtain after the first iteration when $\boldsymbol{\theta}_h$ is not yet estimated a single time.

In practice, in estimating the slope parameters in transition functions it may be useful to apply the transformation $\gamma_{ij} = \exp\{\eta_{ij}\}$, in which case γ_{ij} need not be restricted when η_{ij} is bounded away from $-\infty$. The motivation for this transformation is that estimating η_{ij} instead of γ_{ij} is numerically convenient in cases where γ_{ij} is large, see Goodwin, Holt and Prestemon (2011) or Silvennoinen and Teräsvirta (2016) for discussion. Another alternative, proposed by Chan and Theoharakis (2011), is to redefine the slope parameter as $\gamma_{ij} = 1/\eta_{ij}^2$ and estimate η_{ij}. The authors show that this also alleviates the convergence problems sometimes found when γ_{ij} is large. Ekner and Nejstgaard (2013) aim at the same effect by rescaling γ_{ij} to vary between zero and one.

E Estimated transition equations

This appendix contains the estimated deterministic components in the TV-GARCH equations (standard deviation estimates in parentheses). Note that the intercept is fixed after the first iteration, hence it does not have a standard deviation estimate.

ANZ

$$\begin{aligned}\widehat{g}_{1t} =& 2.28 - \underset{(0.059)}{1.234}(1+\exp\{-\underset{(1.223)}{5.715}(t/T-\underset{(0.003)}{0.404})\})^{-1} \\ &+ \underset{(1.518)}{12.316}(1+\exp\{-\underset{(0.392)}{5.875}(t/T-\underset{(0.002)}{0.571})\})^{-1} \\ &- \underset{(1.514)}{11.704}(1+\exp\{-\underset{(0.0.166)}{4.459}\ (t/T-\underset{(0.004)}{0.623})\})^{-1}.\end{aligned}$$

CBA

$$\begin{aligned}\widehat{g}_{2t} =& 1.35 - \underset{(0.054)}{0.525}(1+\exp\{-\underset{(2.545)}{5.638}(t/T-\underset{(0.007)}{0.407})\})^{-1} \\ &+ \underset{(1.871)}{9.257}(1+\exp\{-\underset{(0.374)}{5.117}(t/T-\underset{(0.004)}{0.574})\})^{-1} \\ &- \underset{(1.867)}{8.944}(1+\exp\{-\underset{(0.252)}{4.504}(t/T-\underset{(0.006)}{0.621})\})^{-1}.\end{aligned}$$

NAB

$$\begin{aligned}\widehat{g}_{3t} =& 1.07 + \underset{(1.273)}{3.843}(1+\exp\{-\underset{(0.130)}{2.518}(t/T-\underset{(0.034)}{0.303})\})^{-1} \\ &- \underset{(1.114)}{3.491}(1+\exp\{-\underset{(0.329)}{3.787}(t/T-\underset{(0.008)}{0.373})\})^{-1} \\ &+ \underset{(5.692)}{20.026}(1+\exp\{-\underset{(0.229)}{4.926}(t/T-\underset{(0.003)}{0.576})\})^{-1} \\ &- \underset{(5.676)}{20.039}(1+\exp\{-\underset{(0.123)}{4.183}(t/T-\underset{(0.006)}{0.609})\})^{-1}.\end{aligned}$$

WBC

$$\begin{aligned}\widehat{g}_{4t} =& 2.45 - \underset{(0.554)}{3.120}(1+\exp\{-\underset{(0.124)}{2.194}(t/T-\underset{(0.034)}{0.534})\})^{-1} \\ &+ \underset{(12.524)}{25.782}(1+\exp\{-\underset{(0.158)}{4.569}(t/T-\underset{(0.006)}{0.585})\})^{-1} \\ &- \underset{(12.682)}{23.616}(1+\exp\{-\underset{(0.375)}{4.767}(1+(t/T-\underset{(0.007)}{0.607})\})^{-1}.\end{aligned}$$

The locations of the transitions are remarkably similar across transitions. The first transition of the WBC equation is very slow. The effect of the transition extends over the whole estimation period and is the reason for the post-crisis decline in the value of $\widehat{g}_{4t}$; see Figure 5.

F Tables

		CEC33			CEC67			CTC50			CTC90		
N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
2	25	0.023	0.076	0.132	0.022	0.069	0.128	0.024	0.074	0.130	0.022	0.070	0.126
	50	0.015	0.063	0.116	0.016	0.064	0.115	0.016	0.064	0.115	0.015	0.062	0.109
	100	0.011	0.056	0.104	0.010	0.054	0.102	0.011	0.056	0.103	0.010	0.051	0.101
	250	0.012	0.055	0.108	0.010	0.054	0.107	0.011	0.055	0.106	0.009	0.053	0.108
	500	0.010	0.051	0.097	0.009	0.049	0.097	0.010	0.050	0.096	0.009	0.050	0.094
	1000	0.010	0.048	0.099	0.010	0.048	0.095	0.010	0.046	0.097	0.010	0.049	0.092
5	100	0.011	0.054	0.112	0.011	0.053	0.110	0.011	0.056	0.112	0.011	0.053	0.111
	250	0.014	0.054	0.099	0.012	0.051	0.099	0.013	0.053	0.100	0.012	0.051	0.101
	500	0.010	0.050	0.104	0.010	0.053	0.106	0.009	0.052	0.101	0.010	0.054	0.105
	1000	0.010	0.056	0.102	0.010	0.052	0.103	0.009	0.053	0.100	0.008	0.053	0.103
10	250	0.013	0.055	0.112	0.013	0.057	0.112	0.013	0.057	0.110	0.012	0.054	0.115
	500	0.009	0.049	0.101	0.010	0.049	0.104	0.008	0.053	0.103	0.010	0.050	0.103
	1000	0.011	0.052	0.102	0.011	0.054	0.105	0.011	0.053	0.099	0.012	0.056	0.103
20	1000	0.012	0.056	0.106	0.012	0.057	0.106	0.013	0.056	0.103	0.012	0.056	0.107

Table 5: Size-study: Test of constant correlations. Data is generated as an MTV-CCC with an equicorrelation coefficient of 0.33 (CEC33) and 0.67 (CEC67) and a Toeplitz structure with a correlation coefficient of 0.5 (CTC50) and 0.9 (CTC90). Tests are based on the first order polynomial approximation. 5000 replications

			CEC33						CEC67					
			kurtosis=4			kurtosis=6			kurtosis=4			kurtosis=6		
	N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
p=0.95	2	500	0.012	0.056	0.108	0.016	0.056	0.106	0.016	0.070	0.122	0.016	0.092	0.122
	2	1000	0.009	0.044	0.103	0.009	0.042	0.097	0.011	0.045	0.093	0.009	0.044	0.097
	2	2000	0.008	0.042	0.094	0.007	0.042	0.090	0.010	0.052	0.099	0.009	0.046	0.092
	5	500	0.006	0.062	0.118	0.006	0.070	0.114	0.018	0.076	0.140	0.018	0.082	0.146
	5	1000	0.016	0.060	0.119	0.016	0.061	0.112	0.016	0.059	0.115	0.018	0.060	0.112
	5	2000	0.010	0.058	0.108	0.008	0.051	0.102	0.016	0.060	0.116	0.010	0.052	0.098
	10	500	0.016	0.058	0.118	0.020	0.064	0.114	0.020	0.068	0.116	0.024	0.080	0.128
	10	1000	0.018	0.053	0.104	0.015	0.051	0.101	0.014	0.061	0.111	0.017	0.063	0.110
	10	2000	0.014	0.072	0.126	0.012	0.060	0.112	0.018	0.082	0.142	0.013	0.062	0.118
p=0.97	2	500	0.010	0.056	0.114	0.012	0.054	0.118	0.020	0.072	0.114	0.014	0.068	0.120
	2	1000	0.011	0.043	0.102	0.011	0.044	0.103	0.012	0.047	0.107	0.013	0.048	0.103
	2	2000	0.009	0.046	0.094	0.007	0.042	0.089	0.010	0.056	0.108	0.012	0.050	0.093
	5	500	0.004	0.066	0.124	0.012	0.056	0.104	0.012	0.088	0.152	0.018	0.086	0.164
	5	1000	0.015	0.063	0.113	0.014	0.067	0.114	0.018	0.063	0.121	0.019	0.060	0.125
	5	2000	0.010	0.060	0.110	0.008	0.050	0.100	0.015	0.060	0.118	0.012	0.050	0.101
	10	500	0.012	0.062	0.108	0.016	0.070	0.112	0.016	0.072	0.112	0.022	0.086	0.148
	10	1000	0.016	0.053	0.100	0.015	0.056	0.107	0.015	0.063	0.113	0.018	0.057	0.110
	10	2000	0.015	0.074	0.132	0.014	0.058	0.108	0.016	0.088	0.142	0.010	0.063	0.112

Table 6: Size-study: Test of constant correlations. Data is generated as an MTV-GARCH-CEC with persistence (p) of 0.95 and 0.97, kurtosis of 4 and 6, and an equicorrelation coefficient of 0.33 and 0.67. Tests are based on the first order polynomial approximation. 2500 replications.

			CTC50						CTC90					
			kurtosis=4			kurtosis=6			kurtosis=4			kurtosis=6		
	N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
p=0.95	2	500	0.010	0.064	0.102	0.010	0.070	0.106	0.018	0.094	0.136	0.026	0.088	0.146
	2	1000	0.009	0.041	0.097	0.011	0.042	0.103	0.014	0.053	0.096	0.020	0.062	0.104
	2	2000	0.008	0.044	0.096	0.009	0.044	0.090	0.017	0.066	0.120	0.014	0.048	0.098
	5	500	0.006	0.062	0.118	0.010	0.058	0.114	0.020	0.120	0.212	0.050	0.134	0.210
	5	1000	0.014	0.060	0.112	0.018	0.064	0.113	0.027	0.076	0.134	0.034	0.093	0.144
	5	2000	0.011	0.057	0.110	0.008	0.052	0.105	0.020	0.075	0.142	0.018	0.058	0.110
	10	500	0.012	0.070	0.120	0.016	0.080	0.128	0.040	0.114	0.172	0.078	0.150	0.230
	10	1000	0.012	0.049	0.100	0.013	0.051	0.102	0.019	0.078	0.127	0.032	0.089	0.147
	10	2000	0.019	0.072	0.127	0.014	0.059	0.111	0.033	0.110	0.178	0.018	0.077	0.140
p=0.97	2	500	0.014	0.066	0.114	0.018	0.068	0.116	0.016	0.082	0.134	0.030	0.104	0.164
	2	1000	0.009	0.044	0.101	0.008	0.042	0.099	0.016	0.051	0.112	0.022	0.063	0.119
	2	2000	0.010	0.050	0.102	0.009	0.044	0.092	0.024	0.070	0.120	0.015	0.052	0.100
	5	500	0.014	0.056	0.128	0.008	0.074	0.130	0.024	0.134	0.208	0.052	0.160	0.256
	5	1000	0.013	0.059	0.112	0.016	0.066	0.123	0.022	0.082	0.157	0.037	0.102	0.164
	5	2000	0.014	0.062	0.112	0.010	0.052	0.101	0.028	0.086	0.145	0.020	0.066	0.116
	10	500	0.018	0.080	0.128	0.022	0.088	0.130	0.040	0.114	0.172	0.100	0.188	0.278
	10	1000	0.012	0.054	0.105	0.013	0.054	0.107	0.019	0.078	0.127	0.030	0.104	0.181
	10	2000	0.016	0.072	0.132	0.016	0.062	0.110	0.033	0.110	0.178	0.026	0.089	0.150

Table 7: Size-study: Test of constant correlations. Data is generated as an MTV-GARCH-CTC with persistence (p) of 0.95 and 0.97, kurtosis of 4 and 6, and a correlation matrix with a Toeplitz structure with a correlation coefficient of 0.5 and 0.9. Tests are based on the first order polynomial approximation. 2500 replications.

G Figures

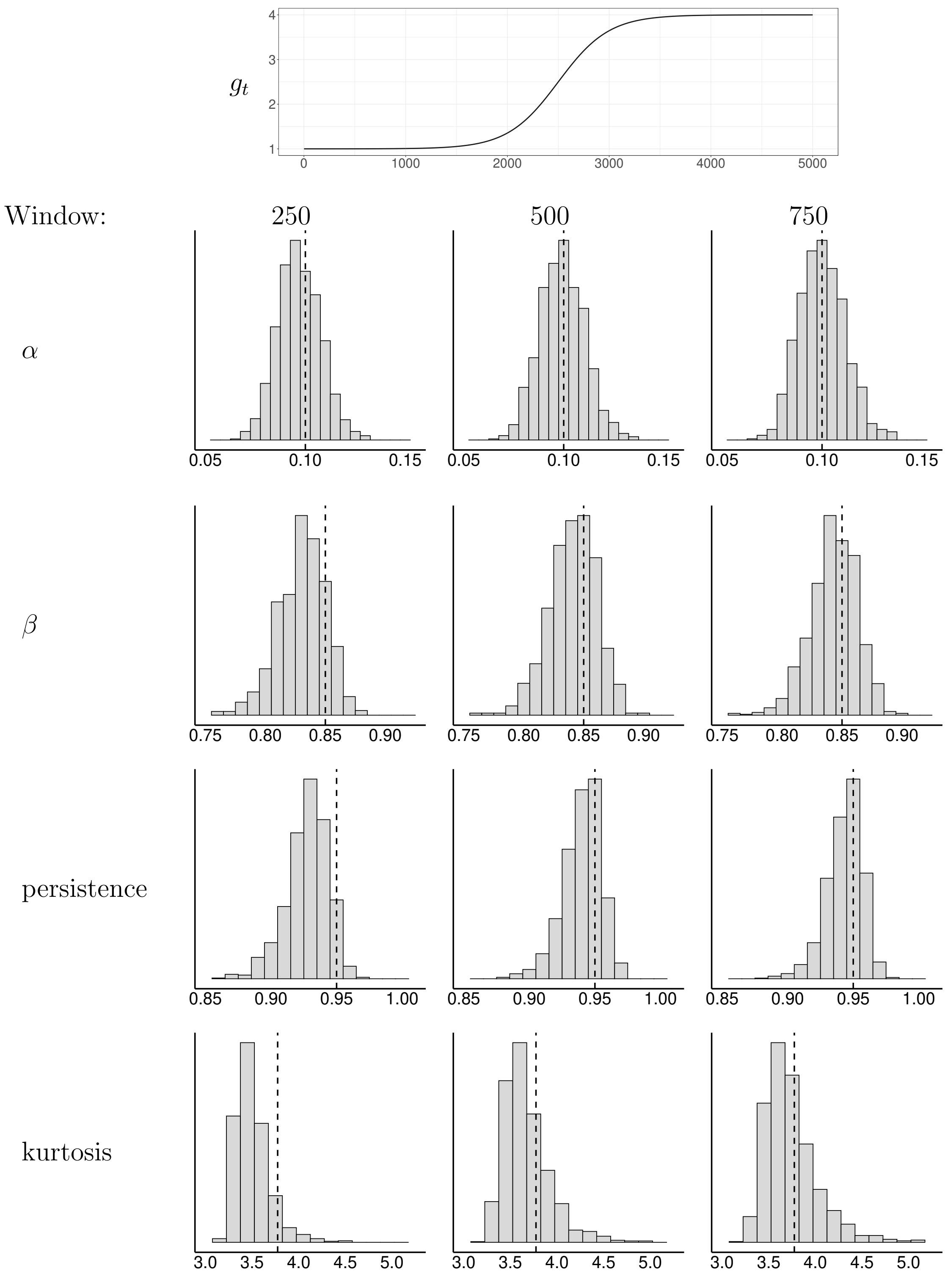

Figure 9: Simulated distributions of GARCH estimates and implied persistence and kurtosis measures for a selection of window widths. The baseline g_t has a single transition. The dotted vertical lines indicate the true values of the parameters α, β, persistence and kurtosis.

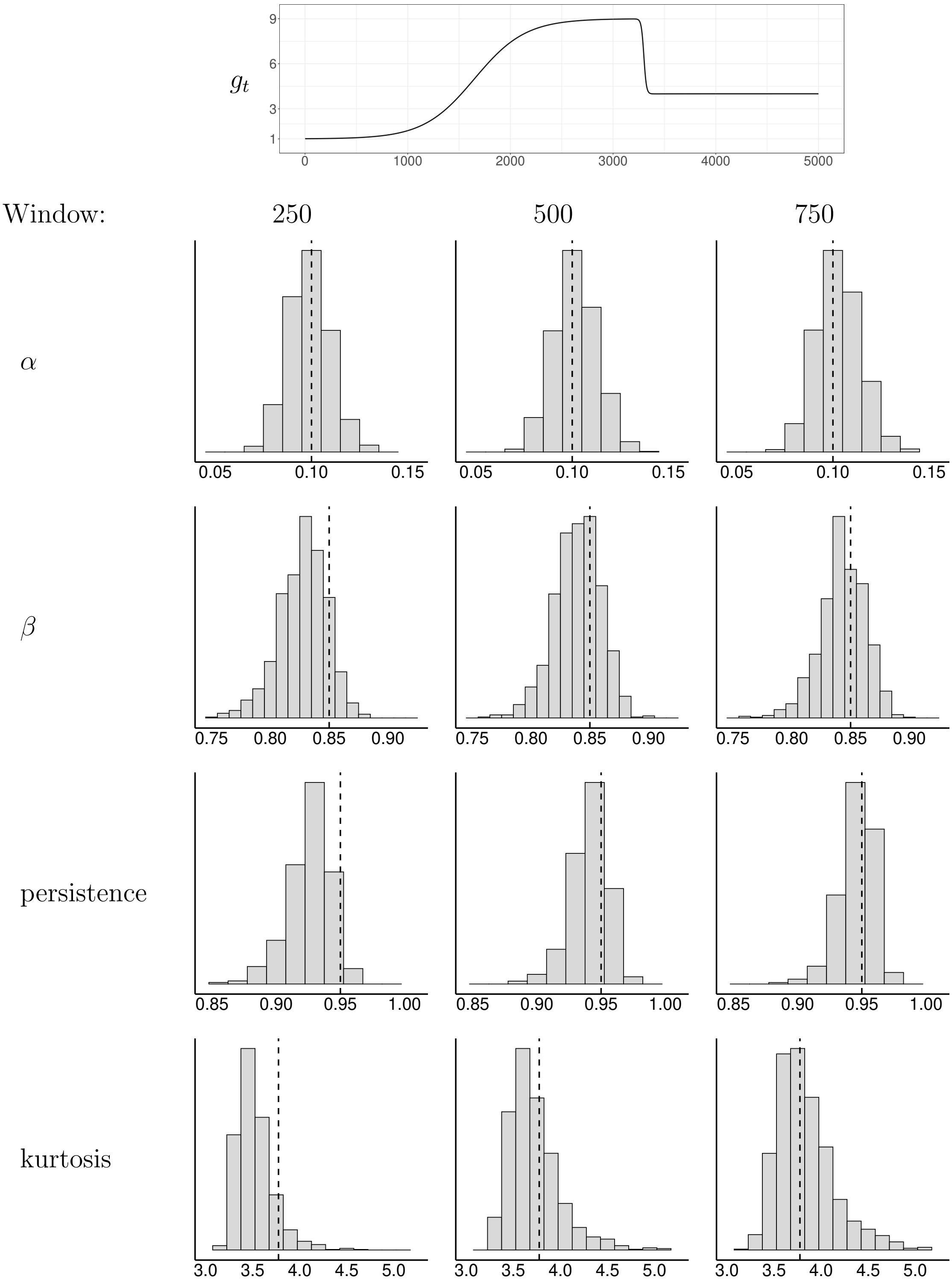

Figure 10: Simulated distributions of GARCH estimates and implied persistence and kurtosis measures for a selection of window widths. The baseline g_t has an asymmetric double transition. The dotted vertical lines indicate the true values of the parameters α, β, persistence and kurtosis.

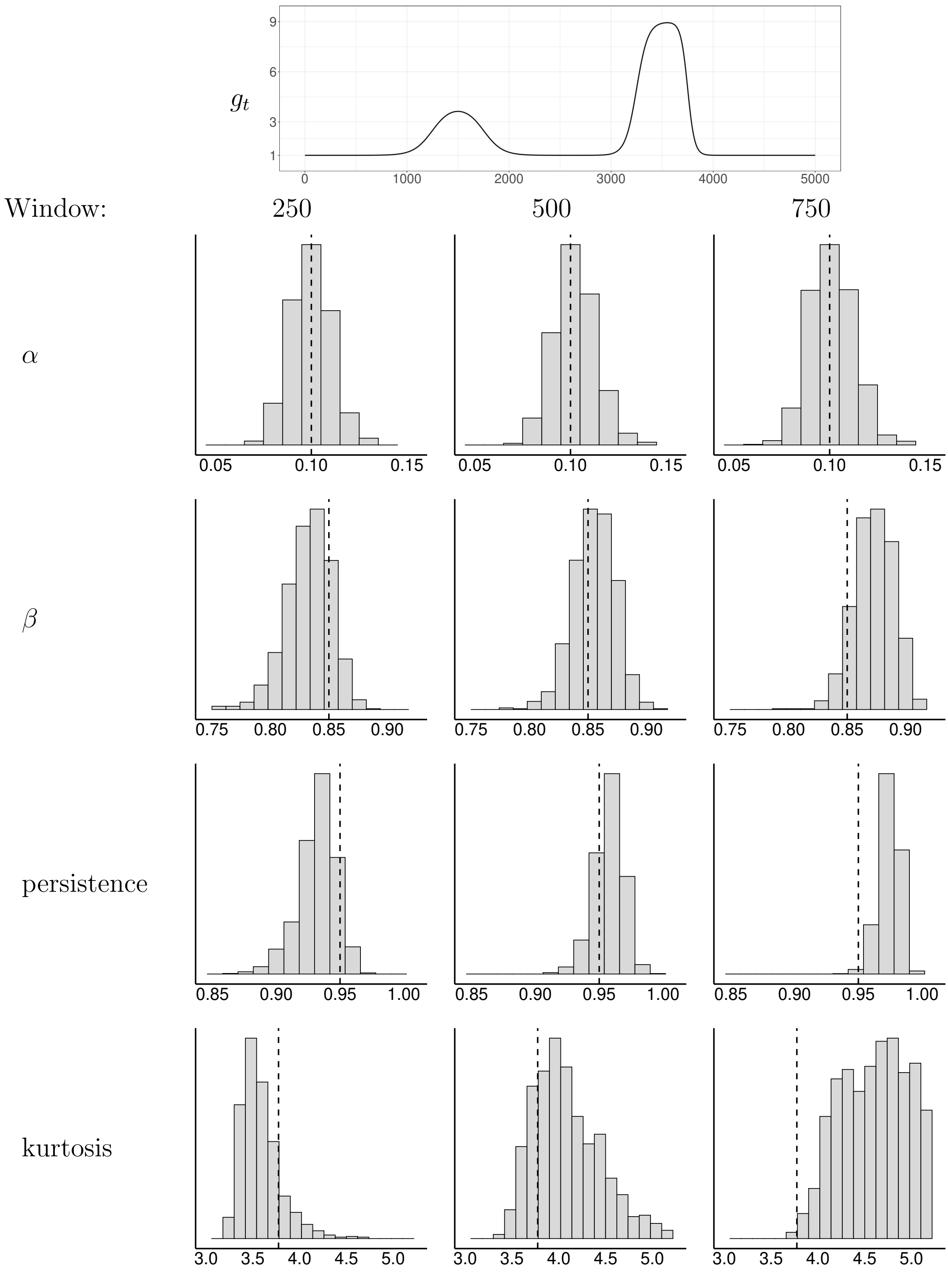

Figure 11: Simulated distributions of GARCH estimates and implied persistence and kurtosis measures for a selection of window widths. The baseline g_t has two double transitions. The dotted vertical lines indicate the true values of the parameters α, β, persistence and kurtosis.

Research Papers 2021

Center for Research in Econometric Analysis of Time Series

2020-17: Jesús-Adrián Álvarez, Malene Kallestrup-Lamb and Søren Kjærgaard: Linking retirement age to life expectancy does not lessen the demographic implications of unequal lifespans

2020-18: Mikkel Bennedsen, Eric Hillebrand and Siem Jan Koopman: A statistical model of the global carbon budget

2020-19: Eric Hillebrand, Jakob Mikkelsen, Lars Spreng and Giovanni Urga: Exchange Rates and Macroeconomic Fundamentals: Evidence of Instabilities from Time-Varying Factor Loadings

2021-01: Martin M. Andreasen: The New Keynesian Model and Bond Yields

2021-02: Daniel Borup, David E. Rapach and Erik Christian Montes Schütte: Now- and Backcasting Initial Claims with High-Dimensional Daily Internet Search-Volume Data

2021-03: Kim Christensen, Mathias Siggaard and Bezirgen Veliyev: A machine learning approach to volatility forecasting

2021-04: Fabrizio Iacone, Morten Ørregaard Nielsen and Robert Taylor: Semiparametric Tests for the Order of Integration in the Possible Presence of Level Breaks

2021-05: Stefano Grassi and Francesco Violante: Asset Pricing Using Block-Cholesky GARCH and Time-Varying Betas

2021-06: Gloria González-Rivera, Carlos Vladimir Rodríguez-Caballero and Esther Ruiz Ortega: Expecting the unexpected: economic growth under stress

2021-07: Matei Demetrescu and Robinson Kruse-Becher: Is U.S. real output growth really non-normal? Testing distributional assumptions in time-varying location-scale models

2021-08: Luisa Corrado, Stefano Grassi and Aldo Paolillo: Modelling and Estimating Large Macroeconomic Shocks During the Pandemic

2021-09: Leopoldo Catania, Alessandra Luati and Pierluigi Vallarino: Economic vulnerability is state dependent

2021-10: Søren Johansen and Anders Ryghn Swensen: Adjustment coefficients and exact rational expectations in cointegrated vector autoregressive models

2021-11: Bent Jesper Christensen, Mads Markvart Kjær and Bezirgen Veliyev: The incremental information in the yield curve about future interest rate risk

2021-12: Mikkel Bennedsen, Asger Lunde, Neil Shephard and Almut E. D. Veraart: Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics

2021-13: Anthony D. Hall, Annastiina Silvennoinen and Timo Teräsvirta: Four Australian Banks and the Multivariate Time-Varying Smooth Transition Correlation GARCH model